IDENTIFYING MELANOMA CHARACTERISTICS USING DIRECTIONAL IMAGING ALGORITHM AND CONVOLUTIONAL NEURAL NETWORK ON DERMOSCOPIC IMAGES

MOHAMMAD ASADUZZAMAN RASEL

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITI MALAYA
KUALA LUMPUR

2024

IDENTIFYING MELANOMA CHARACTERISTICS USING DIRECTIONAL IMAGING ALGORITHM AND CONVOLUTIONAL NEURAL NETWORK ON DERMOSCOPIC IMAGES

MOHAMMAD ASADUZZAMAN RASEL

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITI MALAYA
KUALA LUMPUR

2024

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Mohammad Asaduzzaman Rasel

Matric No: WVA190006/ 17201889

Name of Degree: Doctor of Philosophy

Title of Thesis: Improving the Diagnosis of Melanoma Using Directional Imaging Algorithm and Deep Convolutional Neural Network on Dermoscopic Images

Field of Study: Artificial Intelligence.

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature   Date: 08/07/2024

Subscribed and solemnly declared before,

Witness's Signature   Date:

Name:

Designation:

IDENTIFYING MELANOMA CHARACTERISTICS USING DIRECTIONAL IMAGING ALGORITHM AND CONVOLUTIONAL NEURAL NETWORK ON DERMOSCOPIC IMAGES

ABSTRACT

Melanoma is the deadliest skin cancer worldwide. Advancements in digital dermoscopic image analysis have greatly improved computer-aided Melanoma diagnosis systems. The use of dermoscopic images early detection of Melanoma has gained popularity among researchers due to its non-invasive nature. This thesis aims to enhance the analysis of dermoscopic images for identifying Melanoma. A critical first step for this is to distinguish between healthy and unhealthy skin areas by improving the segmentation process. This is followed by lesion features extraction and analysis (including classification) based on clinically diagnosis criteria including ABCDE rules, 3-point checklist, 7-point checklist, and CASH, to automate the manual process. This research is divided into two phases – 1) Feature Engineering phase explains skin conditions based on lesion segmentation and different dermoscopic feature extraction, while 2) Classification phase detects Melanoma. Multiple deep-learning models are proposed for segmentation. Several imaging, computer vision, and pattern recognition algorithms are employed to describe five dermoscopic features. Subsequently, these features are classified using different proposed deep learning models on various publicly available datasets. To overcome the issues with non-annotated dataset, several techniques are proposed. Both phases of the research outputs are evaluated and compared with the state-of-the-art methods. The proposed algorithms that outperformed the state-of-the-art algorithms contributes to diagnosing early-stage Melanoma. Findings from this study would help dermatologists and patients reduce the time and cost of Melanoma diagnosis, while receiving explanation for such automated diagnosis. Only five most common and important features of many Melanoma-features are analyzed. As part of future work,

incorporating more dermoscopic features such as irregular blotches and regression structures in the analytical section would be promising.

Keywords: Melanoma; dermoscopic features; lesion segmentation; image processing; deep learning.

MENGENAL PASTI CIRI-CIRI MELANOMA MENGGUNAKAN ALGORITMA PENGIMEJAN ARAH DAN RANGKAIAN SARAF KONVOLUSI PADA IMEJ DERMOSKOPI

ABSTRAK

Melanoma adalah kanser kulit yang paling berbahaya di seluruh dunia. Kemajuan dalam analisis imej dermoskopik digital dibantu oleh komputer telah meningkatkan sistem diagnosis Melanoma dengan ketara. Penggunaan imej dermoskopik dalam pengesanan awal Melanoma telah mendapat banyak perhatian daripada komuniti penyelidik kerana sifatnya yang tidak invasif. Tesis ini bertujuan untuk meningkatkan analisis imej dermoskopik untuk pengesanan isu kulit yang lebih baik berkaitan dengan Melanoma. Oleh itu, adalah penting untuk membezakan antara kawasan kulit yang sihat dan tidak sihat dengan memperbaiki proses segmentasi. Tambahan pula, terdapat keperluan untuk pengekstrakan dan analisis ciri lesi (termasuk klasifikasi) berdasarkan kriteria diagnosis Melanoma. Proses analisis ciri ini tidak seharusnya menjadi alternatif kepada keputusan pakar, tetapi akan menjadikan keputusan mereka lebih telus dan kukuh dengan penjelasan yang betul. Teknik analisis ciri berkomputer harus mengikuti kaedah klinikal termasuk peraturan ABCDE, senarai semak 3 mata, senarai semak 7 mata, dan CASH, bagi mengautomasikan proses manual. Ujian saringan adalah perlu untuk mengenali Melanoma pada imej dermoskopik dengan tepat pada masanya. Penyelidikan ini dibahagikan kepada dua fasa di mana fasa pertama menerangkan keadaan kulit berdasarkan segmentasi lesi dan pengekstrakan ciri dermoskopik yang berbeza. Pelbagai model pembelajaran mendalam dicadangkan untuk segmentasi. Pelbagai algoritma pengimejan, penglihatan komputer, dan pengecaman corak digunakan untuk menerangkan lima ciri dermoskopik, dan kemudian ciri-ciri tersebut diklasifikasikan oleh pelbagai model pembelajaran mendalam yang dicadangkan berdasarkan pelbagai set data yang tersedia secara umum. Untuk mengatasi isu dengan set data yang tidak beranotasi,

beberapa teknik dicadangkan. Fasa kedua mengklasifikasikan imej lesi kulit untuk mengesan Melanoma. Kedua-dua fasa hasil penyelidikan dinilai dan dibandingkan dengan kaedah terkini. Algoritma yang dicadangkan mengatasi prestasi algoritma terkini. Beberapa set data lesi kulit yang tersedia secara umum dianalisis untuk mengesan Melanoma. Penyelidikan ini adalah kajian yang berkesan untuk analisis lesi kulit yang menyumbang kepada diagnosis Melanoma peringkat awal. Diharapkan hasil kajian ini akan membantu pakar dermatologi dan pesakit mengurangkan masa dan kos diagnosis Melanoma. Sebagai sebahagian daripada kerja masa depan, menggabungkan lebih banyak ciri dermoskopik seperti tompok tidak teratur dan struktur regresi dalam bahagian analisis memberi potensi yang luas.

Kata kunci: Melanoma; ciri dermoskopik; segmentasi lesi; pemprosesan imej; pembelajaran mendalam.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AHE	:	Adaptive histogram equalization
ANN	:	Artificial Neural Network
APN	:	Atypical Pigment Network
AUC	:	Area Under the Curve
BoF	:	Bag of Features
BWV	:	Blue white veil
CAD	:	Computer Aided Diagnosis
CADs	:	Computer Aided Diagnosis Systems
CASH	:	Color, Architecture, Symmetry, Homogeneity
CIE	:	International Commission on Illumination
CNN	:	Convolutional Neural Network
DNA	:	Deoxyribonucleic acid
FFN	:	Feedforward Neural Network
FPR	:	False Positive Rate
GAN	:	Generative adversarial networks
LIME	:	Locally Interpretable Model-agnostic Explanations
LMBP	:	Levenberg-Marquardt Backpropagation
PCA	:	Principal Component Analysis
PN	:	Pigment Network
RGB	:	Red green blue
ROC	:	Receiver Operating Characteristic
ROI	:	Region of Interest
SGD	:	Stochastic Gradient Descent
SGDM	:	Stochastic Gradient Descent with Momentum

SHAP	:	SHapley Additive exPlanation
SVM	:	Support Vector Machine
TPN	:	Typical Pigment Network
WHO	:	World Health Organization

1.1   Background

According to the World Health Organization (WHO), cancer is the second most significant cause of death and morbidity in Europe and one of the leading causes of death globally. In the following two decades, according to "WHO" predictions (Ferlay et al., 2013), the number of cancer diagnoses will increase by a factor of two. It is crucial to focus research efforts on the creation of early cancer detection strategies because cancer mortality can be significantly decreased if cases are identified and treated in their early stages. Due to its high propensity to metastasize, Melanoma is the most lethal type of skin cancer. According to Ferray et al. (2013), it is the ninth most prevalent type of cancer in Europe. Worldwide, there are about 325000 new cases (174000 males and 151000 females) diagnosed each year. For the past 30 years, Melanoma incidence rates have been steadily rising in line with the general trend. A simple excision can cure a Melanoma that has been detected early, unlike other cancer types where the survival rate is lower in the advanced stages (Lio & Nghiem, 2004). The development of new imaging techniques that enable better visualization of skin lesions and automated methods to diagnose Melanoma has been heavily influenced by these statistics.

One of the most common imaging methods employed by dermatologists is dermoscopy. It permits both the observation of surface and subsurface structures that are invisible to the naked eye and the magnification of skin lesions at the same time (Lio & Nghiem, 2004; Kopf et al., 1997). Dermoscopy significantly improves dermatologists' diagnostic accuracy, according to numerous studies (Mayer, 1997). However, because it depends on the practitioner's visual acuity, this method can only be effectively used by trained medical professionals, and the diagnosis is frequently arbitrary and challenging to replicate (Wolf et al., 1998). These limitations encouraged the development of computer-aided diagnosis (CAD) systems that can be used by novice dermatologists as a second

opinion tool. Figure 1.1 shows different types of dermatoscope devices. Using those devices, the taken images are known as dermoscopy images. Usually, a digital camera or mobile phone camera is used with the help of dermatoscope devices to take pictures of skin lesions so that a certain distance and exact measures are always maintained. Figure 1.2 shows some images taken using dermatoscope devices.

Figure 1.1: Different types of dermatoscope devices used by mobile phone cameras and digital cameras.
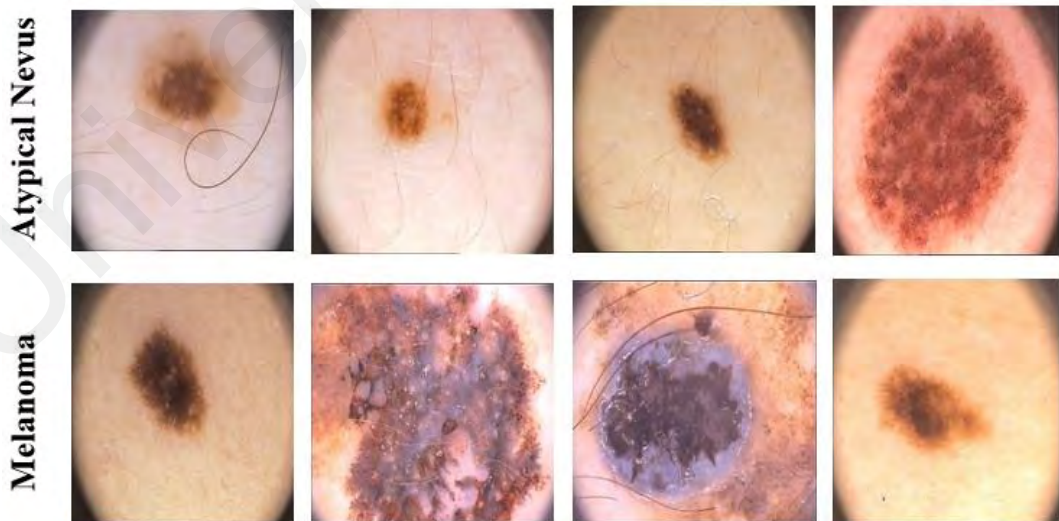
Figure 1.2: Different dermoscopy images of skin lesions such as Melanoma and Atypical Nevus.

Artifact removal, lesion segmentation, feature extraction, and lesion binary classification (benign or malignant) are the most common sequential steps used by CAD systems in the literature (Korotkov & Garcia, 2012; Oliveira et al., 2016). These systems perform well under test conditions, but they typically extract features that are either unrelated to medicine or difficult to explain to doctors (Frühauf et al., 2011).

Additionally, the feature extraction processes are frequently not sufficiently described (Celebi et al., 2007), which makes it difficult for other users to use the algorithms. Alternately, systems that concentrate on the extraction of pertinent dermoscopic features, the foundation of medical diagnosis, could be created. These systems have received less attention in the literature, despite being close to the point of diagnosis in medicine and having consequent clinical relevance. Additionally, these suggested methods typically do not attempt to diagnose the lesions based on the clinical data, instead attempting to detect only one or two dermoscopic criteria. For these two different categories of CAD systems, this thesis suggests new contributions.

1.2   Skin Cancer: Melanoma

Cancer develops when the body's cells begin to multiply out of control. Nearly any area of the body can develop cancerous cells, which can then spread to other organs (Tarver, 2012). Skin cancer is a condition in which uncontrolled growth of abnormal skin cells occurs. It takes place when irreparable DNA (Deoxyribonucleic acid) damage to skin cells, most often brought on by ultraviolet radiation from sunlight or tanning beds, results in mutations or genetic defects that cause the skin cells to multiply rapidly and form malignant tumors (Schadendorf et al., 2023).

The identification of skin lesion cases is complicated by the abundance of benign (harmless) and malignant (harmful) Melanoma. Basal cell carcinoma, squamous cell carcinoma, and Melanoma are the three primary types of abnormal skin cells that are seen in clinical settings (Tarver, 2012; Cavalcanti & Scharcanski, 2011). The Skin Cancer Foundation (Schadendorf et al., 2023) describes three additional abnormal cell types that are less common, including actinic keratosis, Merkel cell carcinoma, and atypical moles. Six different categories of skin lesions are shown in Figure 1.3. Atypical moles are the second most dangerous cells after Melanoma cases, according to research.

(a) Actinic Keratosis   (b) Merkel cell carcinoma   (c) Basal cell carcinoma

(d) Squamous cell carcinoma   (e) Atypical moles   (f) Melanoma
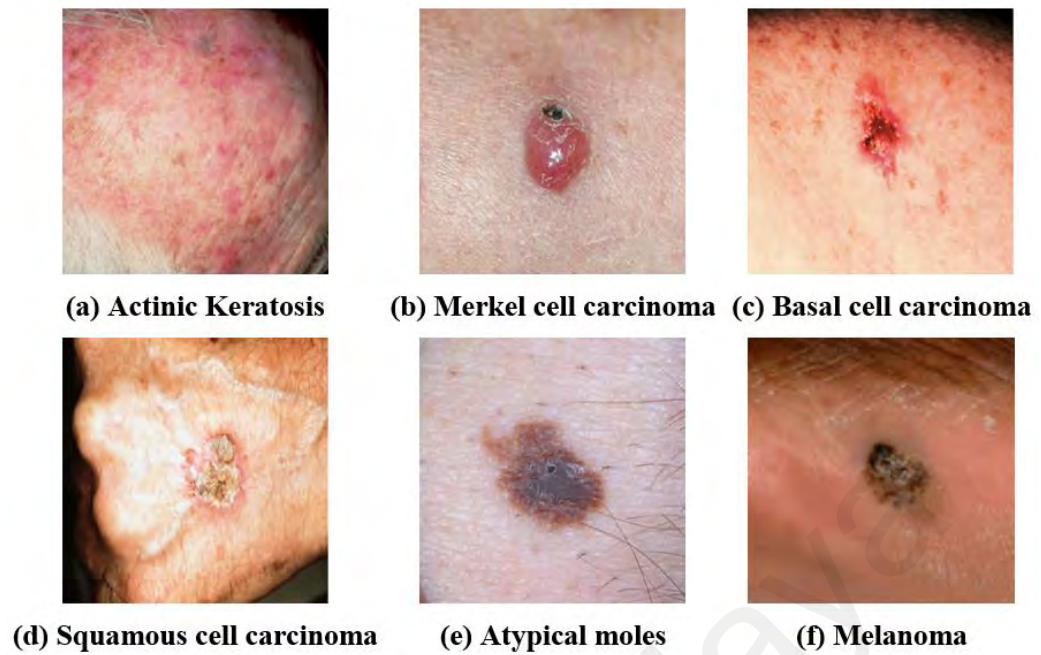
Figure 1.3: Example of different skin cancers, classified by the Skin Cancer Foundation.

According to Skin Cancer Foundation (Schadendorf et al., 2023), the differences between these abnormal lesions are:

(a) Actinic keratosis, also referred to as solar keratosis, is an ecchymose growth. It is regarded as a pre-cancer because, if left untreated, skin cancer could develop from it.

(b) Merkel cell carcinoma is a rare, severe, and aggressive form of skin cancer that has a high risk of returning and metastasizing (spreading) to other parts of the body. It is, however, 40 times more uncommon than Melanoma.

(c) The most common type of skin cancer is Basal cell carcinoma. Open sores, red patches, pink growths, shiny bumps, or scars are some of the common appearances. Rarely does this skin cancer spread.

(d) As far as skin cancer goes, squamous cell carcinoma is the second most prevalent type. It frequently resembles warts, elevated growths with a central depression, open sores, or scaly red patches.

(e) Atypical moles, also referred to as dysplastic nevi, are benign moles with unusual appearances. People who have them have a higher risk of developing Melanoma in a mole or elsewhere on the body because they may resemble Melanoma. They are at least ten times more likely to develop Melanoma.

(f) The most common cause of Melanoma, a dangerous type of skin cancer, is unrepaired DNA damage to skin cells due to ultraviolet radiation from sunlight or tanning beds. This damage triggers genetic defects which cause skin cells to quickly multiply and form malignant tumors. Melanomas can appear in a variety of colors including skin-colored, pink, red, purple, blue, white, black, or brown. Early detection and treatment are critical for a successful cure; otherwise, the cancer may spread to other parts of the body and become difficult to treat.

It is noticed that the first cause of these skin cancer types is caused by damage to skin tissue from exposure to ultraviolet radiation (Tarver, 2012; Schadendorf et al., 2023; Al-Ahmad et al., 2013; Boustany, 2013). Among different types of skin cancers, Melanoma is particularly important to diagnose due to its potential for aggressive growth and metastasis. Rapid progression, high mortality rate, visible signs, and limited treatment options for advanced stages highlight the potential severity of Melanoma and its impact on patient outcomes. Early diagnosis is crucial to provide timely treatment, prevent metastasis, and improve long-term survival rates.

1.3   Clinical Diagnosis

In the medical field, dermatologists compare and analyze multiple signs of each tumor to recognize Melanoma cases. Mostly, four scoring systems have been used based on visual characteristics to recognize the Melanoma tissue, which are the ABCDE rule (Schadendorf et al., 2023; Cavalcanti & Scharcanski, 2011; Friedman et al., 1985; Stolz, 1995; Gadeliya Goodson & Grossman, 2009), 7-point checklist (Cavalcanti & Scharcanski, 2011; Maglogiannis & Doukas, 2009; Argenziano et al., 1998; Mackie & Doherty, 1991), 3-point checklist (Soyer et al., 2004), and CASH (Henning et al., 2008). Medical professionals are aware of these four scoring systems, and image-processing professionals use them as their primary means of comparison. Most of the developed and explored methods are based on these systems, such as Bareiro et al. (2016) and She et al. (2007) for the ABCDE rule (asymmetry, border regularity, color, diameter or size, and

evolve), and Argenziano et al. (2011) for the 7-point checklist (pigment network, blue white veil, vascular pattern, streak, dots-globules, blotches, and regression structure). The features used in these systems are also explored separately for Melanoma recognition, for example, pigment network (Arroyo & García Zapirain, 2014; Garcia-Arroyo & Garcia-Zapirain, 2018) and vascular structure (Argenziano et al., 2004). These systems are summarized below and detailed in Chapter 2: Literature Review. Figure 1.4 shows the four most common dermoscopic features to diagnose Melanoma in the early stage.
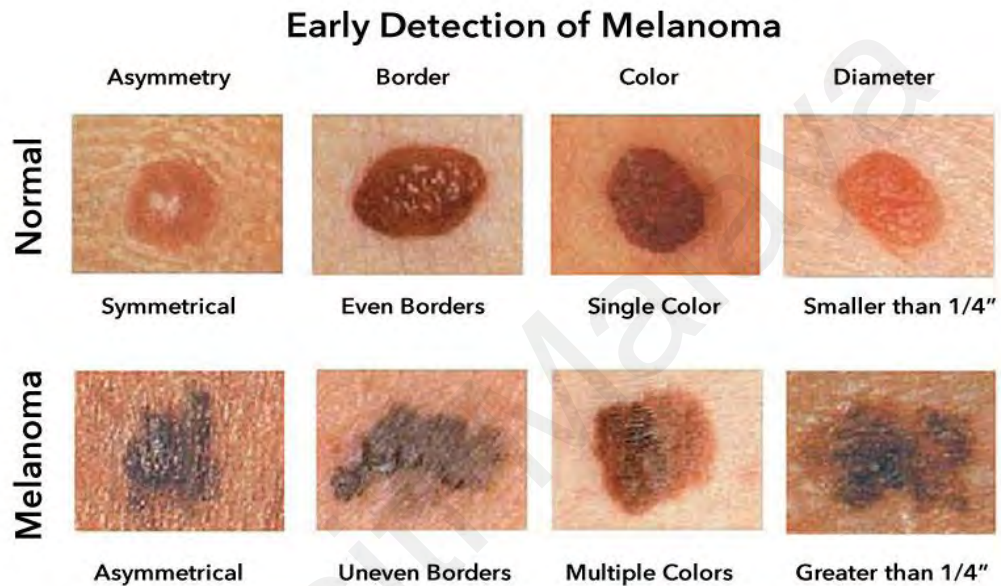
Figure 1.4: Clinically practiced the most common features for diagnosing Melanoma.

Method 1: The ABCDE rule has a 95% success rate (Schadendorf et al., 2023; Cavalcanti & Scharcanski, 2011; Friedman et al., 1985; Stolz, 1995; Gadeliya Goodson & Grossman, 2009).

1. A is for asymmetrical shape.

2. B is for irregular borders.

3. C is for color variation (i.e., white, red, light brown, dark brown, blue-gray).

4. D is for diameter.

5. E is for evolving.

If any of the signs and symptoms are noticed can be suspected as Melanoma.

Method 2: The 7-point checklist has a 95% success rate (Cavalcanti & Scharcanski, 2011; Maglogiannis & Doukas, 2009; Argenziano et al., 1998; Mackie & Doherty, 1991)

1. Atypical pigment network (2)

2. Blue white veil (2)

3. Atypical vascular pattern (2)

4. Irregular streaks (1)

5. Irregular dots/globules (1)

6. Irregular blotches (1)

7. Regression structures (1)

The number inside of the first parenthesis indicates the score. If the total score of all points is 3 or more, then it is considered Melanoma.

Method 3: The 3-point checklist has a 96.3% success rate (Soyer et al., 2004).

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

2. Atypical network: pigment network with irregular holes and thick lines.

3. Blue white structures: any type of blue and/or white color, i.e., a combination of the blue white veil and regression structures.

If any lesion has any two symptoms from these three, then it is considered a malignancy (Melanoma or Basal cell carcinoma).

Method 4: The CASH has a 95.6% success rate (Henning et al., 2008).

1. Color: few vs. many. (Single point for each color)

2. Architecture: order vs. disorder. (0 to 2)

3. Symmetry vs. asymmetry. (0 to 2)

4. Homogeneity vs. heterogeneity. (Single point for each structure).

If a score of 7 or less is likely benign. A score of 8 or more is considered suspicious of Melanoma.

By studying these four clinical methods, the most common dermoscopic features or criteria (key indicators to diagnose Melanoma) are found. Improved accuracy, early

detection, differentiation from benign lesions, evaluation of lesion progression, guiding biopsy decisions, and patient reassurance are the reasons why dermoscopic features are important in the diagnosis of Melanoma.

These common features are presented in Table 1.1 and the details are available in Chapter 2: Literature Review.

Table 1.1: The commonalities of different clinical methods.

Features	Sub-Features	ABCDE	7-point checklist	3-point checklist	CASH
Asymmetry	Symmetry, Asymmetry	Yes	No	Yes	Yes
Pigment network	Atypical, Typical	No	Yes	Yes	Yes
Blue white veil	Present, Absent	No	Yes	Yes	No
Colors	Red, Light Brown, Dark Brown, Black, White, Blue-Gray	Yes	No	No	Yes
Dots-globules	Atypical, Typical, Absent	No	Yes	No	Yes

1.4 Computer-Aided Diagnosis Systems

In the field of medical imaging, computer-aided diagnosis (CAD) systems have been created (Doi, 2007; Eadie et al., 2012; Greenspan et al., 2013). These CAD systems' goal is to offer a computer output as a second opinion to help radiologists with interpretation to increase the precision of diagnosis and shorten the reading time for images (Doi, 2005). According to its primary function, a computer-aided detection (CADe) system or computer-aided diagnosis (CADx) system, a CAD system is typically used to distinguish between two types of systems. It is also investigated for early detection and diagnosis (Korotkov et al., 2015). Various tumor images, including those from radiography, dermoscopy, MRIs, and mammograms, are subjected to a CAD system. A CAD system was created in the 1960s, and in the 1980s it developed into an intriguing topic that was the subject of numerous investigations and research projects carried out in some labs and hospitals (Cavalcanti & Scharcanski, 2011). The advancements in melanoma detection and diagnosis have increased the number of papers in a CAD system since the 2000s (Doi,

2005; Doi, 2007). The five main steps of the CAD system used in medical imaging, from image acquisition to classification and diagnosis, are as follows:

1. **Image acquisition:** This stage of image processing is regarded as a data source for the following ones. Additionally, there is a wealth of pertinent and consistent information from the acquisition that can give image-processing specialists a starting point and a natural working direction.

2. **Image enhancement:** This step's goal is to minimize all artifacts and noise in the images. These artifacts, which represent things like numbering in mammogram images and hair and blood vessels in dermoscopic images, depend, for the most part, on the type of image acquisition technique used.

3. **Segmentation:** In the CAD system, segmenting the Region of Interest (ROI) is a crucial step. The segmentation task is made more challenging by the abundance of skin cancer images. It turns into one of the most difficult and difficult tasks in the CAD system.

4. **Feature extraction:** The goal of this step is to provide the best descriptor to extract features that separate the database into two or more classes after the ROI has been isolated.

5. **Classification and detection:** The proposed system is observed and examined in the final steps of classification and diagnosis. The selection of the classifier is essential to achieve better results on the detection and depends on the feature extracted and the desired number of classes. The classification rate is assessed using various performance metrics, including accuracy, sensitivity, specificity, and receiver operating characteristic (ROC), depending on the type of classifier that was used.

Numerous techniques have been adopted or created specifically for Melanoma detection after the system's five steps. The literature review on the methods for detecting Melanoma is summarized in Chapter 2: Literature Review.

1.5   Research Challenge

Challenge 1 (Artifacts): Preserving the quality of dermoscopic images is a big challenge. In most cases, dermoscopic images come with different artifacts such as hair, gel, and light reflection. The presence of artifacts produces difficulty in extracting information from the lesion or region of interest.

Challenge 2 (Color constancy): Color constancy is another issue when dermoscopic images are acquired from two different sources such as images captured by different devices. This is contributed by different environmental conditions such as the light intensity and pixel resolution, hence, producing varying image quality. Different image quality can pose challenges in diagnosing Melanoma through dermoscopy for several reasons: resolution and clarity, distortion and artifacts, color accuracy, artifact suppression, and image consistency.

Challenge 3 (Feature extraction): Another big research challenge is extracting information from dermoscopic images to diagnose specific diseases. Often it is difficult to find the exact feature information due to the lack of appropriate computer-aided programs such as image processing and pattern recognition algorithms.

Challenge 4 (Automation): Currently, the entire dermoscopic image analysis process is either handcrafted or semi-automatic for diagnostic purposes. This analysis method is time-consuming and requires skilled manpower. A research challenge lies in making this process faster and fully automated using machine learning techniques such as deep learning and transfer learning.

Challenge 5 (Interpretation and Explanation): Deep learning algorithms, while highly effective in many cases, can be regarded as classifiers, making it challenging to interpret and explain their decisions. Developing methods to provide explanations or

justifications for the diagnostic decisions made by these algorithms is an active area of research.

1.6　Problem Statement

Several phases are necessary for automating the diagnosis of Melanoma based on dermoscopy images. These are identified as segmentation, feature extraction, and classification. Each has its challenges as described below:

1. Dermoscopic Image Segmentation: The first research problem addressed in this study is the need to automate the process of differentiating healthy and unhealthy (disease-affected area) skin, which is currently a sensitive and complex task due to the similarities in texture and pattern of the skin. The existing approach relies heavily on manual lesion isolation performed in diagnostic centers and clinics, which introduces a high probability of human error. To overcome this limitation, there is a pressing need to develop an automated system or methodology that can offer a more efficient and precise approach to diagnosing skin conditions. The automation of this process would enable accurate identification and isolation of diseased areas, reducing the reliance on subjective human judgment and improving the overall diagnostic accuracy and reliability in the field of dermatology.

2. Feature Extraction: The second research problem addressed in this thesis is the need to develop an efficient and accurate feature extraction method for dermoscopic segmented images, specifically focusing on the identification of asymmetry, pigment network, blue white veil, dots-globules, and colors, which are crucial for diagnosing Melanoma. Current methods for feature extraction involve a combination of different approaches; however, there is a lack of comprehensive and standardized techniques that effectively capture and extract these important features. Consequently, there is a pressing need to investigate and propose novel algorithms or frameworks that can robustly extract

these specific features from dermoscopic images, enabling more accurate and reliable Melanoma diagnosis.

3. Classification: The third research problem addressed here is the need to enhance the classification accuracy and reliability of dermoscopic image analysis for diagnosing skin conditions, particularly Melanoma, by developing an innovative classification framework that effectively utilizes dermoscopic features such as asymmetry, pigment network, bluish veil, colors, and dots-globules. While the current classification methods exist, they suffer from limitations in terms of robustness and standardization in incorporating these dermoscopic features as essential parameters. This limitation hampers the ability to accurately differentiate between typical and atypical lesions, leading to potential misdiagnosis. Thus, there is an urgent need to explore and propose novel approaches or algorithms that can overcome these challenges and improve the accuracy and reliability of classification, thereby enhancing the diagnostic capabilities for Melanoma and other skin conditions based on dermoscopic image analysis.

These challenges are regarded as open issues in Melanoma detection. Thus, these are the key challenges that will be addressed in this thesis.

1.7 Research Question

1. How can advanced image segmentation techniques be utilized to accurately delineate lesion boundaries in dermoscopic images for effective characterization of Melanoma? **RO-1**

2. What dermoscopic features are most indicative of Melanoma, and how can they be accurately extracted from segmented lesions to facilitate automated analysis? **RO-1**

3. How can a novel CNN architecture be designed and optimized to effectively classify dermoscopic images for Melanoma detection, considering the complexity and variability of lesion appearances? **RO-2**

4. What state-of-the-art techniques can be integrated into the CNN model to enhance its performance and robustness in detecting Melanoma across diverse dermoscopic image datasets? **RO-2**

5. How do the performance metrics of the feature-based analysis and CNN-based methods compare to those of state-of-the-art methods in terms of accuracy, sensitivity, specificity, and computational efficiency when applied to Melanoma detection in dermoscopic images from different datasets? **RO-3**

6. What are the potential limitations and challenges associated with both the feature-based analysis and CNN-based approaches in accurately detecting Melanoma in dermoscopic images, and how can these limitations be addressed or mitigated? **RO-3**

7. How do the proposed methodologies perform in handling various confounding factors such as image quality, patient demographics, and lesion characteristics that may influence Melanoma detection accuracy? **RO-3**

8. Can the insights gained from this research contribute to the development of more effective and reliable computer-aided diagnostic tools for Melanoma screening and early detection in clinical practice? **RO-3**

1.8   Research Objectives

1. To perform feature engineering for five dermoscopic features from dermoscopic images based on lesion segmentation and feature-specific deep learning models.

2. To experiment with diverse CNN models for automatic computer-based image classification for Melanoma detection from dermoscopic images.

3. To evaluate feature-engineered and automated feature-identified CNN methods against state-of-the-art techniques in determining the efficacy of Melanoma detection.

1.9   Scope of Research

The scope of this research aims to investigate and evaluate the explanation-based and non-explanation approaches for Melanoma detection. The analysis will focus on analyzing dermoscopy images of a diverse group of patients, including individuals with different skin colors, ages, and genders, to enhance the understanding of Melanoma detection across various demographic groups (different geographic locations). The research will primarily utilize retrospective data available from publicly available online databases, allowing for a comprehensive analysis of a large and diverse set of cases. Furthermore, this study will prioritize the comparison of technical algorithms-based approaches for Melanoma detection, with a lesser emphasis on clinical laboratory-based comparisons. The aim is to explore the effectiveness and potential advantages of algorithmic methods in improving Melanoma detection accuracy and efficiency.

1.10   Research Significant

This research holds significant importance for Melanoma diagnosis as it provides an effective and demystified study for analyzing this form of cancer. It contributes to the early detection of Melanoma, a critical factor in improving patient outcomes. Additionally, the research has the potential to reduce time delays and expenses associated with Melanoma diagnosis, streamlining the diagnostic process and enabling timely interventions.

Furthermore, the significance of this research extends beyond Melanoma alone. The methods and techniques developed in this study can be applied to diagnose other sensitive diseases, including tumors, skin infections, skin rashes, and damaged cells. This widens the scope of the application and enhances the overall impact of the research.

Lastly, this study incorporates various image processing, pattern recognition algorithms, and deep learning models (CNN-based), applying them to real-life scenarios. By practicing these approaches in practical applications, the research ensures their relevance, reliability, and effectiveness in diagnosing Melanoma and potentially other critical diseases.

1.11   Thesis Contribution

Through the development of feature engineering-based and automated methods for dermoscopic image analysis, this research presents a novel and innovative approach to improving the efficacy and accuracy of skin cancer diagnosis, advancing the field of dermatology's diagnostic capabilities. The research contributions are-

1. A novel algorithm for the segmentation of lesions, using semantic segmentation deep network.

2. A novel algorithm for the measurement of lesion asymmetry, using geometry-based shape analysis.

3. A novel algorithm for pigment network detection, using a directional imaging filter and connected component analysis.

4. A novel algorithm for the detection of the Blue white veil, using a specific range of RGB pixel values.

5. A novel algorithm for the classification of dots-globules, using a nested deep learning model.

6. A novel algorithm for the identification of different colors in a lesion, using the regular RGB pixel value.

7. Developing an explainable method for identifying Melanoma based on various feature extraction.

8. Developing a novel deep learning model for classifying skin lesions using several publicly available datasets.

1.12   Thesis Overview

The structure of this thesis is as follows:

Chapter 1 presents an introduction to the proposed Melanoma detection methods and the motivation for this research work. The research background, causes of skin cancer, types of skin cancer, clinical methods for Melanoma diagnosis, CAD system for Melanoma, challenges in the research, and a problem statement are then presented, followed by the goals and scope of the study.

Chapter 2 presents a literature review of two different Melanoma diagnosis methods such as Feature Engineering and Classification-based. In Feature Engineering-based, various image enhancement, segmentation, feature extraction, and classification methods are discussed to detect Melanoma. In Classification-based, different deep learning models are reported on analyzing dermoscopic images to diagnose Melanoma. After the deep analysis of existing works, the research gaps are mapped based on research questions and research objectives.

Chapter 3 represents the research architecture which contains two phases. The first phase has two steps which are segmentation and different feature analysis. The second phase has only one step named Classification (deep learning based). The descriptions of different dermoscopic datasets are given with data preprocessing and data augmentations process. As well as the different evaluation metrics are reported for the proposed methods.

Chapter 4 presents the first step of the first research phase which is the segmentation process to distinguish the healthy and unhealthy areas. Different semantic segmentation models are proposed on different datasets. Also, the proposed methods are compared and discussed with other existing models.

Chapter 5 presents the second step of the first research phase which is the feature extraction, analysis, and classification process. Five different Melanoma diagnosis features are extracted from the dermoscopic images and later those features are analyzed and classified into different classes. Finally, those classification results support the idea of Feature Engineering-based dermoscopic image analysis.

Chapter 6 presents the second research phase which is Classification-based Melanoma detection. Here, a proposed convolutional neural network is presented and trained on different datasets with different activation functions. Various ablation studies are done and compared with other conventional methods to present a Melanoma screening test method.

Chapter 7 presents a summary of the whole research work including two phases of Melanoma detection by different image processing, pattern recognition, and deep learning-based methods, together with the contributions, limitations, recommendations for future works, and conclusion of the proposed algorithms.

CHAPTER 2: LITERATURE REVIEW

2.1 Background

Early Melanoma detection can be achieved through two primary methods: standard clinical methods (ABCDE, 7-point checklist, 3-point checklist, and CASH) and computer-aided methods. The computer-aided methods can be further categorized into two approaches: (a) different dermatological features (asymmetry, pigment network, blue white veil, dots-globules, lesion colors etc.) extraction based to explain lesion condition, also known as *feature engineering-based a*pproach; and (b) the deep learning-based approach, often referred to as the *classification-based* approach. For both approaches, the input is skin lesion images. Therea are mainly two types of skin lesion images for these two approaches, namely, dermoscopic and non-dermoscopic (macroscopic) images. In the literature, several research are done by analyzing these two types of input images. These different approaches for diagnosing Melanoma are illustrated in Figure 2.1.
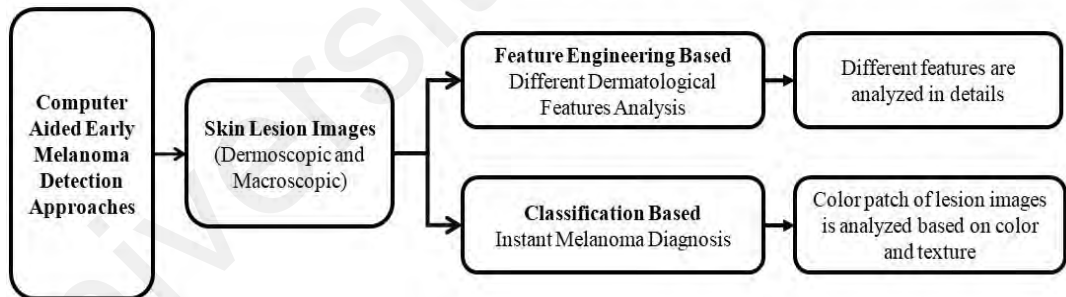
Figure 2.1: Computer-aided early Melanoma diagnosis two approaches.

In Chapter 1: Introduction, the standard clinical methods (rules-based methods) have been discussed, which include ABCDE, 3-point checklist, 7-point checklist, and CASH in section 1.3 Clinical Diagnosis. These methods combine various dermoscopic key features, such as asymmetry, pigment network, blue white veil, dots-globules, and lesion colors (as mentioned in Chapter 1: Introduction).

The computer-aided clinically inspired methods involve these dermoscopic features either individually (single feature) or altogether (multiple features) to detect Melanoma where the input is skin lesion images (both dermoscopic and macroscopic). Section 2.2 "Skin Lesion Image Modalities for Early Detection of Melanoma" has been discussed in detail.

The feature engineering-based approach relies on extracting these dermoscopic key features and subsequently involves a classification process to diagnose Melanoma by interpreting the current condition of the skin lesion. For example, features such as asymmetry, pigment network, blue white veil, dots-globules, and colors can be identified from skin lesion images to characterise Melanoma. More details can be found in Section 2.3: Feature Engineering Based Melanoma Detection. This feature-based detection explains clinically practiced rules-based detection and the decision-making process of deep learning models.

Alternatively, the classification-based method directly detects Melanoma without requiring detailed information about the skin lesions. This approach is particularly suitable for delivering instant results and facilitating mass screening purposes, making it an efficient option for early detection. Section 2.4 "Classification Based Melanoma Detection" has presented an overview.

Overall, these three steps (including input and two approaches) play crucial roles in early Melanoma detection, offering complementary advantages for accurate and timely diagnosis. This chapter will discuss the existing research works related to these steps separately. After analyzing the existing literature, this research aims to identify and highlight the research gaps, thereby emphasizing the scope and contribution of this study.

2.2 Skin Lesion Images Modalities for Early Detection of Melanoma

Melanoma diagnosis using clinically inspired computer-aided systems involves the application of well-established and widely practiced clinical methods to aid in the early detection and assessment of potential skin cancer cases. These methods include the ABCDE criteria (Schadendorf et al., 2023; Cavalcanti & Scharcanski, 2011; Friedman et al., 1985; Stolz, 1995; Gadeliya Goodson & Grossman, 2009), 3-point checklist (Soyer et al., 2004), 7-point checklist (Cavalcanti & Scharcanski, 2011; Maglogiannis & Doukas, 2009; Argenziano et al., 1998; Mackie & Doherty, 1991), and CASH (Henning et al., 2008) to assess different characteristics of skin lesions, helping to identify possible indicators of Melanoma. However, these methods are applied on skin lesion images which are either dermoscopic or non-dermoscopic images. A list of relevant clinically inspired research works is reported in Table 2.1 with a systematic review. The main objective of this table is on the analysis of input images of different computer-aided works and their accuracy (equation 1) in terms of Melanoma detection.

$$Accuracy = \frac{Predicted\ number\ of\ Melanoma}{Total\ number\ of\ Melanoma} \qquad (1)$$

Table 2.1: A review of input images for computer-aided Melanoma detection.

Authors	Features	Dataset	Data type	Techniques	Accuracy	Limitation
Di Leo et al., 2010	7-point checklist	Not reported	Macroscopic	Image Processing, Machine Learning	90.00%	Small dataset, Manual
Raikar et al., 2013	7-point checklist	Not reported	Dermoscopic	Image Processing, Machine Learning	Not reported	Small dataset, Manual
Abbas et al., 2013	CASH	Atlas	Macroscopic	Image Processing, Pattern Recognition, Machine Learning	91.30%	Small dataset, Manual, and Local Features
Amelard et al., 2015	ABCD	DermQuest	Macroscopic	Image processing, Pattern Recognition	Not reported	Small dataset, Diameter
Krishna and Remya Ajai, 2019	3-point checklist	PH2	Dermoscopic	Image Processing, Machine Learning	99.5%	Small dataset, Manual
Kawahara et al., 2019	7-point checklist	Derm7pt	Dermoscopic	Deep Convolutional Neural Networks	89.20%	Single dataset, Difficult to implement
Thanh et al., 2019	ABCD	ISIC	Dermoscopic	Image processing, Pattern Recognition	100.00%	Semi-automatic
Hurtado and Reales, 2021	ABCD	Own dataset	Macroscopic	Machine Learning	86.30%	Limited number of data
Bansal et al., 2022	ABCD	HAM10000 PH2	Dermoscopic	Hand Crafted, Deep Learning	94.90% 98.00%	Not fully automated
Alwakid et al., 2023	Color and texture	HAM10000	Dermoscopic	Deep Learning	91.26%	Single dataset

Table 2.1 shows that different research works followed different features of clinical methods. In these works, there is no evidence found that choice of clinical methods used influences the success rate of Melanoma detection on skin lesion images. Mainly the

effectiveness of detection depends on the chosen computer-based techniques. Additionally, it is found that dermoscopic images are more reliable and effective than macroscopic images (Raikar et al., 2013; Krishna and Remya Ajai, 2019; Kawahara et al., 2019; Thanh et al., 2019; Bansal et al., 2022; Alwakid et al., 2023). The dermoscopic images give us specific size and ratio of lesions with constant focus length whereas macroscopic images do not.

Di Leo et al. (2010) introduced a diagnostic method that defines a set of seven features based on color and texture parameters to describe the malignancy of a lesion. To further enhance the efficacy of this approach, future research should focus on developing suitable algorithms for the detection of Atypical Vascular Patterns and Irregular dots-globules. Raikar et al. (2013) implemented a diagnostic method to automatically detect Melanoma, providing support to clinicians. This method also defined a set of seven features based on color and texture parameters and relies on an image-based measurement system. Abbas et al. (2013) developed an innovative pattern classification method for assisting dermatologists' diagnosis. This method utilized color symmetry and multiscale texture analysis to classify various tumor patterns in a perceptually uniform color space. Amelard et al. (2015) proposed a set of high-level intuitive features to quantitatively describe Melanoma in standard camera images, aiming to model the ABCD criteria commonly used by dermatologists. Krishna & Remya Ajai (2019) devised an efficient feature extraction algorithm for the classification of Melanoma and non-Melanoma skin lesions. Their work involved extracting the ABCD and 3-point checklist features from preprocessed images. A multitask convolutional neural network was presented by Kawahara et al. (2019) and trained on multimodal data, including clinical and dermoscopic images, as well as patient metadata. This network diagnosed skin lesions and categorized the 7-point Melanoma checklist criteria. Their model was robust to missing data at the time of inference because it uses multiple multitask loss functions.

The final model categorized the 7-point checklist, and the diagnosis of skin conditions produced multimodal feature vectors appropriate for image retrieval and localizes clinically discriminant regions. Thanh et al. (2019) proposed a method for the automatic detection of Melanoma skin cancer using image processing techniques. Their approach involved three stages: pre-processing images of skin lesions through adaptive principal curvature, segmenting skin lesions through color normalization, and extracting features based on the ABCD rule. Hurtado and Reales (2021) introduced a novel skin cancer classification system utilizing images from standard cameras, investigating the impact of smoothed bootstrapping for dataset augmentation. Eight classifiers, including KNN, ANN, and SVM, were compared with and without data augmentation, with the augmented ANN demonstrating the highest performance and balance, achieving an AUC of 87.1%, an improvement from 84.3% with the original dataset. Bansal et al. (2022) addressed the labor-intensive process of manual Melanoma detection in dermoscopic images by proposing three hair removal methods and integrating handcrafted and deep learning model features for classification using ResNet50V2 and EfficientNet-B0. The approach achieved significant improvements in the HAM10000 and PH2 datasets, surpassing existing state-of-the-art classifiers and demonstrating promise for enhancing Melanoma detection in clinical settings. Alwakid et al. (2023) proposed a method by leveraging Inception-V3 and InceptionResnet-V2 models for Melanoma detection, fine-tuning feature extraction layers after training new top layers. Utilizing data augmentation to address dataset discrepancies from a public dataset, the proposed models achieved improved performance, with an effectiveness of 0.89% for Inception-V3 and 0.91 for InceptionResnet-V2, showcasing the potential of deep learning in enhancing Melanoma diagnosis.

Overall, these studies present significant advancements in computer-aided Melanoma diagnosis by utilizing various features and techniques. However, further research could

focus on refining these methods and exploring additional features for even more accurate and reliable detection. These works were developed solely to detect Melanoma; however, they do not adequately address the reasoning behind how skin lesions are identified as Melanoma. Various features, such as asymmetry property, pigment network, blue white veil, dots-globules, and different colors of the lesions, should be individually analyzed and classified based on their atypical and typical properties to understand the skin lesion condition. Unfortunately, these aspects are not thoroughly covered in these research works. Additionally, clinically inspired computer-aided Melanoma diagnosis methods have been developed and validated using datasets from specific geographic regions or populations. This raises the concern for evaluating the performance of these methods on multi-ethnic and diverse datasets to ensure their generalizability across different populations and skin types.

2.3   Feature Engineering Based Melanoma Detection

Feature engineering-based means that the computer-based Melanoma detection relies on different dermoscopic features. A considerable amount of existing research (Barın & Güraksın, 2022; Talavera-Martínez et al., 2022; Pathan et al., 2018; Cacciapuoti et al., 2020; Nambisan et al., 2022; Wang et al., 2023) has been dedicated to analyzing various features of skin lesions for diagnosing Melanoma. These features include asymmetry, pigment network, blue white veil, dots-globules, and lesion colors. Researchers have extracted and classified these features into different classes, such as typical (regular), atypical (irregular), present (specific feature is available), and absent (specific feature is not available), in their studies to describe skin condition for better understanding. However, before conducting these feature analyses, an essential step called segmentation is required to identify the region of interest in the skin lesion image. This step is crucial to avoid confusion between healthy and unhealthy (disease-affected) skin areas.

The segmentation process precisely delineates the boundary between healthy and unhealthy skin, allowing for a focused analysis of the lesion. Once the segmentation or isolation process is completed, the subsequent feature analysis becomes more straightforward. Figure 2.2 presents the illustration of the segmentation process followed by the feature analysis process.
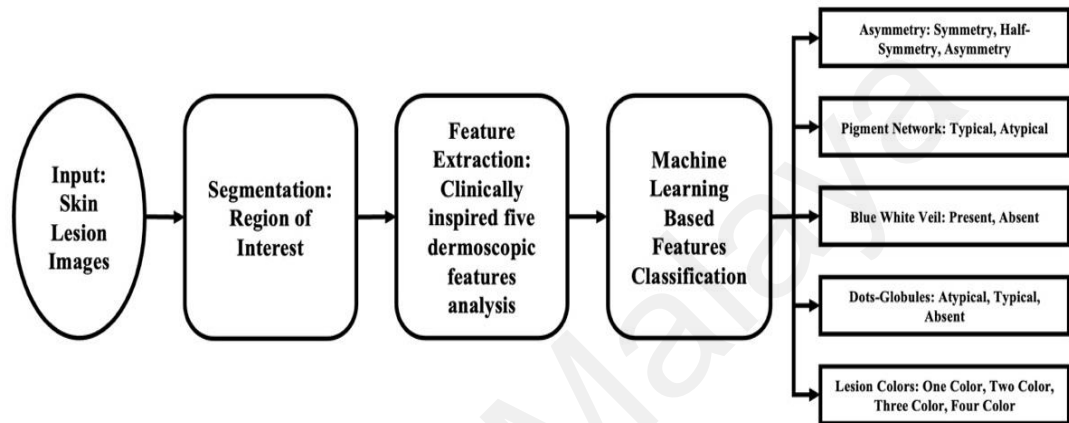
Figure 2.2: Melanoma detection process with explaining skin condition by features analysis.

2.3.1   Skin Lesion Segmentation

A proper lesion segmentation algorithm can limit the region of interest, saving time in the analysis process and enabling more specific extraction of feature information for diagnosing Melanoma from dermoscopic images. The clinical treatment of skin lesions primarily relies on timely detection and accurate delimitation of lesion boundaries to localize cancerous regions effectively. Given the high prevalence of skin cancer, especially aggressive Melanoma with its high metastasis rate, timely diagnosis is critical to initiate treatment before malignancy sets in. In response to these challenges, medical imaging techniques are employed to analyze and segment lesion boundaries from dermoscopic images.

In Figure 2.3, we can observe the different stages of this process. (a) displays the input dermoscopic image, while (b) shows the isolated lesion, representing the disease-affected area. Finally, (c) illustrates the output in the form of a binary image after the segmentation process, utilizing a publicly available dataset (PH2).

(a)   (b)   (c)
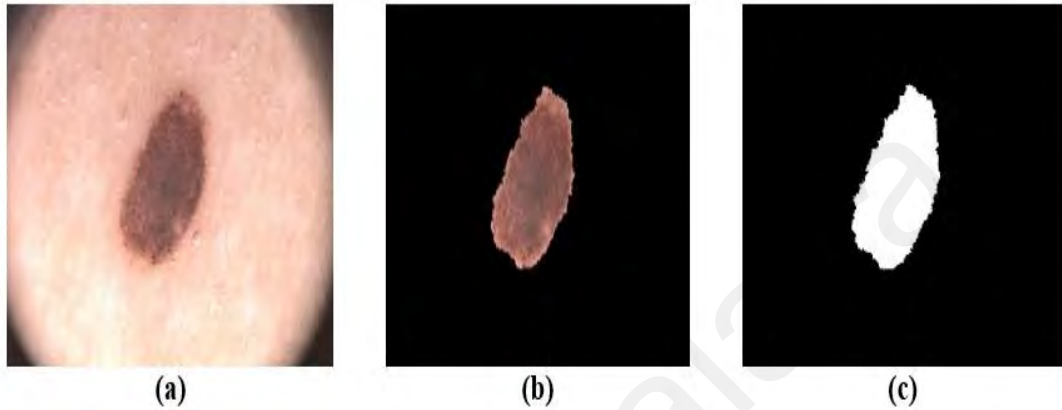
Figure 2.3: The segmentation of disease-affected areas on the dermoscopic image from the PH2 dataset.

In the literature, numerous skin lesion segmentation techniques are available, employing either conventional image processing methods or advanced deep learning approaches, as shown in Table 2.2. In this table, IoU or the Jaccard similarity coefficient (equation 2) is the ratio of correctly classified pixels to the total number of ground truth and predicted pixels in the lesion class.

$$IoU = \frac{True\ Positive}{True\ Positive + False\ Positive + \ False\ Negative} \tag{2}$$

The process of image segmentation plays a crucial role in extracting meaningful features from dermoscopic images of skin lesions. These extracted features, including asymmetry, pigment network, blue white veil, dots-globules, and lesion colors, provide valuable insights into the current skin condition.

Table 2.2: A review of lesion segmentation research works.

Author	Technique	Model	Database	Accuracy	Automatic/ Manual	Limitation/ Future works
Sumithra et al., 2015	Machine Learning	Region Growing (pixel-based)	Own	77.80%	Semi-automatic	Brief and small dataset
Yuan et al., 2017	Deep Learning	Deep Fully Convolutional Neural Networks	ISIC2016 PH2	95.50% 93.80%	Automatic	Complex and small dataset
Goyal et al., 2020	Deep Learning: Ensemble	Mask RCNN, DeepLabV3+	ISIC2017 PH2	94.08% 93.38%	Automatic	Complex, Brief & Small dataset
Adegun & Viriri, 2020	Deep Learning	Deep Convolutional Encoder-Decoder Networks	ISIC2017 PH2	95.00% 95.00%	Automatic	Brief and small dataset
Khan et al., 2021	Deep Learning	Deep Saliency Method: DCNN	ISIC2016 ISIC2017 ISIC2018 PH2	95.38% 95.79% 92.69% 98.70%	Automatic	Complex and time-consuming
Barın & Güraksın, 2022	Deep Learning	Fully Convolutional Neural Networks: AlexNet, ResNet18	ISIC2017 ISIC2018	93.47% 94.65%	Automatic	Complex and small dataset

Sumithra et al. (2015) proposed an automatic segmentation and classification method for skin lesions. Their approach involved filtering skin images to remove unwanted elements and using region-growing segmentation with automatic seed point initialization. Color and texture features were extracted, and SVM and k-NN classifiers were employed, achieving promising results on a dataset of 726 samples. Yuan et al. (2017) presented a fully automatic method for skin lesion segmentation using deep convolutional neural networks. Despite being challenging to implement, their approach demonstrated effective learning with limited training data and introduced a novel loss function based on Jaccard distance to handle class imbalance. Goyal et al. (2020), introduced fully automated deep-learning ensemble methods for accurate lesion boundary segmentation in skin cancer

images. Their ensemble methods outperformed popular segmentation models like FrCN, FCNs, U-Net, and SegNet in terms of sensitivity and specificity. However, the implementation details provided were brief and complex. Adegun & Viriri (2020) proposed a deep learning-based method for automatic Melanoma lesion detection and segmentation. Their approach utilized an enhanced encoder-decoder network with skip pathways and a multi-stage, multi-scale approach. Although their method addressed complex visual characteristics, the experimental results on small datasets were insufficient to demonstrate their effectiveness fully. Khan et al. (2021) addressed the need for automated and accurate methods for multiclass skin lesion segmentation and classification. Their approach involved image enhancement, saliency estimation, and thresholding for segmentation. The use of deep pre-trained CNN models for feature extraction and an improved optimization algorithm was noted. However, the methodology was complex, and training required a longer time. Barın & Güraksın (2022) focused on computer-aided automatic diagnostic systems for diagnosing skin lesions in skin cancer. They proposed a hybrid FCN-ResAlexNet architecture, which outperformed other architectures in accuracy, dice, and Jaccard index metrics when trained with different datasets. However, implementing this deep learning model for lesion segmentation was found to be complex.

Overall, these research works showcase various approaches to skin lesion image segmentation, highlighting the advancements and challenges in accurately detecting and diagnosing skin lesions. Each method has its strengths and limitations, and further research is necessary to refine and improve the segmentation techniques for better clinical application. Most of these research works are based on pre-trained CNN models for lesion detection and segmentation. Using pre-trained networks for skin lesion segmentation may lead to limited generalization on lesion data, domain shift issues due to differences in imaging characteristics, and the risk of overfitting on relatively small dermatological

datasets. Additionally, interpreting the deep and complex pre-trained models can be challenging in diagnosing Melanoma. While the primary focus of these studies was to improve the accuracy of the system, they did not address the system's training and validation time.

Hence, there is a need for a new deep neural network-based semantic segmentation network that provides a detailed understanding of the lesion by labeling each pixel with its corresponding object class. Such a network could facilitate pixel-based classification and segment the lesion more efficiently than existing methods. Moreover, a robust and fully automated segmentation system is essential to ensure consistent and reliable results. Furthermore, researchers should explore different datasets and employ various pre-processing approaches, optimization algorithms, and loss functions to develop a successful skin lesion segmentation system. This comprehensive investigation will help address the challenges faced when segmenting skin images, from non-lesion to lesions, and advance the state-of-the-art in skin lesion segmentation for Melanoma diagnosis. Therefore, this whole segmentation process will be the reflection of the first step of the first research objective of this thesis (see section 1.8, Chapter 1: Introduction).

2.3.2   Dermoscopic Feature Extraction

Melanoma is a dangerous type of skin cancer, and its diagnosis depends heavily on dermatological features. Asymmetry, pigment network, blue white veil, dots-globules, and lesion colors are the five frequent features (based on standard clinically practiced four methods: ABCD, 3-point checklist, 7-point checklist, and CASH) that offer important insights into the nature and development of skin lesions. Asymmetry, or unevenness in shape, is frequently a sign of malignancy. It is possible to distinguish between benign and malignant lesions using the pigment network, a reticular pattern of pigmentation. A blue

white veil that appears as a bluish tint over the lesion implies greater invasion and aggression. Small black spots or globules within the lesion are known to enhance the chance of developing Melanoma. Finally, examining the lesion's various hues can help distinguish between benign and potentially cancerous growths. Dermatologists and other medical professionals can quickly and accurately identify Melanoma by carefully assessing these dermatological features, enabling early intervention and potentially life-saving treatments. The diagnosis process will support the idea of extracting features for optimal classification which is the second step of the first research objective of this thesis (see section 1.8, Chapter 1: Introduction). The five important features are-

2.3.2.1 Lesion Asymmetry

The shape of a lesion on the dermoscopic (deeper skin structures not visible to the naked eye) image carries a clue of skin disease. Lesion shape has a vital role in analyzing dermoscopic structures. In clinically practiced methods, the asymmetric lesion is one of the criteria for diagnosing Melanoma. This is done by measuring asymmetric properties. The asymmetrical lesion shape is a common feature in ABCD, CASH, and 3-point checklist methods. The definition of asymmetry is one half of the lesion or mole or spot that appears, unlike the other half as can be seen in Figure 2.4 (a). An example of a symmetrical lesion is shown in Figure 2.4 (b). Dermatological asymmetric properties depend on lesion color, structure, and border.

Figure 2.4: (a) asymmetrical lesion, and (b) symmetrical lesion from ISIC dataset.

The analysis of dermatological asymmetry has been the subject of investigation in various research works found in the literature. Table 2.3 provides a systematic review of such research, revealing that only a limited number of studies have focused on this feature extraction process. This could be attributed to the perception that the individual lesion shape alone holds little diagnostic value for disease diagnosis. However, it is crucial to recognize that lesion shape plays a pivotal role in diagnosing Melanoma, as it serves as a vital clue for accurate identification. While most research works concentrate on disease diagnosis, the significance of this critical clue should not be overlooked in the quest for improved diagnostic methods. In this table, accuracy (equation 1) is the ratio of correctly classified lesions to the total number of lesions in the asymmetric class, according to the ground truth.

$$Accuracy = \frac{Predicted\ number\ of\ Melanoma}{Total\ number\ of\ Melanoma} \qquad (1)$$

Table 2.3: A review of dermatological lesion asymmetry analysis research works.

Authors	Aim: Asymmetry	Database	Techniques	Accuracy	Gap
Chakravorty et al., 2016	Classification	PH2	Image Processing, Machine Learning	87.00%	Single dataset, Semiautomatic
Milczarski, 2017	Analysis, Classification	PH2	Image Processing	95.80%	Single dataset, Manual
Ali et al., 2020	Analysis, Classification	ISIC2018	Image Processing, Machine Learning	80.00%	Single dataset, Semiautomatic
Damian et al., 2021	Classification	MED-NODE, PH2	Image Processing, Artificial Neural Network	Not mentioned	Limited data, Complex
Zhang and Guo, 2021	Analysis, Classification	Not mentioned	Image Processing, Pattern Recognition	Not mentioned	Supervised
Talavera-Martínez et al., 2022	Classification	SymDerm	Pre-trained CNN, Transfer Learning	64.50%	Low Accuracy

A study by Chakravorty et al. (2016) focused on Melanoma asymmetry as a characteristic of early diagnosis. The authors employed the Kullback-Leibler Divergence (WANG, 2008) of the color histogram and Structural Similarity metric to measure irregularities in color and structure distribution within the lesion area. Several classifiers utilizing these features were evaluated on a dataset, demonstrating improved asymmetry classification compared to existing literature. However, it is important to note that the method used in this experiment was not fully automated, and for evaluation purposes, only one dataset was used, highlighting a research gap.

In a paper by Milczarski (2017), a dermatological asymmetry measure in hue (DASMHue) was introduced and discussed as part of the 3-point checklist for skin lesions. The focus of this paper was on the hue distribution asymmetry of segmented skin lesions, and new dermatological asymmetry measures based on hue distribution were defined. The presented DASMHue measure showed stronger overestimating results but achieved a better total ratio (95.8%) of correctly and overestimated cases compared to

considering shape alone, based on results from the dataset. However, it's worth noting that this experiment didn't involve many datasets or any automation process.

A paper by Ali et al. (2020) addressed the need for an objective computer vision system to aid in the early detection of Melanoma by evaluating asymmetry, color variegation, and diameter. The proposed approach involved training a decision tree on extracted asymmetry measures to predict the asymmetry of new skin lesion images. Suspicious colors for color variegation were derived, and Feret's diameter was used to determine the lesion's diameter. The decision tree achieved an 80% accuracy in determining asymmetry, while the number of suspicious colors and diameter values was objectively identified. However, this percentage rate of accuracy couldn't demonstrate any promising achievement.

A paper by Damian et al. (2021) introduced a feedforward neural network (FFN) with the Levenberg-Marquardt Backpropagation (LMBP) training algorithm to analyze skin lesions. The model utilized different combinations of inputs and desired outputs related to skin lesion types, databases, and asymmetry computation methods. The FFN-LMBP model was validated and tested on 24 images each, focusing on the asymmetry feature extracted using geometric characteristics and histogram projection algorithms. Two datasets were used for skin lesion detection in this study. However, the proposed idea was presented briefly, leaving the implementation somewhat challenging.

Zhang and Guo (2021) introduced an approach that integrates the ABCD rule with clinical characteristics of Melanoma in dermoscopic images to differentiate between benign and malignant melanocytes. Initially, the method employed image processing and pattern recognition techniques to obtain and preprocess dermoscopic images, extracting feature points from the images. Subsequently, both the microscopic and macroscopic

asymmetry of the images were analyzed. Finally, a symmetry score was determined using total dermoscopic score to assess the benign or malignant nature of melanocytes.

A study by Talavera-Martínez et al. (2022) proposed a novel approach using deep learning techniques to classify skin lesion symmetry in dermoscopic images. A CNN model was employed to classify lesions as fully asymmetric, symmetric concerning one axis, or symmetric concerning two axes. The introduction of a new dataset with 615 labeled skin lesions and the evaluation of transfer learning and traditional learning-based methods contributed to the development of a simple, robust, and efficient classification pipeline, outperforming traditional approaches and pre-trained networks with a weighted-average F1-score of 64.5%. However, the proposed algorithm could not demonstrate any significant achievement in the analysis of dermoscopic asymmetry while maintaining a relatively complex algorithm.

These studies focusing on shape-based analysis of asymmetry lesions have made significant contributions. However, as of the current state of research, there has been limited exploration of deep learning methods in this field, and no existing works have offered a detailed analytical explanation for detecting dermatological asymmetric lesions. Additionally, there is a lack of geometry-based lesion shape analysis that would facilitate comprehension for non-experts regarding the properties of these lesions. In this thesis, these gaps are aimed to be addressed by proposing several algorithms that will not only enhance the understanding of dermatological shape properties for experts but also provide accessible explanations to non-experts. By incorporating deep learning techniques and detailed analytical explanations, this study aims to make substantial advancements in the detection and comprehension of dermatological asymmetric lesions.

2.3.2.2 Pigment Network

An irregular fine network or netlike structure made of intersecting brown lines that are known as the pigment network is a skin condition. The pigment network of a lesion helps to spot abnormalities in the dermoscopic structure. The detection of these abnormalities is required to diagnose Melanoma. The unusual pigment network on dermoscopic images is one of the most important criteria in Melanoma diagnosis. Finding the pigment network on a dermoscopic image is difficult due to its thin topological pattern. Distinguishing regular (typical) and irregular (atypical) pigment networks is even more difficult. The nature of grid distribution and color are the primary differences needed to distinguish between the atypical and typical networks. The irregular distribution of an atypical network is usually characterized in color as dark-brown, gray, or black. The regular or typical network has a light-brown structure that is evenly spread. Atypical or dysplastic (abnormal or atypical appearance of a cell or tissue) naevi and certain Melanomas exhibit an atypical network on the lesion. These properties of the atypical network and typical network are an obvious indication for distinguishing melanocytic lesions from non-melanocytic lesions. Figure 2.5 shows a lesion with typical and a lesion with atypical networks taken from a publicly available dataset.
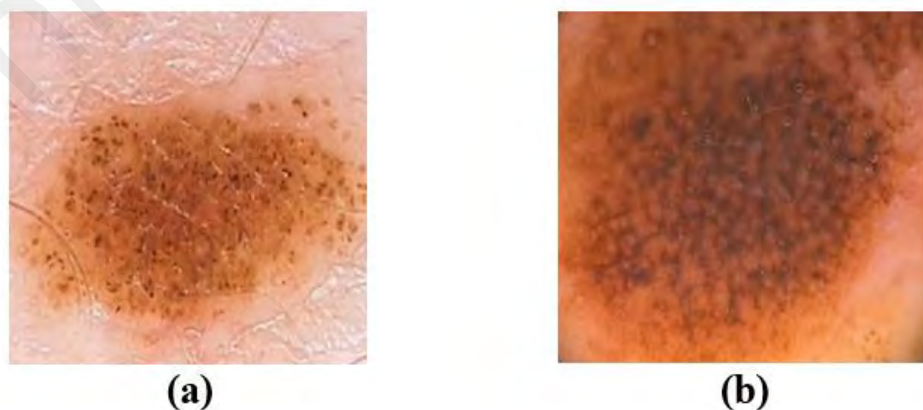
Figure 2.5: (a) is a typical network with an evenly spread light-brown color, and (b) is an atypical network with a dark-brown color from the MED-NODE dataset.

The atypical network is one of the key characteristics used in the 3-point checklist and 7-point checklist methods for identifying Melanoma. Several approaches for studying dermoscopic images and detecting pigment networks have been developed. The color, directionality, and topological features of the pigment networks are used clinically to classify them as an atypical pigment network or typical pigment network. Being the most important feature, the pigment network is investigated in different research works in the literature. Table 2.4 presents a systematic review of the pigment networks of lesions. In the second column of this table, "Detection" means only abnormality of the pigment network was detected, and "Classification" means the pigment network was classified into typical and atypical pigment networks. In this table, accuracy (equation 1) is the ratio of correctly classified lesions to the total number of lesions in the atypical pigment network class, according to the ground truth.

$$Accuracy = \frac{Predicted\ number\ of\ Melanoma}{Total\ number\ of\ Melanoma} \qquad (1)$$

Table 2.4: A review of pigment network detection research works.

Authors	Aim: Pigment Network	Database	Techniques	Accuracy	Gap
Sadeghi et al., 2011	Detection	Atlas	Cyclic graphs	94.30%	Classification
Barata et al., 2012	Detection, Classification	PH2	Bank of Directional Filters, Machine Learning	92.10%	Dataset, Semiautomatic
García Arroyo & García Zapirain, 2014	Detection	Own	Binary Image Processing, Machine Learning	83.64%	Classification
Alfed et al., 2015	Detection	PH2	Image Thresholding, Machine Learning	94.00%	Classification, Small dataset
Eltayef et al., 2017	Detection, Classification	PH2	Bank of Directional Filters, Machine Learning	92.30%	Dataset, Semiautomatic
Pathan et al., 2018	Detection, Classification	PH2	Bank of 2D Gabor Filters, Machine Learning	96.70%	Dataset, Semiautomatic

In this study (Sadeghi et al., 2011), a novel approach was introduced to detect and visualize pigment network structures in dermoscopic images. The method involved image enhancement, edge detection, and conversion of the resulting binary edge image into a graph representation. Cyclic subgraphs corresponding to skin texture structures were identified, filtered to remove non-network structures, and used to create a high-level graph. Classification based on the density ratio of the graph achieved an accuracy of 94.3% on a dataset of 500 images. However, it should be noted that the overall process is time-consuming and requires supervision.

In their paper (Barata et al., 2012), the authors presented an automatic system for detecting pigment networks in dermoscopy images. The system employed pre-processing, directional filters, and connected component analysis to detect the lines of the pigment network. Features extracted from the detected network were used to train an

AdaBoost algorithm for lesion classification. Testing the algorithm on a dataset of 200 medically annotated images achieved a sensitivity of 91.1% and a specificity of 82.1%.

The study by García Arroyo and García Zapirain (2014) introduced a detection algorithm for the pigment network in skin lesion images, which served as an indicator in diagnosing Melanoma. The algorithm consisted of two blocks: a machine learning process to generate rules for constructing a mask of pigment network candidate pixels, and an analysis of structures within this mask to diagnose the presence or absence of the pigment network. Testing the method on a dataset of 220 images resulted in 86% sensitivity and 81.67% specificity, demonstrating the algorithm's reliability.

In the paper by Alfed et al. (2015), an efficient system for skin cancer detection using dermoscopic images was proposed. The system analyzed the statistical characteristics of the pigment network, which proved effective in discriminating against cancer. Evaluation on a dataset of 200 dermoscopic images achieved detection accuracy based on cross-validation.

Eltayef et al. (2017) proposed an automatic system for detecting pigment networks in dermoscopy images. The algorithm consisted of four stages: pre-processing, directional filters, connected component analysis, feature extraction, and classification using a feed-forward neural network. Testing the method on a dataset of 200 dermoscopy images achieved improved results compared to previous studies.

The empirical study by Pathan et al. (2018) aimed to develop an automatic system for detecting pigment networks and differentiating between typical and atypical patterns. The algorithm used Gabor filters to detect pigment network masks and extract features to diagnose lesions based on the pigment network. The proposed system achieved high accuracy and reliability in lesion diagnosis (96% sensitivity, 100% specificity, and 96.7%

accuracy) and pigment network classification (84.6% sensitivity, 88.7% specificity, and 86.7% accuracy) on 200 skin lesion images, which was a significantly low number of images for pigment network classification.

Despite considerable progress in the field of dermoscopy for Melanoma diagnosis, there remain research gaps in the accurate detection and characterization of the pigment network in dermoscopic images. Current methods often struggle to precisely segment and analyze the intricate patterns of the pigment network, leading to potential misdiagnosis and suboptimal treatment decisions. Moreover, there is a lack of standardized methodologies and consensus on defining and quantifying the characteristics of the pigment network. Developing robust and automated algorithms that can accurately detect, segment, and quantify the pigment network will greatly enhance the accuracy and reliability of dermoscopic analysis, ultimately improving Melanoma detection and patient outcomes. Notably, one of the primary gaps is the limited dataset usage, with most of the work focusing only on the PH2 dataset. Additionally, isolated pigment network classification from the lesions is rare among research works. Addressing these issues calls for a new algorithm that is automatic and deep learning-based, integrating image processing and pattern recognition methods for comprehensive explanations.

2.3.2.3 Blue White Veil

The bluish or blue-whitish or blue white veil (BWV) constitutes a significant symptom of Melanoma. It appears as a tiny veil within the skin lesion, exhibiting a mix of gray, blue, and white colors. This veil presents itself as a blue and focal structure-less zone, accompanied by an overlying white "ground-glass" haze. Notably, the BWV occurs exclusively in raised/palpable areas of a lesion. Its presence serves as a strong indicative sign of Melanoma, being visible solely in cases of this condition. Detecting the BWV in

dermoscopic images usually demands a keen eye due to its subtle appearance. Regrettably, during the examination of a lesion, the BWV is often overlooked by observers owing to its mild manifestation. Figure 2.6 illustrates two skin lesions with BWV from a publicly available dataset called MED-NODE.

Figure 2.6: (a) and (b) are the two skin lesions with BWV from the MED-NODE dataset.

This key feature is investigated in different research works from the literature. Limited research has been conducted to detect the BWV on dermoscopic images, likely due to the complexity and challenges in precisely identifying and differentiating it from other structures. Table 2.5 presents a systematic review of the detection and analysis of BWV. In the second column of this table, "Analysis" means measuring the total area of BWV, and "Detection" means just recognizing the BWV's presence. In this table, accuracy (equation 1) is the ratio of correctly classified lesions to the total number of lesions in the BWV class, according to the ground truth.

$$Accuracy = \frac{Predicted\ number\ of\ Melanoma}{Total\ number\ of\ Melanoma} \qquad (1)$$

Table 2.5: A review of lesion BWV detection research works.

Authors	Aim: BWV	Database	Techniques	Accuracy	Gap
Celebi et al., 2008	Analysis, Detection	Atlas	Color Extraction, Decision Tree	82.94%	Single dataset
Madooei and Drew, 2013	Detection	Own	Decision Tree	87.00%	Limited data
Kropidlowski et. al., 2016	Analysis, Detection	Own	Color Analysis Processing	86.50%	Limited data, Manual
Madooei et al., 2019	Detection	Atlas, PH2	Multiple Instance Learning	72.63%, 84.50%	Limited data, Semiautomatic
Cacciapuoti et al., 2020	Detection	Own	Machine Learning	92.00%	Complex

The article (Celebi et al., 2008) presented a machine-learning approach to detect BWV, a significant feature for diagnosing Melanoma in dermoscopy images. The method utilized contextual pixel classification with a decision tree classifier. Testing on a single dataset of 545 dermoscopy images achieved a sensitivity of 69.35% and a specificity of 89.97% for detecting blue white veils, with a higher sensitivity of 78.20% when identifying the blue veil as a primary feature for Melanoma recognition.

Madooei and Drew (2013) proposed two schemas for automatic detection of BWV feature in dermoscopic images. They first proposed a revised threshold-based method with results comparable to state-of-the-art, with much-reduced computation. The second approach, their main contribution, set out an innovative method that attempted to mimic human interpretation of lesion colors. The latter outperformed prior art and moreover introduced a perceptually and semantically meaningful new approach that can serve as a scaffolding for new color investigations in dermoscopic, for example for detection and recognition of common colors under dermoscopic.

This paper (Kropidlowski et al., 2016) addressed the need for image analysis techniques to detect BWV, atypical vascular patterns, and regression structures in

dermatoscopic images for Melanoma assessment. The proposed method utilized color-based image features and showed promising preliminary results, achieving 78% accuracy in detecting BWV, 84% accuracy in detecting atypical vascular patterns, and 86.5% accuracy in detecting regression structures. However, only 84 BWV images were used in this experiment.

In this paper (Madooei et al., 2019), a novel approach to identifying the BWV, an important dermoscopic criterion for diagnosing cutaneous Melanoma, was presented. The method utilized a multiple-instance learning framework and image-level labels to predict the presence or absence of BWV. The approach outperformed state-of-the-art techniques, achieving improved performance in BWV detection and providing a framework for identifying dermoscopic local features from weakly labeled data. The analysis and implementation methods are challenging, even though this research helped advance computerized image processing of skin lesions and has the potential to improve Melanoma detection.

This paper (Cacciapuoti et al., 2020) introduced an image-based measurement and classification system that scores pigmented skin lesions based on the 7-point checklist (BWV is one of seven features) diagnostic method. The system focused on measuring biological structures of only 270 skin lesions associated with atypical nevi and is evaluated for its performance in supporting dermatologists of varying experience levels during clinical activities.

In the realm of dermoscopy for Melanoma detection, a significant research gap exists regarding the accurate and reliable detection of BWV on skin lesion images. These methodologies often struggle to precisely identify and differentiate BWV from other structures, leading to potential misclassification and diagnostic errors. There is a need to use a deep learning model to distinguish the presence and absence of a BWV on the lesion

to provide a more acceptable and accurate classification than conventional methods. Additionally, only a limited number of image data were used to evaluate the different proposed BWV detection algorithms in Table 2.5. It is essential to use multiple sources (different datasets) to demonstrate the adaptability and viability of the detection algorithm from various origins, such as dermoscopy and non-dermoscopy datasets.

2.3.2.4 Dots-Globules

Dots-globules are essential dermoscopic properties and represent a significant feature in the 7-point checklist and CASH methods. Globules, classified as large dots, indicate localized pigment concentrations. Their color reveals their origin, with black dots formed by free melanin in the stratum corneum, brown globules produced by junctional nests of melanocytes, and blue or grey dots attributed to melanophages in the dermis. These distinct characteristics contribute to the differentiation of various lesion types, with some lesions displaying regular dots-globules, while others exhibit irregular patterns. In melanocytic lesions, the pattern and color of dots-globules differ from those in benign lesions, as depicted in Figure 2.7.

(a) in benign lesion   (b) in melanocytic lesion   (c) in non-melanocytic lesion
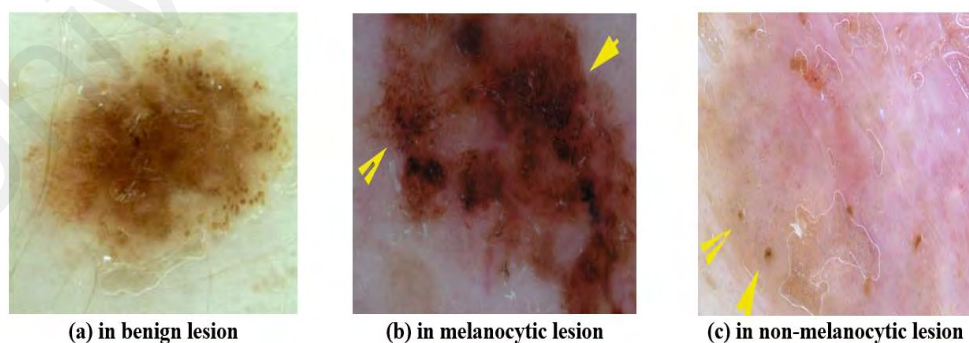
Figure 2.7: Examples of dots-globules in different skin lesions from the ISIC dataset.

Benign lesions commonly exhibit central black dots-globules, whereas Melanoma is characterized by clusters of tiny brown dots. Dots associated with a pigment network may stem from pigment on ridges (superimposed on the network grid) or papillae (in the holes). Classification into three categories - absent, atypical, and typical dots-globules - is based on their pattern. The presence or absence of irregular dots-globules stands as one of the most crucial dermoscopic criteria for diagnosing malignant Melanoma. Researchers have extensively explored this dermatological pattern property in the literature, and Table 2.6 provides a systematic review of dots-globules in dermoscopic images. In the second column of this table, "Analysis" means measuring the dots-globules area on the lesion, and "Detection" means only recognizing or identifying small and large dots. In this table, accuracy (equation 1) is the ratio of correctly classified lesions to the total number of lesions in the atypical dots-globules class, according to the ground truth.

$$Accuracy = \frac{Predicted\ number\ of\ Melanoma}{Total\ number\ of\ Melanoma} \qquad (1)$$

Table 2.6: A review of lesion dots-globules detection research works.

Authors	Aim: Dots-globules	Database	Techniques	Accuracy	Limitation
Jaworek-Korjakowska & Tadeusiewicz, 2013	Analysis, Detection	Own	Histogram Equalization, Otsu algorithm	93.00%	Limited data, Manual
Maglogiannis & Delibasis, 2015	Analysis, Detection	Own	Inverse non-linear diffusion, Machine Learning	76.92%	Single dataset, Semiautomatic
Nowak et al., 2015	Detection	Own	Multi-stage Image Filtering	Not mentioned	Manual, No comparison
Nambisan et al., 2022	Analysis, Detection	Own	Deep Learning	70.00%	Single dataset, Complex

This paper (Jaworek-Korjakowska & Tadeusiewicz, 2013) proposed a novel algorithm that encompasses image enhancement, hair removal, lesion segmentation, and irregular dots-globules detection for pigmented skin lesion assessment. The algorithm demonstrated improved accuracy in detection and classification compared to existing methods, tested on a small dataset of 100 dermoscopic images.

In this work (Maglogiannis & Delibasis, 2015), a methodology for detecting and segmenting dark dots-globules in dermoscopy images was presented. The algorithm employed a multi-resolution approach based on inverse non-linear diffusion for segmentation and extracted features from the segmented dots-globules. While the results showed the effectiveness of the proposed algorithm in segmenting dark dots-globules, the accuracy was significantly low, and the classification of dots-globules was not conducted.

This paper (Nowak et al., 2015) introduced a method for detecting melanin globules in melanocytic skin lesion images. The method involved multi-stage image filtering to extract objects matching the globule structure pattern, followed by classification based on shape and size. However, there was no mention of the accuracy and dataset used in this research.

In this study (Nambisan et al., 2022), the detection of dots-globules in dermoscopic images was explored using deep learning techniques. Two encoder-decoder architectures, UNET and UNET++, were compared, along with different pipeline configurations. The proposed UNET++ technique with checkpoint ensembles achieved high F1 scores for dots-globules detection, outperforming statistical correlation scores among specialists. This method showed potential to improve diagnostic accuracy and provide visible structure detection to better interpret deep learning results, addressing the black-box nature of deep learning in dermatology. However, only one dataset was used in this research, and the accuracy of detection was not high enough.

After analyzing these works, several gaps were identified, like other criteria for Melanoma diagnosis methods. These include the usage of a limited number of skin lesion images, the complexity of color and pattern analysis of dots-globules, and the reliance on supervised analysis processes, suggesting further research in this area. To improve the analysis process, a combination of color analysis and pixel counting algorithms could be explored. Additionally, to achieve a fully automated and time-saving classification process, a deep learning model would prove beneficial.

2.3.2.5 Lesion Colors

The number of colors present in a lesion serves as a crucial criterion for detecting Melanoma, as specified by the 7-point checklist and CASH methods in diagnosing the condition. In these two methods, the presence of multiple colors on the skin lesion indicates malignant Melanoma, with the total number of colors ranging from one to six. These colors include light brown, dark brown, red, yellow, white, and blue/gray, as observed in the PH2 dataset. The total number of colors is further categorized into four classes: one color, two colors, three colors, and four colors (comprising four, five, and six colors). Figure 2.8 showcases various colors observed in dermoscopic images from the ISIC dataset.

(a) One color: Light brown   (b) Two colors: Black and dark brown   (c) Three colors: Dark brown, Light brown, and Blue-gray.

Figure 2.8: The presence of different colors in skin lesion images of the ISIC dataset.

This dermatological color property has been the subject of investigation in various research works from the literature, holding significant importance for computer-aided diagnosis systems. However, limited research on lesion color detection can be attributed to several factors. Firstly, accurately capturing and quantifying skin lesion color consistently and objectively pose challenges due to variations in lighting conditions, imaging techniques, and subjective color perception. Moreover, the complexity of color analysis in the context of skin lesions, often featuring diverse colors and overlapping features, necessitates the integration of multiple clinical factors. Additionally, the lack of access to comprehensive and well-curated datasets of annotated skin lesion images may impede the development and validation of color detection algorithms. These factors collectively contribute to the scarcity of research solely focused on lesion color detection. Table 2.7 presents a systematic review of the dermatological color property research. In the second column of this table, "Analysis" means lesion analysis for other features, "Detection" means identifying different colors, and "Classification" means categorizing lesions based on the total number of colors. In this table, accuracy (equation 1) is the ratio of correctly classified lesions to the total number of lesions in the specific color class, according to the ground truth.

$$Accuracy = \frac{Predicted\ number\ of\ Melanoma}{Total\ number\ of\ Melanoma} \qquad (1)$$

Table 2.7: A review of lesion color detection research works.

Authors	Aim: Color	Database	Techniques	Accuracy	Gap
Stanley et al., 2007	Analysis, Detection	Atlas	Color Histogram, Machine Learning	87.70%	Limited data, No classification, Semiautomatic
Nezhadian, & Rashidi, 2017	Analysis Detection	ISIC	Active Counter Model, Machine Learning	97.00%	Limited data, No classification, Semiautomatic
Almubarak et al., 2017	Analysis, Detection	Atlas	Fuzzy Logic-based Color Histogram, Machine Learning	92.60%	No classification, Semiautomatic
Sabbaghi Mahmouei et al., 2019	Analysis, Detection, Classification	Own	Color Enhancement, QuadTree Segmentation	81.00%	Less number of colors, Semiautomatic

A research study (Stanley et al., 2007) explored the use of relative color histogram analysis to discriminate skin lesions based on color features in dermoscopy images. The study examined different training set sizes and regions of the skin lesion for feature calculations. Experimental results demonstrated improved discrimination capability when focusing on the interior lesion region, achieving high recognition rates for malignant Melanoma and dysplastic nevi. While the presented techniques, including relative color analysis and the determination of benign and malignant regions, hold potential applications for any set of benign and malignant lesion images, this research did not perform classification for the number of lesion colors.

An algorithm (Nezhadian & Rashidi, 2017) was introduced for classifying dermoscopic images into malignant and benign skin lesions, with a specific focus on Melanoma diagnosis. The algorithm utilized image segmentation, texture-based features, and color components for classification. Experimental results on the international skin imaging collaboration dataset demonstrated a high accuracy of 97% using a support vector machine classifier. However, like the previous research, this study also did not

consider the number of lesion colors, which is an important feature for diagnosing Melanoma.

A fuzzy logic-based color histogram analysis technique (Almubarak et al., 2017) was proposed for distinguishing between benign and malignant skin lesions in dermoscopy images. The method utilized fuzzy sets and a fuzzy clustering ratio to quantify the skin lesion color feature. Experimental results on a dataset of 517 dermoscopy images demonstrated that the proposed approach achieved a recognition rate of 92.6% for Melanomas, with a false-positive rate of approximately 13.5%. The results highlighted the significance of colors in the periphery of the lesion and the relevance of the fuzzy logic-based description for the clinical assessment of malignant Melanoma. However, like the previous studies, color classification of the dermoscopic lesions was not conducted.

A QuadTree-based Melanoma detection system (Sabbaghi Mahmouei et al., 2019) was introduced in this paper, inspired by dermatologists' color perception. The system addressed challenges in clinical color assessment in dermoscopy images by employing color enhancement and automatic color identification techniques based on QuadTree segmentation. The proposed approach accurately modeled expert color assessment, identifying more colors in Melanomas compared to benign skin lesions and delineating the locations of Melanoma colors. Despite these advancements, this research also did not classify or identify the number of colors in the skin lesions.

Although these research works detected the RGB color of lesions and analyzed them to diagnose malignant Melanoma, they were individually applied to a single dataset, and the process of color classification (such as light brown, dark brown, dark, red, orange, and white) was absent. Dermoscopy color classification in different colors is challenging due to several reasons: subjectivity and variability, complex color patterns, overlapping

color features, atypical colors, and technical limitations. Additionally, most of the detection methods are supervised learning techniques (without deep learning). Utilizing deep learning-based models can potentially provide a solution for this non-automatic process, enabling automatic color detection and classification.

Finally, the Feature Engineering based Melanoma detection process incorporates a comprehensive analysis of dermoscopic features, including asymmetry, pigment network, blue white veil, dots-globules, and lesion colors. The combined utilization of image processing and deep learning techniques allows for a robust and accurate assessment, offering promising advancements in the field of Melanoma diagnosis. By leveraging these state-of-the-art methodologies, early detection is aimed to enhance and provide valuable insights for improved patient care and management.

2.4   Classification Based Melanoma Detection

Classification or deep learning-based early Melanoma detection on dermoscopic images holds significant importance in the fields of dermatology and healthcare. Melanoma, a deadly form of skin cancer, has experienced a global increase in incidence, underscoring the criticality of early detection for successful treatment and improved patient outcomes. While dermoscopic images offer valuable insights into skin lesions, their accurate diagnosis often necessitates specialized expertise, which may not be universally accessible in all healthcare facilities. By harnessing the power of deep learning algorithms, these systems can autonomously analyze and classify dermoscopic images, enabling the early identification of suspicious lesions, including those that are subtle and challenging to detect. This technology has the potential to democratize healthcare, extending accurate Melanoma screening to a broader population and thereby reducing healthcare disparities. Moreover, the speed and efficiency of these automated

systems can assist dermatologists in making prompt and well-informed decisions, leading to timely interventions and potentially saving lives. As research and technology continue to advance, deep learning-based early Melanoma detection on dermoscopic images emerges as a highly promising tool in the ongoing fight against skin cancer, paving the way for more efficient and effective Melanoma management. Some of the most recent deep learning-based Melanoma detection research works are presented in Table 2.8. In this table, accuracy (equation 1) is the ratio of correctly classified lesions to the total number of lesions in the Melanoma class, according to the ground truth.

$$Accuracy = \frac{Predicted\ number\ of\ Melanoma}{Total\ number\ of\ Melanoma} \qquad (1)$$

Table 2.8: A review of deep learning-based Melanoma detection research works.

Authors	Aim	Database	Tech- niques	Accuracy	Model	Limitation / Future Work
Ma & Yin, 2018	Melanoma Diagnosis	ISIC2017	Deep Learning	79.00%	Deep attention network	Briefly described and color-based only
Xiao & Wu, 2020	Multiclass Lesion Classification	ISIC2017	Deep Learning	85.00%	Visual saliency by DenseNet121	Semiautomatic and Single dataset
Hosny et al., 2020	Multiclass Lesion Classification	ISIC2018	Transfer Learning	98.70%	AlexNet	Dataset and Explanation
Kassem et al., 2020	Multiclass Lesion Classification	ISIC2019	Deep Learning	94.92%	CNN	Dataset and Explanation
Gessert et al., 2020	Multiclass Lesion Classification	HAM10000 ISIC2018 Atlas	Deep Learning	84.00% 83.90% 83.70%	CNN (Patch- Based Attention)	Complexity and Imbalance Data
Zhao et al., 2021	Melanoma Diagnosis	ISIC 2018 ISIC 2019	Deep Learning	93.64%	StyleGAN + DenseNet201	Complexity and Explanation
Sharma et al., 2022	Melanoma Diagnosis	HAM10000	Deep Learning	98.30%	CNN	Class Number and Explanation
Ayas, 2022	Melanoma Diagnosis	ISIC2019	Swin Trans- former	97.20%	Transformer + CNN	Dataset and Explanation
Anand et al., 2023	Lesion Classification	HAM10000	Deep Learning	97.96%	U-Net + CNN	Single Dataset and two steps
Wang et al., 2023	Lesion Classification	ISIC2017 ISIC2018	Deep Learning	92.20%	Efficient- NetB4	Global features based

The paper by Ma & Yin (2018) introduced a deep attention network designed for classifying dermoscopy images in Melanoma detection. This network effectively combined the attention mechanism and the Fisher criterion to optimize its performance. Through extensive experimentation, the results showcased that this innovative neural network significantly improved recognition rates, particularly in cases where the sample size was limited. Furthermore, the study investigated the incorporation of color constancy algorithms for color calibration, revealing their potential to further enhance the model's

recognition rate. However, it is essential to note that the proposed model was described briefly, and the evaluation was conducted using only one dataset. To establish the robustness and reliability of this deep attention network for Melanoma detection, future research should involve the validation of multiple datasets.

Xiao & Wu (2020) addressed the necessity for integrated frameworks that combine traditional and deep learning features for skin cancer classification. They proposed a novel model that effectively merged traditional local binary pattern features with deep convolutional features, enabling the extraction of representative global-local features. To enhance the performance of the Global-DNN and Local-DNN, the authors applied fusion strategies. Additionally, they employed a target segmentation method based on visual saliency detection to eliminate background interference. Although the experimental results were obtained from a single dataset, they demonstrated promising potential for skin cancer classification. However, further validation on multiple datasets is essential to establish the generalizability and robustness of the proposed model.

Hosny et al. (2020) presented a method for skin lesion classification to detect Melanoma and other types of skin lesions. Their proposed approach utilized transfer learning by initializing the weights of the last three replaced layers in a pre-trained AlexNet. When tested on a dataset, the method achieved remarkable accuracy (98.70%), sensitivity (95.60%), specificity (99.27%), and precision (95.06%) in classifying the skin lesions into seven different classes. The research highlighted the significance of multiple datasets during the training period to improve the model's performance and generalization capabilities.

The paper by Kassem et al. (2020) addressed the classification of skin lesions, particularly focusing on Melanoma detection to facilitate early diagnosis and potentially save lives. The authors proposed a model that utilized transfer learning with a pre-trained

GoogleNet model, fine-tuning the parameters through training. The model was rigorously tested on a dataset, exhibiting high accuracy, sensitivity, specificity, and precision percentages for classifying eight different classes of skin lesions. Remarkably, the model also effectively identified images that did not belong to any specific class and categorized them as unknown images. However, it is important to note that this model, too, was trained on a single dataset, which may limit the model's overall robustness and generalizability. Further research that includes diverse datasets would be crucial to validate and enhance the model's performance in real-world scenarios.

Gessert et al. (2020) tackled two significant challenges in skin lesion classification: effectively using high-resolution images and addressing the class imbalance. The proposed approach introduced a patch-based attention architecture to handle high-resolution images without downsampling, resulting in a remarkable 7% improvement in sensitivity. Moreover, the authors employed class balancing techniques and a novel diagnosis-guided loss weighting method, which contributed to an additional 3% sensitivity improvement over traditional methods. Notably, the implications of these methods extended beyond skin lesion classification, as they could be adapted for other clinical applications facing similar challenges with high-resolution images and class imbalance. However, the lack of detailed explanations might make implementing this proposed model complex.

Zhao et al. (2021) confronted the challenges of skin lesion classification using deep learning techniques, particularly the limited availability of labeled images and imbalanced datasets. To overcome these obstacles, their proposed approach harnessed a skin lesion augmentation style-based GAN in conjunction with DenseNet201. The framework featured a redesigned generator and discriminator, along with a new loss function that significantly improved multiclass accuracy. Experimental results on

multiple datasets showcased the effectiveness of the proposed method, achieving an impressive balanced multiclass accuracy of 93.64%. Nonetheless, the complexity of understanding and reproducing this research, coupled with the limited explanation, may pose challenges for researchers trying to implement the model.

Sharma et al. (2022) presented a cascaded ensembled network that skillfully combined CNN with handcrafted features to enhance the efficiency of skin cancer classification. By integrating non-handcrafted image features extracted by CNN with handcrafted features like color moments and texture features, the ensembled deep learning model achieved a substantial accuracy improvement, reaching 98.3%, compared to the CNN model's accuracy of 85.3%. However, a limitation of this algorithm lies in its reliance on a single dataset, and a more detailed description of the model implementation would be beneficial for researchers looking to replicate the study.

Ayas (2022) proposed a novel approach combining transformer and CNNs, termed Swin Transformer, for multiclass skin lesion classification. This approach leverages end-to-end mapping and did not require prior knowledge while also addressing class imbalance through weighted cross-entropy loss. Evaluation on the dermoscopic dataset demonstrated superior performance, achieving high sensitivity, specificity, accuracy, and balanced accuracy compared to existing state-of-the-art methods.

Anand et al. (2023) presented a fusion model integrating the U-Net and CNN models. The U-Net model was employed for disease segmentation in skin images, while the CNN model was proposed for multi-class classification of the segmented images. The model was simulated and analyzed using dermoscopic images across seven different classifications of skin diseases. Evaluation of this model was conducted using two optimizers, namely Adam and Adadelta, over 20 epochs and with a batch size of 32 for

skin disease classification. Remarkably, the model performed exceptionally well with the Adadelta optimizer.

Wang et al. (2023) proposed a deep learning method aimed at enhancing both intra-class consistency and inter-class discrimination in automatic skin lesion classification. The method introduced a CAM-based global-lesion localization module to improve inter-class discriminative feature learning and proposed a global features-guided intra-class similarity learning module to enhance intra-class feature concentration. By collaborating with these techniques, the method effectively improved classification performance. Evaluation on two dermoscopic datasets demonstrated the method's generalizability and adaptive focus on discriminative lesion regions, as shown through experiments with various backbones.

Despite remarkable advancements in deep learning-based Melanoma diagnosis, several research gaps persist in prominent works. Various methods have developed powerful CNN models, but unfortunately, these models are complex to implement. The research works mentioned in Table 2.8 are all complex to implement and train on different dermoscopic datasets. The literature lacks hints about the selection criteria for different parameters (activation function, stride, masks, learning rate, number of epochs, etc.) and the minimum number of training datasets required for these CNN models. Selecting the right parameters in different layers is crucial as the effectiveness of the CNN model's training depends on it. Moreover, what factors contribute to their success in achieving greater accuracy and efficiency remains unclear. Furthermore, most available approaches perform well on specific datasets but struggle with different origins of datasets. These concerns cause other challenges, such as dealing with the minimum number of training samples, particularly in Melanoma.

An in-depth review reveals additional questions that warrant investigation, such as how to design the best CNN for skin lesion classification, which parameters yield the best performance, what the minimum number of training images required for significant results, and how to improve the CNN architecture and parameters to make it suitable for different origins of datasets (dermoscopy, non-dermoscopy, and microscopy). Therefore, the analytical research goal is to design a CNN model using the best-performing parameters in different layers and with a minimum number of training images to achieve higher accuracy and time efficiency on multiple types of datasets. This goal will support the evaluation process against current conventional methods, which aligns with the second research objective of this thesis (see section 1.8, Chapter 1: Introduction).

In conclusion, given the remarkable advancements in Artificial Intelligence and the potential for data to evolve and come from diverse sources, the deep learning method holds significant promise for Melanoma detection. Its ability to handle complex datasets and learn from various origins makes it an attractive avenue for further investigation. Leveraging the advantages of the deep learning approach in Melanoma detection can lead to more accurate and efficient models, ultimately improving early diagnosis and patient outcomes in the fight against skin cancer.

2.5   Chapter Summary

This literature review discusses various computer-aided Melanoma detection methods in terms of their effectiveness, usefulness, and accuracy. A substantial amount of research has been conducted to diagnose Melanoma using dermoscopic images. However, upon analyzing clinically inspired, Feature Engineering-based, and Classification-based methods, a list of research gaps has been identified, offering opportunities to advance Melanoma detection research. To address these research gaps, a proposed method should

be introduced which will be divided into two phases: the Feature Engineering phase and the Classification phase. The Feature Engineering phase will encompass lesion segmentation and feature extraction processes to provide explanations about the skin condition, aiding both experts and non-experts in understanding the diagnosis. On the other hand, the Classification phase will focus on utilizing only color patches (lesion color and texture) to enable instant Melanoma diagnosis, facilitating remote control treatment and mass screening tests. These two phases will be implemented and detailed in Chapter 4, Chapter 5, and Chapter 6 of this thesis. The research plan for detecting Melanoma is outlined in Chapter 3: Research Methodology. By following this structured approach, the proposed method aims to bridge the existing research gaps and contribute to the field of Melanoma diagnosis.

CHAPTER 3: RESEARCH METHODOLOGY

3.1   Background

The preservation of dermoscopic image quality poses a significant challenge due to various artifacts, including hair, gel, and light reflection. These artifacts make it difficult to segment the skin lesion (region of interest) and extract information from that lesion. Additionally, issues such as color constancy arise when dermoscopic images are obtained from different sources, leading to changes in image quality due to variations in light intensity and pixel resolution. These differences in image quality can hinder the accurate diagnosis of Melanoma through dermoscopy, affecting resolution, clarity, distortion, artifacts, color accuracy, artifact suppression, and image consistency. Extracting disease-specific information from dermoscopic images is also challenging, as it often requires suitable computer-aided programs and algorithms for image processing and pattern recognition. Lastly, the manual or semi-automatic nature of the current dermoscopic image analysis process necessitates further research to achieve full automation using machine learning techniques like deep learning and transfer learning. This chapter presents the research phase, research pipeline, research objectives mapping of the proposed methods, and dataset that will be adopted in the proposed work. The research phase describes the multiple steps of the proposed methods, and the research pipeline shows how those steps of the proposed methods are achieved along the possible outcomes.

3.2   Proposed Method

This section discusses the outline of the thesis, which is based on the research gaps identified in sections 2.2, 2.3, and 2.4. The existing clinically inspired Melanoma detection methods are limited in providing proper explanation for diagnosing Melanoma. To address this gap, feature engineering-based Melanoma detection methods are

considered. However, these methods also limited in segmenting lesion and extracting feature from dermoscopic images. Resolving lesion segmentation challenges is crucial for accurately identifying and delineating abnormal skin lesions, which is essential for diagnosis, treatment planning, and monitoring disease progression. Overcoming feature extraction challenges is important for effectively capturing and utilizing relevant patterns and characteristics from the skin lesion images to improve the early Melanoma detection method. To fill these research gaps, the first research objective is proposed, focusing on improving lesion segmentation and feature extraction. Achieving this objective will help to detect Melanoma by following clinical rules-based method.

Additionally, there is a need for a diagnosis technique capable of providing instant diagnostic results for remote control treatment and mass screening tests. To address this need, the second research objective is introduced.

Finally, there is a requirement of validation of decision-making process for early Melanoma detection approaches on multiple datasets. In third research objective, feature engineering-based approach and instant diagnosis technique are evaluated against clinically approved ground truth and compared with state-of-the-arts.

Upon determining the three research objectives, the proposed methods are expected to address most of the research gaps concerning clinically inspired approaches, segmentation, features, and classification of early Melanoma on dermoscopic images. A comparative analysis between the proposed methods and existing methodologies (based on Table 2.1 to Table 2.8 from Chapter 2) for different focused areas is presented in Table 3.1. Since first and second research objectives propose two individual methods to detect early Melanoma and third research objective present the evaluation of these two methods against state-of-the-arts, this research is divided into two phases: Phase 1 is combined

with research objectives first and third, and Phase 2 is combined with research objectives second and third.

Table 3.1: A comparison mapping of research gaps between existing and proposed methods.

Focused Area (Based on)	Feature Engineering Based Detection						Classification Based Detection
	Segmen-tation	Asymmetry	Pigment Networks	BWV	Dots-globules	Lesion Colors	Deep Learning: Color Patch
Clinical Approach (**Table 2.1**)	x	√	x	x	x	√	x
Segmentation (**Table 2.2**)	√	x	x	x	x	x	x
Asymmetry (**Table 2.3**)	x	√	x	x	x	x	x
Pigment Network (**Table 2.4**)	x	x	√	x	x	x	x
Blue white veil (**Table 2.5**)	x	x	x	√	x	x	x
Dots and globules (**Table 2.6**)	x	x	x	x	√	x	x
Colors (**Table 2.7**)	x	x	x	x	x	√	x
Classification (**Table 2.8**)	√	x	x	x	x	x	√
Proposed Method	√	√	√	√	√	√	√
	Ob 01 + Ob 03						**Ob 02 + Ob 03**

Based on this analysis, the research plan is divided into three chapters such as Chapter 4: Skin Lesion Segmentation, Chapter 5: Dermatological Feature Analysis, and Chapter 6: Classification Based Melanoma Detection. These chapters will collectively address the identified research gaps and contribute to the advancement of Melanoma detection methods.

The proposed method is divided into two phases which are the Feature Engineering phase and the Classification phase. The feature engineering phase will be the combination of segmentation and feature extraction by following clinically practiced methods' features which are based on image processing, computer vision, pattern recognition techniques, and deep learning algorithms. This feature engineering phase should cover the research gap of the clinically inspired approaches, segmentation, asymmetry, pigment network, blue white veil, dots-globules, and lesion colors (see Table 3.1). The Classification phase will be only the deep learning-based computer-aided diagnosis system that will be able to detect Melanoma instantly based on the skin lesion's color and texture (color patch). This will cover the research gap of the Classification (see Table 3.1) which will be useful for mass screening tests. The overall proposed framework's graphical presentation is shown in Figure 3.1. Details of this proposed framework are available in 3.3 and 3.4 sections.
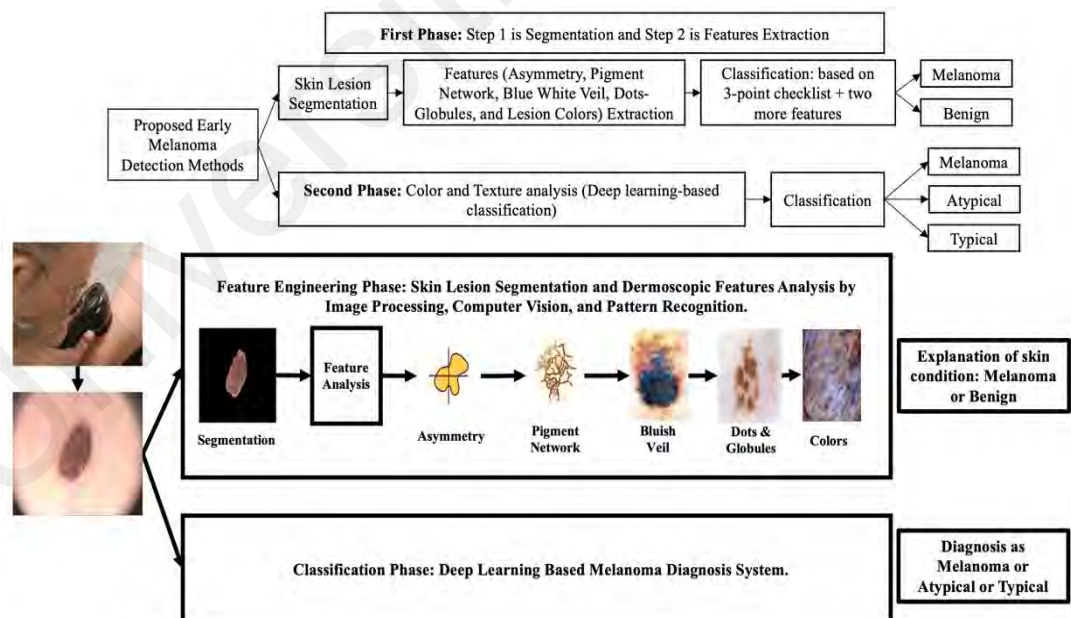
Figure 3.1: The overall proposed method for early diagnosing Melanoma.

3.3   Research Phases

In Chapter 2: Literature Review, Melanoma detection-related existing works were analyzed to identify the research gap and problem statements. Based on the findings, the research aims and objectives are determined. As a result, the dermoscopic image analysis is the main research focus point identified in this thesis. To detect Melanoma on the dermoscopic images, two distinct methods are proposed in this thesis. The first method is in Phase 1 which is Feature Engineering-based, and the second method is Phase 2 which is Classification-based methods. Those two phases are presented in Figure 3.2.
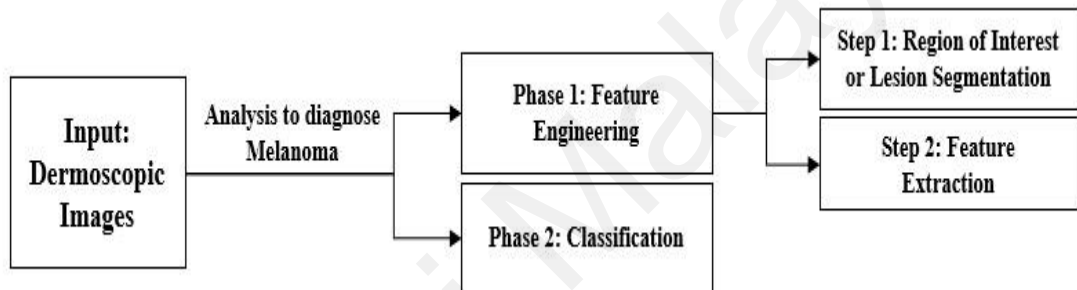
Figure 3.2: Research phases of the proposed work for detecting Melanoma.

3.3.1   Phase 1: Feature Engineering

Feature engineering-based melanoma detection is crucial as it offers a transparent and interpretable approach to identifying Melanoma cases. By using dermoscopic features and diagnostic criteria, clinicians can justify their decisions, build trust, and enhance diagnostic accuracy, fostering continuous learning and improvement. Moreover, such systems aid less-experienced clinicians, reducing errors, and leading to earlier detection for improved patient outcomes.

In this phase, the research focuses on dermoscopic images as the primary input. To extract useful information from these images, a pre-processing step is necessary, which includes image segmentation. The purpose of image segmentation is to identify and

differentiate the disease-infected areas from the non-disease-infected areas. The identified disease-infected areas are then used as input for the feature extraction process.

However, it is important to note that not all features require segmentation, certain characteristics and patterns within the lesion can only be effectively identified and analyzed once the lesion has been accurately delineated from the surrounding healthy skin. For the complex features that depend on the specific boundaries of the lesion, like lesion asymmetry, irregular borders, or internal structures, accurate segmentation is essential. For some features, such as overall shape, size, and color distribution, segmentation may not be necessary as they can be extracted directly from the whole lesion (including healthy and unhealthy regions), and a comparison between healthy and unhealthy regions on the skin is required. Since lesion asymmetry analysis depends on lesion shape and border irregularity, the segmentation process is required in this proposed method. On the other hand, color analysis of dermoscopic images does not depend on segmented images. The segmented image solely represents the color of the lesion (disease-infected area) and does not provide information about the color of healthy (non-disease-infected) skin. Consequently, Phase 1 of this research involves two distinct steps.

Step 1 (Lesion Segmentation): The region of interest is segmented on the input dermoscopic image in this step of Phase 1. Using semiautomatic (involves user interaction, where the algorithm requires some initial input or guidance from the user before completing the segmentation process) and automatic (does not require user intervention, and the entire segmentation process is carried out automatically without any human input) algorithms, the disease-infected area becomes isolated. The isolated image is known as a segmented image. Figure 3.3 demonstrates the segmentation of skin lesions from a dermoscopic image.
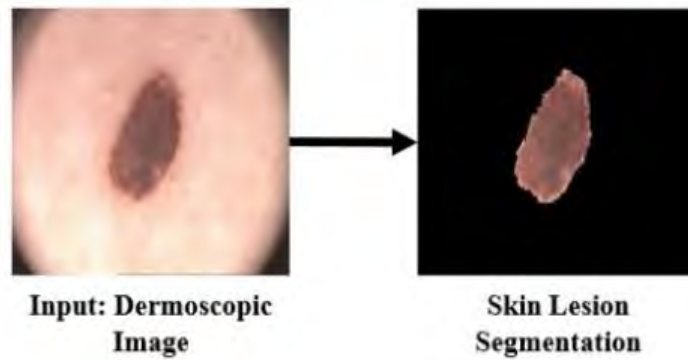
Figure 3.3: An example of a skin lesion segmentation.

The segmented image makes a distinct line between disease and non-disease areas. This distinct line helps to do further processes such as asymmetry analysis on dermoscopic images for Melanoma diagnosis. Lesion segmentation in dermoscopic images is selectively required for extracting specific features that heavily rely on the precise delineation of the lesion boundaries, enabling more accurate and meaningful analysis for Melanoma diagnosis and classification. Four distinct multiracial datasets such as PH2 (Mendonca et al., 2013), ISIC2016 (Gutman et al., 2016), ISIC2017 (Codella et al., 2017), and HAM10000 (Tschandl et al., 2018) are employed in this step. These datasets provide ground truth about the region of interest (disease-infected area) where other datasets do not. The segmentation process of a skin lesion generates a binary mask that outlines the boundaries of the lesion, effectively separating it from the surrounding healthy skin. This output aids in further analysis, classification, and quantification of the lesion's characteristics for accurate diagnosis and treatment planning.

Step 2 (Feature Extraction): In this step of Phase 1, Melanoma-diagnosing features are extracted from both dermoscopic images and segmented (output of Step 1) dermoscopic images. Dermatological asymmetry (lesion segmentation is required), pigment networks,

blue white veil, dots-globules, and lesion colors are detected and classified based on their topological properties. Figure 3.4 presents different dermatological features.
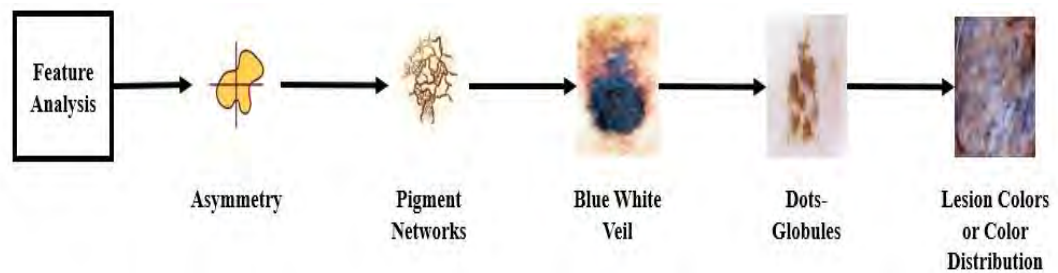
Figure 3.4: Different dermatological features that are required to diagnose Melanoma.

Due to the various characteristics of the different dermoscopic features, several image processing, shape analysis, and pattern recognition algorithms are applied to those dermoscopic images for analysis, detection, and extraction processes. For the classification process, different deep learning models are used on those extracted features. Then, individually relevant feature classification results are compared with the existing works from the literature. Analysis of these features for the Melanoma diagnosis is presented in a separate chapter (i.e., Chapter 5: Dermatological Feature Analysis). To show the flexibility of the proposed algorithms, multiple datasets (see Section 3.4: The Source of Data) are utilized in the analytical experiments.

3.3.2   Phase 2: Classification

Classification-based Melanoma detection on dermoscopic images is essential as it offers automated and efficient diagnosis without manual intervention. These algorithms excel at capturing complex patterns, ensuring accurate and objective assessments, enabling large-scale screening for early detection, and optimizing healthcare resources.

The method is necessary to save time, avoid clinical feature-based algorithms, and provide rapid results for diagnosing Melanoma on dermoscopic images without the need for manual handcrafted features. It primarily focuses on analyzing intricate patterns, textures, shapes, and other discriminative information present in the lesion images. Therefore, a deep learning-based computer program has been trained using various datasets to provide immediate results for diagnosing Melanoma on dermoscopic images, serving as a general screening tool.

The process involves analyzing small color patches containing intricate patterns, textures, shapes, and other discriminative information about the skin lesion. The entire input image is divided into small color patches, aiding the deep-learning models in making decisions. The output of this Melanoma detection algorithm is a classification-based result without any accompanying explanation. To obtain an explanation for the decisions made by this Classification-based Melanoma detection method, Phase 1 will be useful.

3.4   Research Pipeline

The mapping of each research objective with contribution is presented in Table 3.1, where research methods with different algorithms and expected outcomes are briefly described.

Table 3.2: The contribution of these research works based on each objective.

Objectives		Methods/Techniques	Contribution	Outcomes
Ob 01. To perform feature engineering for five dermoscopic features from dermoscopic images based on lesion segmentation and feature-specific deep learning models. **Ob 02.** To experiment with diverse CNN models for automatic computer-based image classification for Melanoma detection from dermoscopic images. **Ob 03.** To evaluate feature-engineered and automated feature-identified CNN methods against state-of-the-art techniques in determining the efficacy of Melanoma detection.	**Ob 01 + Ob 03**	1. Segmentation: Semantic Segmentation. Classification: Binary Classifier (Region of interest and background).	A novel semantic segmentation network to segment regions of interest from dermoscopic images.	Lesion Segmentation: The ROI (lesion) is isolated from the healthier area.
		2. Asymmetry: Analysis: Geometrical shape. Classification: Binary Classifier	New analytical approach by using different image and pattern recognition techniques and algorithms for the dermatological structures of dermoscopic images.	Symmetry Classification: Asymmetry or symmetry.
		3. Pigment Network: Analysis: Directional imaging filter to detect and isolate the pigment network. Classification: Binary Classifier		Pigment Network Classification: Atypical and typical network.
		4. Blue White Veil (BWV) Analysis: Histogram analysis, and pixel count. Classification: Binary Classifier.		BWV Classification: Present or Absent.
		5. Dots-Globules Analysis: Color and shape. Classification: Multi-class classifier.		Dots-Globules Classification: Atypical, Present, and Absent.
		6. Number of Colors Analysis: RGB, HSV, and L*a*b* color spaces. Classification: Multi-class classifier.		Color Classification: White, Black, Blue, Red, and Brown.
	Ob 02 + Ob 03	7. Classification Based Feature Extraction: Convolutional Neural Network Classification: Convolutional Neural Network	A novel Convolutional Neural Network model for the local features of dermoscopic images.	Deep Learning Classification: Melanoma, Atypical Nevus, and Common Nevus.

To achieve the stated research objectives, two Melanoma detection methods are proposed in two phases- Phase 1: Feature Engineering, and Phase 2: Classification. In this thesis, the first step (lesion segmentation) and second step (features extraction) of

Phase 1 represents research objectives 01 and 03. The lesion segmentation helps to detect the disease area and non-disease area on the dermoscopic image. For this process, the semantic segmentation method is used (see Chapter 4: Skin Lesion Segmentation). The second step of Phase 1 represents the five dermoscopic features (asymmetry, pigment network, blue white veil, dots-globules, and lesion colors) extraction process where different algorithms are used including directional imaging, pattern recognition, and deep learning algorithms. For the dermoscopic asymmetry analysis, the segmented image from the first step of this phase is used as input. For the rest of the four dermoscopic features (pigment network, blue white veil, dots-globules, and lesion color), directly a dermoscopic image (without segmentation) is used as input. All five features are analyzed individually (see Chapter 5: Dermatological Feature Analysis). Finally, the second phase (research objectives 02 and 03) represents the Classification-based Melanoma detection where a convolution neural network is designed and trained to extract color patches including intricate patterns, textures, shapes, and other discriminative information automatically and classify the dermoscopic images as either Melanoma or Benign (see Chapter 6: Classification Based Melanoma Detection). The relation between two phases is- first phase is the explanation of second phase where first phase explains skin condition according to clinically practiced methods and second phase provides instance detection result. The architecture of the proposed work (including two phases and three research objectives) is shown in Figure 3.5.
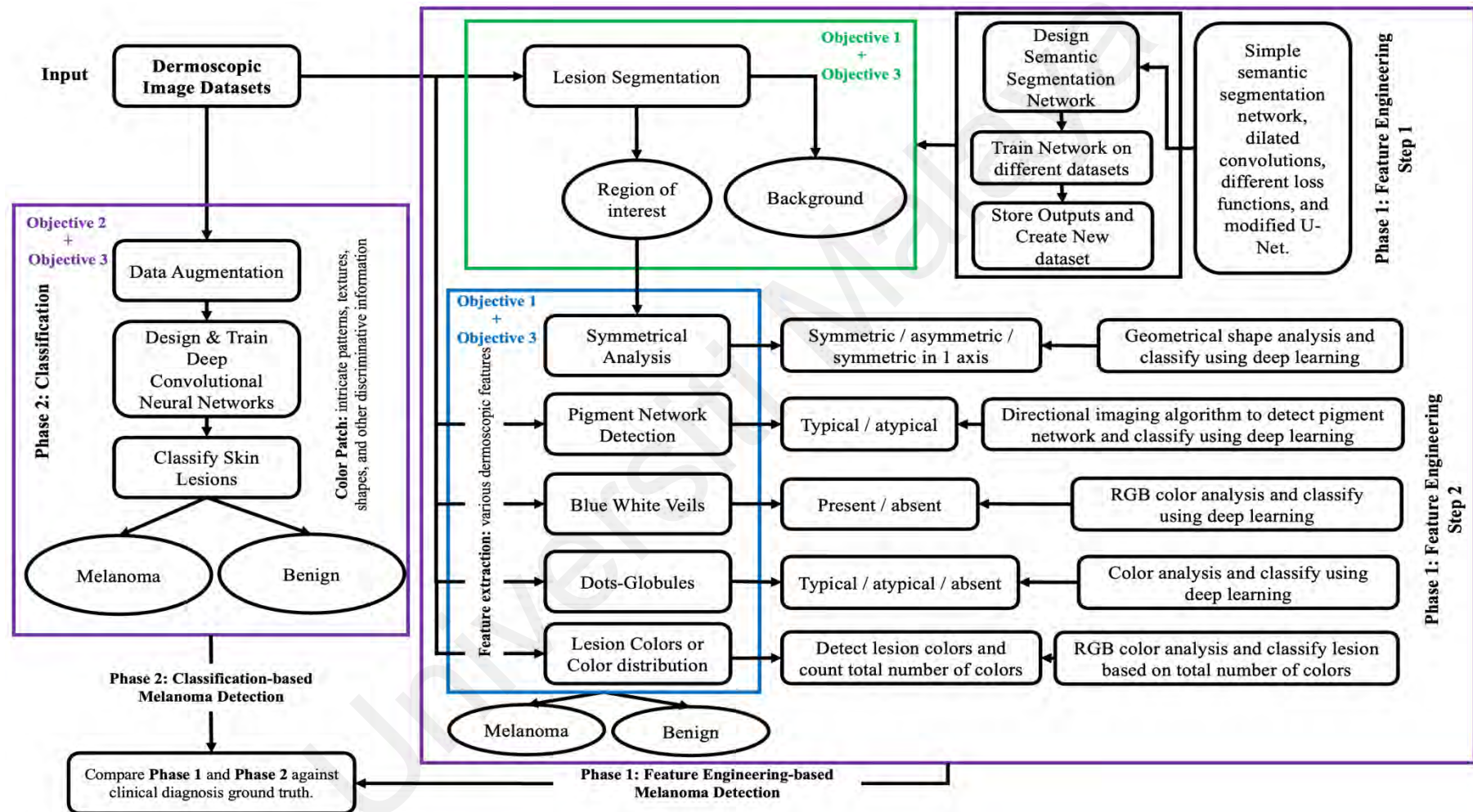
Figure 3.5: A conceptual framework of the overall proposed methodology.

3.5 The Source of Data

The multiple sources of dermoscopic images will be used to make this research concept and idea more robust. The data sources are-

1. PH2 consists of 200 high-resolution images with various types of skin diseases were downloaded from the PH2 database: a joint research collaboration between the Universidade do Porto, Instituto Superior Tecnico Lisboa, and the Dermatology Service of Hospital Pedro Hispano in Matosinhos, Portugal (Mendonca et al., 2013). All images were acquired under the same environment through the Tuebinger Mole Analyzer process by magnifying them 20 times. The image data is divided into three classes: Atypical Nevus, Common Nevus, and Melanoma.

2. MED-NODE is a non-dermoscopy (captured images without a dermoscopy device) dataset (Giotis et al., 2015). It includes 100 naevus and 70 melanoma images from the digital image archive of the Department of Dermatology of the University Medical Center Groningen used for the development and testing of the MED-NODE system for skin cancer detection from macroscopic images.

3. ISIC Archive is the combination of ISIC2016 (Gutman et al., 2016), ISIC2017 (Codella et al., 2017), ISIC2018 (Codella et al., 2018), ISIC2019 (Combalia et al., 2019), and ISIC2020 (Rotemberg et al., 2021) datasets. In the ISIC archive, thousands of dermoscopic images of various skin diseases are available for research purposes.

4. ISIC2016 means the International Skin Imaging Collaboration 2016 (Gutman et al., 2016). This is an academia and industry partnership designed to facilitate the application of digital skin imaging to help reduce melanoma mortality. The ISIC Grand Challenges have been run annually since 2016. ISIC2016 contains 1279 dermoscopic images with benign and malignant lesions.

5. ISIC2017 means the International Skin Imaging Collaboration 2017 (Codella et al., 2017). The initial focus of ISIC was on Melanoma. This dataset contains 2750 dermoscopic images with 374 Melanoma, 254 Seborrheic keratosis, and 1372 Nevus lesions.

6. HAM10000 means "Human Against Machine with 10000 training images" (Tschandl et al., 2018). This is a collection of dermoscopic images from different populations with different modalities. The dataset consists of 10015 dermoscopic images which can serve as a training set for academic machine learning purposes. Cases include a representative collection of all important diagnostic categories in the realm of pigmented lesions: Actinic keratoses and intraepithelial carcinoma, Basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, Melanoma, melanocytic nevi, and vascular lesions. More than 50% of lesions are confirmed through histopathology, the ground truth for the rest of the cases is either follow-up examination, expert consensus, or confirmation by in-vivo confocal microscopy.

7. Derm7pt is a database for evaluating computerized image-based prediction of the 7-point skin lesion malignancy checklist (Kawahara et al., 2019). The dataset includes 2013 clinical and dermoscopic color images, along with corresponding structured metadata tailored for training and evaluating computer-aided diagnosis systems (CADs).

8. A **"Test Set"** consists of 100 dermoscopic images from the ISIC2018 dataset (Codella et al., 2018), with clinically diagnosed ground truth, as well as features such as pigment network and dots-globules. The ground truth for dermoscopic asymmetry and blue-white veil features was established by the Dermatology division of Universiti Malaya, Malaysia. Additionally, the ground truth for the number of colors in the lesions was provided by the Cancer Research Lab of Bangladesh Cancer Research Institute,

Bangladesh. This test set is utilized to evaluate all proposed methods, including Phase 1 and Phase 2 of this research.

The description of these datasets is available in Table 3.3 based on their annotated information. In this table (-) means the ground truth is not available in the specific dataset.

Table 3.3: The annotated information of different datasets.

Name of Datasets	Segmentation	Feature Extraction and Classification					Classification
	Lesion Area	Asymmetry	Pigment Network	Blue White Veil	Dots-Globules	Number of Colors	Disease Diagnosed
ISIC2016	1279	-	-	-	Yes (detection)	-	2 classes
ISIC2017	2750	-	Yes (detection)	-	-	-	3 classes
PH2	200	Yes	Yes	Yes	Yes	Yes	3 classes
MED-NODE	170	-	-	-	-	-	2 classes
HAM10000	10015	-	-	-	-	-	7 classes
Derm7pt	2013	-	Yes	Yes	Yes	-	15 classes
Test Set	100	Yes	Yes	Yes	Yes	Yes	3 classes

The methodology of this research involves multiple steps divided into two phases. To address the lack of ground truth, multiracial datasets are utilized at different stages of the research. Several image data pre-processing techniques are applied in these steps. Table 3.4 presents the multiracial datasets used in the various stages of the two phases.

Table 3.4: The various datasets at different steps of two research phases.

Stage of Methodology	Name of Dataset	Number of Image Data
1. Segmentation	PH2	200
	ISIC2016	1279
	ISIC2017	2750
	HAM10000	10015
	Test Set	100
2. Asymmetry	PH2 mask	200
	Augmented PH2**	1585
	MED-NODE mask*	170
	Augmented MED-NODE**	354
	ISIC2016	1279
	ISIC2016 mask	1279
	Test Set	100
3. Pigment Network	PH2	200
	New PH2*	200
	MED-NODE	170
	ISIC (2019 + 2020)	100
	Test Set	100
4. Blue white veil	PH2	200
	Augmented PH2**	1300
	MED-NODE *	170
	Augmented MED-NODE**	1105
	Derm7pt	2013
	Test Set	100
5. Dots-Globules	PH2	200
	Augmented PH2**	400
	Derm7pt	2013
	Test Set	100
6. Lesion Colors	PH2	200
	MED-NODE*	170
	Combine**	370
	Test Set	100
7. Classification-based	PH2	200
	Augmented PH2**	400
	ISIC Archive	300
	Test Set	100

Note. * Refers without labels, ** refers to new variations of datasets created in the proposed work based on image processing techniques.

In Table 3.4, the Name of Dataset column shows datasets with a single asterisk (*) and

with a double asterisk (**) sign. A single asterisk means those datasets did not provide

any ground truth (i.e., no classification labels). However, these ground truths can be acquired upon performing data pre-processing on the dermoscopic images in different steps of the methodology. These ground truths are based on computer vision and pattern recognition algorithms. Those algorithms are discussed and reported in Chapter 5: Dermatological Feature Analysis. A double asterisk (**) means those datasets are formed with the data augmentation process and/or the combination of multiple datasets.

3.6 Data Pre-Processing

Image data pre-processing is a crucial step in computer vision and deep learning projects. It refers to the techniques used to prepare the images (by resizing, scaling, gray-scaling, removing noise, enhancing contrast, and detecting edges) for analysis or model training. The following are some of the common image pre-processing techniques applied in Phase 1 (including the process of segmentation, asymmetry, pigment networks, blue white veil, number of colors, and dots-globules) and Phase 2.

1. Resizing: Resizing images to a standard size is important to ensure that all images in the dataset have the same dimensions. This is important for models to work properly. The input image size for this research is 256x256 pixels except for the pigment network detection process. Higher resolution is required to detect pigment networks on the dermoscopic images. Hence, the image size is 512x512 pixels for the pigment detection process only.

2. Image Augmentation: Image augmentation is a technique used to artificially increase the size of the dataset by creating variations of existing images. This can help to reduce overfitting in deep learning models. However, there is no requirement to do data augmentation for the lesion segmentation process. This is because the original datasets

consist of sufficient image data where each image data has been predefined with a lesion mask. A lesion mask is a binary image where the black (0) pixel represents healthy skin, and the white (1) pixel represents unhealthy skin.

For dermatological asymmetry analysis, the PH2, MED-NODE, and ISIC2016 datasets are used. In this analysis process, the lesion mask (binary image) is required. PH2 and ISIC2016 datasets provide mask images. For the MED-NODE dataset, as will be discussed (see Chapter 5: Dermatological Feature Analysis), the mask images needed to be created by the new proposed algorithm. After creating the mask dataset for MED-NODE, all images are classified into three classes (asymmetry, symmetry, and half symmetry) by using another algorithm which also is reported in Chapter 5. Finally, the augmented PH2 and augmented MED-NODE datasets are created by using different augmentation techniques such as rotation (5, 10, 90, 180, and 270 degrees), adjusting brightness, and flipping. For the PH2 dataset, 200 images (117 symmetrical, 31 partial-symmetrical, and 52 asymmetrical) to 1585 images (571 symmetrical, 496 partial-symmetrical, and 520 asymmetrical) are created. For the MED-NODE dataset, 170 images (58 symmetrical, 19 partial-symmetrical, and 93 asymmetrical) to 354 images (120 symmetrical, 114 partial-symmetrical, and 120 asymmetrical) are created.

There is no requirement for the augmentation process for the pigment network analysis. However, a directional imaging algorithm (see Chapter 5: Dermatological Feature Analysis) is required to create a New PH2 dataset. The directional imaging algorithm refers to a computational or mathematical technique used to process and enhance images in specific directions or orientations. This algorithm is designed to identify and extract features or information from images along predefined directions, such as horizontal, vertical, diagonal, or custom angles, enabling targeted analysis and improvement of image quality based on directional characteristics.

For the blue white veil analysis, the PH2, MED-NODE, and Derm7pt datasets are used. The augmentation process is applied to PH2 and MED-NODE datasets. MED-NODE has no ground truth about the blue white veil. Hence, a specific algorithm is employed to classify the image data of MED-NODE into two classes (44 blue white veil images and 126 non-blue white veil images). After that, the augmentation process is done using various angles, flipping, and zooming on PH2 and MED-NODE datasets. This process increased the PH2 dataset to 1300 data from 200 data and the MED-NODE dataset was increased to 1105 data from 170 data.

The PH2 and Derm7pt datasets are used for analyzing dermatological feature dots-globules. For this analysis, only the PH2 dataset is applied (due to the small amount of data) for the augmentation process by randomly rotating 90, 180, and 270 degrees. Since the PH2 dataset has ground truth, image pre-processing is not required. Upon the augmentation, the total number of data is 400 for this dataset, which is divided into three classes such as absent, atypical, and typical.

For the lesion color analysis, PH2 and MED-NODE datasets are used. Since MED-NODE does not provide any information about the lesion color, a specific algorithm is created to detect the total number of colors on a lesion. This algorithm provides the total number of colors on the different lesions of the MED-NODE dataset (see Chapter 5: Dermatological Feature Analysis). Based on the number of colors, all image data are classified into six classes. Then the PH2 and MED-NODE datasets are combined to increase the amount of data. After the combination of these datasets, the total number of data is 370 (200 of PH2 and 170 of MED-NODE).

Finally, for the Classification-based Melanoma detection method (Phase 2), PH2 and ISIC archive datasets are used to classify the disease into three classes (atypical nevus, typical nevus, and Melanoma) in Chapter 6: Classification Based Melanoma Detection.

The augmentation process is used to increase the amount of data from 200 to 400 by applying rotation techniques including 90, 180, and 270 degrees. This increases atypical nevus images by 80 to 160, typical nevus images by 80 to 160, and Melanoma images by 40 to 80.

It is important to note that not all these pre-processing techniques will be required for every stage of methodology, and the specific techniques used will depend on the nature of the problem and the data.

3.7   Evaluation Metrics for Proposed Methods

Several algorithms are proposed in this research. Most of them are segmentation and classification-based. For evaluating those proposed methods, different evaluation metrics are required. There are various evaluation metrics for the segmentation and classification process. They are-

A. Segmentation: To evaluate the lesion segmentation process different evaluation metrics are used such as mean accuracy (mAC), global accuracy (gAC), mean intersection-over-union ($mIoU$), weighted intersection-over-union ($wIoU$), and mean boundary F1 score (mBF). Mathematically they are presented as-

$$mAC = \frac{TP}{TP + FN} \tag{1}$$

$$gAC = \frac{TP}{TP + FN} \tag{2}$$

$$IoU = \frac{area\ of\ overlap}{area\ of\ union}$$

$$\tag{3}$$

$$= \frac{TP}{TP + FP + FN}$$

$$mIoU = \frac{\sum(each\ class\ IoU)}{no.\ of\ total\ class} \qquad (4)$$

$$wIoU = \frac{\sum(no.\ of\ each\ class\ sample \times each\ class\ IoU)}{no.\ of\ total\ samples} \qquad (5)$$

$$mBF = \frac{2\ \times area\ of\ overlap}{total\ no.\ of\ pixels\ in\ ground\ truth\ and\ predicted\ images}$$
$$(6)$$
$$= \frac{2TP}{2TP + FN + FP}$$

Here, *mAC* is the average accuracy of all classes in all images; and *gAC* is the ratio of correctly identified classifier pixels, regardless of class, to the total number of pixels.

B. Classification: The experimental results of the classifiers report these metrics-sensitivity (*SE*), specificity (*SP*), precision (*PR*), accuracy (*AC*), and F1 score (*F1*). The experimental datasets containing more than two classes and data is imbalanced. To calculate the average evaluation metrics for the different classes (weighted metrics are recommended to use over the standard ones when data is imbalanced), the weighted-sensitivity (*wSE*), weighted-specificity (*wSP*), weighted-precision (*wPR*), weighted accuracy (*wAC*), and weighted-F1 score (*wF1*) are counted. Mathematically they are presented as-

$$SE = \frac{TP}{TP + FN} \qquad (7)$$

$$SP = \frac{TN}{FP + TN} \qquad (8)$$

$$PR = \frac{TP}{TP + FP} \qquad (9)$$

$$AC = \frac{TP + TN}{P + N} \qquad (10)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{11}$$

$$wSE = \frac{\sum(no.\,of\,each\,class\,sample \times each\,class\,SE)}{no.\,of\,total\,samples} \tag{12}$$

$$wSP = \frac{\sum(no.\,of\,each\,class\,sample \times each\,class\,SP)}{no.\,of\,total\,samples} \tag{13}$$

$$wPR = \frac{\sum(no.\,of\,each\,class\,sample \times each\,class\,PR)}{no.\,of\,total\,samples} \tag{14}$$

$$wAC = \frac{\sum(no.\,of\,each\,class\,sample \times each\,class\,AC)}{no.\,of\,total\,samples} \tag{15}$$

$$wF1 = \frac{\sum(no.\,of\,each\,class\,sample \times each\,class\,F1)}{no.\,of\,total\,samples} \tag{16}$$

For both segmentation and classification evaluation metrics, TP is true positive, TN is true negative, FN is false negative, FP is false positive, P is actual positive, and N is actual negative number of the image data. Equation (1) to equation (16) all are used to evaluate the proposed algorithms in Chapter 4: Skin Lesion Segmentation, Chapter 5: Dermatological Feature Analysis, and Chapter 6: Classification Based Melanoma Detection.

3.8   Required Tools

This experimental research will be implemented in MATLAB environment. The MATLAB 2019, 2020, 2021, 2022, 2023, and 2024 versions are engaged as a platform for developing the coding sections. Three different tools will be involved in the MATLAB environment. They are-

1. Image processing tool and computer vision tool,

2. Pattern recognition tool, and

3. Deep learning tool.

3.9   Methodological Stages in Different Chapters

The different steps of analysis are reported sequentially in different chapters. Table 3.5 presents the location of those reported analysis processes.

Table 3.5: Chapter-based location of the analysis process.

Objectives	Focus	Chapter: Sub-section	
$1^{st}+3^{rd}$	Lesion Segmentation	Chapter 4	All
$1^{st}+3^{rd}$	Feature 1- Asymmetry	Chapter 5	5.2
$1^{st}+3^{rd}$	Feature 2- Pigment Network	Chapter 5	5.3
$1^{st}+3^{rd}$	Feature 3- Blue white veil	Chapter 5	5.4
$1^{st}+3^{rd}$	Feature 4- Dots-Globules	Chapter 5	5.5
$1^{st}+3^{rd}$	Feature 5- Lesion Colors	Chapter 5	5.6
$2^{nd}+3^{rd}$	Automated detection (Classification)	Chapter 6	All

3.10   Chapter Summary

This chapter briefly describes the pipeline of the proposed dermoscopic image analysis research for Melanoma diagnosis. The two distinct phases of the research and the several stages of the research methodologies are described here which present the overview of this thesis. In the following chapters, a wide discussion and explanation of these methodological stages are available.

CHAPTER 4: SKIN LESION SEGMENTATION

4.1 Background

Lesion segmentation is the first step of the feature engineering phase in this thesis (see Figure 4.1). Skin lesion segmentation for Melanoma is the process of separating the suspicious areas of skin lesions from the surrounding normal skin to extract dermoscopic features which help to diagnosis Melanoma. This is typically done using computer vision algorithms that analyze images of the skin to identify regions that may be indicative of Melanoma.
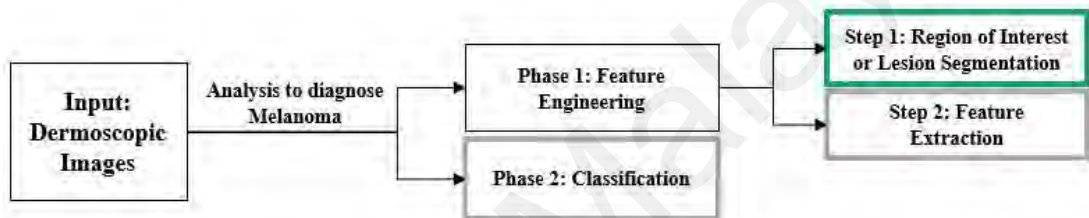
Figure 4.1: From the different steps of this thesis, region of interest or lesion segmentation (Phase 1: Step 1) is analyzed here (green color box).

A proper skin lesion segmentation algorithm can limit the region of interest (ROI) to save time for the analysis process and to extract the feature's information for diagnosing Melanoma more specifically and precisely from the dermoscopic images. Early diagnosis and delineation of the lesion boundaries are crucial for precise malignant region localization and clinical treatment of skin lesions. Skin cancer, particularly malignant Melanoma with its significant potential for metastasis, is more prevalent than commonly believed. Therefore, early detection is essential for treating it before malignancy develops. Medical imaging is utilized to analyze and segment lesion boundaries from dermoscopic images to solve these issues. The approaches range from visual inspection to textural analysis of the images (human and automated). However, due to the sensitivity involved in surgical procedures or drug administrations, the accuracy of these systems is

poor for proper clinical therapy. To improve patient outcomes and healthcare productivity, accurate segmentation helps to develop automated diagnostic tools, reduces human subjectivity, and ensures reliable data for quantitative analysis. This presents an opportunity to develop an automated model with good accuracy so that it can be used to extract dermoscopic features. Here, four popular publicly available dermoscopic datasets are used which were annotated by the dataset creators. These datasets with ground truth are- PH2, ISIC2016, ISIC2017, and HAM10000 (already described in Chapter 3: Research Methodology). The lesion masks in all these datasets are in binary images (consisting of pixels that can have one of exactly two colors, usually black and white). The black pixels represent non-disease, and the white pixels represent the disease-infected area on those dermoscopic images. Figure 4.2 shows one example of these images from the PH2 dataset.
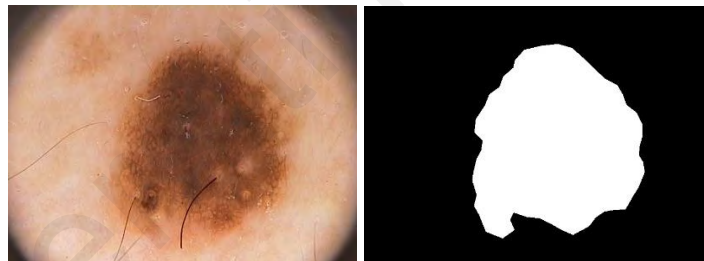
Figure 4.2: The left-side image is a dermoscopic image and the right-side image is a lesion mask. The black pixel indicates the non-disease area, and the white pixel indicates the disease-infected area.

4.2 Segmentation Method Overview

From the literature, the existing segmentation methods are complex to implement and not able to provide remarkable success. Due to these reasons, a simple and effective automatic segmentation process is required. That is why, to segment the disease-affected area on dermoscopic images, five different deep learning-based semantic segmentation networks are proposed. All of them are convolutional neural network-based models such

as simple semantic segmentation, semantic segmentation using dilated convolutions, semantic segmentation with Tversky loss function, semantic segmentation with Dice loss function, and modified U-Net. Then, their results are compared with each other's results as well as existing models' results by using different evaluation metrics (Equations 1 to 6, detailed in Chapter 3: Research Methodology). Figure 4.3 presents the overall architecture of the segmentation process.
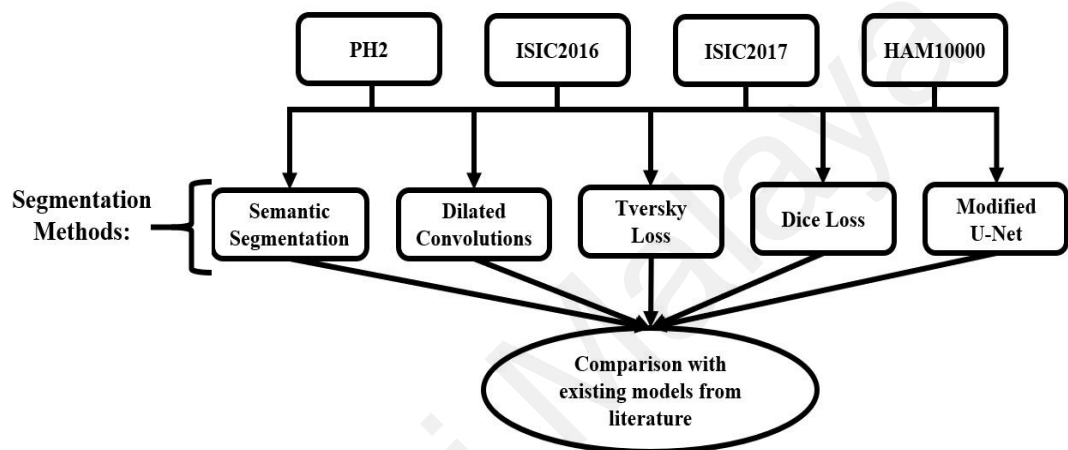
Figure 4.3: The different deep learning approaches for lesion segmentation on dermoscopic images.

4.3 A Simple Semantic Segmentation Network (Proposed Method: 01)

A semantic segmentation network can be an effective solution to isolate lesions (disease-infected areas or regions of interest) automatically from the healthier skin. Semantic segmentation describes the process of associating each pixel of an image with a class label (such as flower, person, road, sky, ocean, or car). Road segmentation for autonomous driving and cancer cell segmentation for medical diagnosis are two examples of applications for semantic segmentation (Singh & Rani, 2020).

Following the fundamental design of the semantic segmentation network, a straightforward semantic segmentation network is developed, and common layers of the

segmentation network are investigated. The downsampling of an image between convolutional and ReLU layers and subsequent upsampling of the output to match the input size is a prevalent behavior in semantic segmentation networks. The conventional scale-space analysis utilizing picture pyramids is equivalent to this process. However, a network executes the operations during this process using non-linear filters that are optimized for a particular set of classes that need to be segmented.

The image input layer, which establishes the smallest picture size the network can handle, is the first layer in a semantic segmentation network. Since most semantic segmentation networks are completely convolutional (their layers are largely convolutional), they can process images larger than the allowed input size. Here, the network processes 256x256 RGB images with an input image size of [128 128 3]. That means the network can process larger images, such as 256x256 RGB images, by leveraging the fully convolutional nature of the network. Besides, this network can accept inputs of different sizes while preserving the ability to perform semantic segmentation.

The basic idea behind designing semantic segmentation is to accurately assign a class label to each pixel in an image, thereby dividing the image into meaningful segments corresponding to different objects or regions. The process involves leveraging deep learning techniques, particularly convolutional neural network, to learn and infer the semantic information present in the image. Here are the key steps involved in designing semantic segmentation such as dataset preparation, network architecture, training, feature learning, upsampling and skip connections, and inference. By following these steps, semantic segmentation networks can effectively learn to understand the semantic content of images.

The convolution and ReLU layers serve as the foundation for building a downsampling network. The convolution layer padding is chosen so that the input size and output size

of the convolution layer match. As we move through the network, the input and output sizes between the majorities of layers stay constant, making network construction simpler. A max-pooling layer is used for the downsampling. The Max Pooling layer is a typical type of layer that is often added between convolutional layers in CNN. Its goal is to keep the most important components while reducing the input volume's spatial dimensions (width and height). By setting the 'Stride' option to 2, the max pooling layer is made to downsample the input by a factor of 2. A network is built by stacking the convolution, ReLU, and max pooling layers. This network downsamples its input by a factor of four. The transposed convolution layer, often known as the 'deconv' or 'deconvolution' layer, is then used to perform the upsampling. When a transposed convolution is used for upsampling, both filtering and upsampling are done simultaneously. The next step is to generate a transposed convolution layer with an upsample of two. To make the output size equal to twice the input size, the 'Cropping' parameter is set to 1. All ReLU layers as well as the transposed convolution are layered. This collection of layers upsamples an input by 4. An ablation study is done before determining the value of different parameters for different layers of this CNN. The best result is found from the above combination of parameters.

The last group of layers oversees classifying pixels. The input for these latter layers has the same spatial (height and width) dimensions as the input image. The third dimension's channel count is higher and is equal to the number of filters in the final transposed convolution layer, though. We must reduce the third dimension to fit the number of classes we want to divide. This can be accomplished by using a 1-by-1 convolution layer with, for instance, three filters. The third dimension of the input feature maps is combined to reduce the number of classes using a convolution layer. The softmax and pixel categorization layers come after this 1-by-1 convolution layer. These two layers work together to forecast the categorical label for each image pixel. Finally, all layers are

stacked to complete the semantic segmentation network. Figure 4.4 presents the architecture of this semantic network for dermoscopic image segmentation.
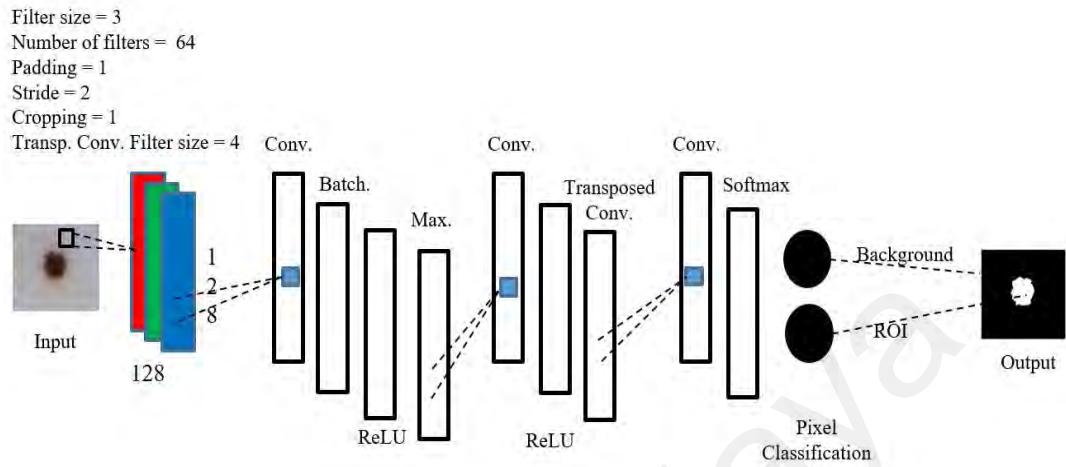
Figure 4.4: Architecture of the proposed semantic segmentation network.

4.4 Semantic Segmentation Network with Dilated Convolution (Proposed Method: 02)

The simple semantic segmentation network (Proposed Method: 01) may struggle to capture fine details and handle object boundaries due to limited receptive fields, being prone to over-smoothing of segmentation masks, and lacking the ability to capture context information effectively. The second proposed method for this research, semantic segmentation network with dilated convolution can be the solution to these gaps of the first proposed method. Dilated convolution, also known as atrous convolution, is a variant of the standard convolution operation used in CNN. It introduces the concept of dilation, which modifies the spacing between the values (or receptive fields) in the filter kernel.

In a standard convolution operation, a filter/kernel slides over the input image or feature map, and at each position, it performs a dot product between the filter weights and

the corresponding input values. The filter moves in a contiguous manner, with a fixed stride of 1, meaning it processes each pixel in the input.

Dilated convolution, on the other hand, allows for the expansion of the filter's receptive field by introducing gaps between the values in the filter kernel. These gaps are determined by a dilation rate, which specifies how much spacing there should be between the kernel values. Instead of a stride of 1, dilated convolution uses a stride equal to the dilation rate.

The main difference between dilated convolution and standard convolution lies in the way they sample the input. In standard convolution, all input pixels are sampled, and the receptive field is relatively limited. Dilated convolution, however, can have a larger receptive field by skipping input pixels based on the dilation rate. This enables the network to capture a broader context while reducing the number of parameters and computations compared to traditional methods like increasing the filter size or using pooling layers.

Dilated convolution has proven to be a crucial parameter/structure in various computer vision tasks for several reasons:

Increased receptive field: Dilated convolution allows models to capture larger contextual information without significantly increasing the number of parameters. This is particularly important when dealing with tasks such as semantic segmentation, where precise localization and understanding of objects require capturing long-range dependencies.

Fine-grained detail preservation: By utilizing dilated convolution with carefully chosen dilation rates, models can preserve fine-grained details while maintaining a

broader receptive field. This is especially valuable in tasks like image segmentation, where retaining precise object boundaries is crucial.

Efficient multi-scale processing: Dilated convolutions can be used in combination with multiple dilation rates in parallel, creating a multi-scale pyramid of receptive fields. This enables the network to analyze and integrate information at different scales, facilitating robust feature extraction and hierarchical reasoning.

Parameter efficiency: Dilated convolutions provide a way to increase the receptive field without increasing the number of parameters, which can be advantageous in scenarios with limited computational resources or memory constraints.

Overall, dilated convolution has become an essential tool in deep learning, particularly in computer vision tasks that involve capturing context, preserving details, and efficiently processing multi-scale information. Its ability to expand the receptive field while maintaining efficiency makes it a valuable technique in various state-of-the-art models.

Dilated convolutions, also referred to as atrous convolutions, are frequently used in semantic segmentation networks like DeepLab (Chen et al., 2018), as they can increase the receptive field of the layer (the area of the input that the layers can see) without adding more parameters or computations. The use of an image input layer with input size corresponding to the size of the input images has resulted in the creation of a network for pixel categorization. The specification of three convolution blocks, batch normalization, and ReLU layers follows. Inputs are padded to be the same size as outputs by setting the "Padding" option to "same" for each convolutional layer, which uses 128 3-by-3 filters with increasing dilation factors. A convolutional layer with K 1-by-1 convolutions, where K is the number of classes, is used to categorize the pixels. This layer is then followed by a softmax layer and a pixel classification layer with the inverse class weights. Figure 4.5

presents the architecture of this semantic segmentation network with dilated convolutions.
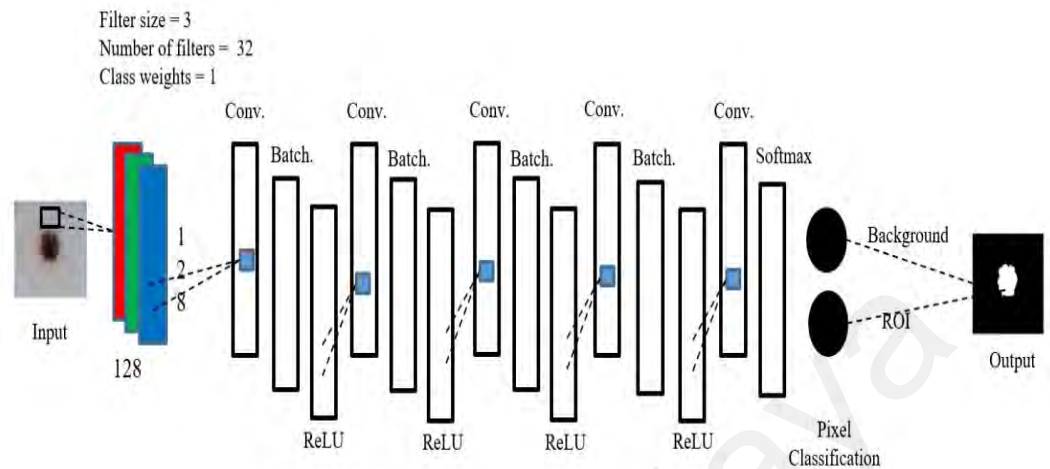
Figure 4.5: Architecture of the proposed semantic segmentation network with dilated convolutions.

4.5 Custom Pixel Classification Layer with Tversky Loss Function for Semantic Segmentation Network (Proposed Method: 03)

The semantic segmentation network with dilated convolution (Proposed Method: 02) may require more computational resources compared to simple architectures due to the increased receptive fields and training such networks might be slightly more complex due to the introduction of dilated convolutions. To overcome these issues of the first and second proposed methods, the third proposed method was introduced.

A Custom Pixel Classification Layer with a Tversky Loss Function for Semantic Segmentation Network refers to a specific configuration used in the field of computer vision, particularly in semantic segmentation tasks. In a semantic segmentation network, the pixel classification layer is responsible for assigning class labels to individual pixels in an image. It takes the output of the network (usually a feature map) and performs pixel-wise classification, predicting the class label for each pixel. A loss function is a

mathematical function used in various machine learning and optimization algorithms to measure the difference between the predicted values generated by a model and the actual target values in a dataset. The purpose of a loss function is to quantify how well or poorly a model is performing its task.

The Tversky loss function is a variation of the Dice loss, which is commonly used in semantic segmentation tasks. It measures the dissimilarity between the predicted segmentation map and the ground truth by computing the overlap or similarity between the two. The Tversky loss function introduces two parameters, namely α and β, which control the importance given to false positives (FP) and false negatives (FN) respectively. By adjusting these parameters, the Tversky loss can be tailored to address the specific requirements of a segmentation task. It provides a flexible way to penalize certain types of errors more or less than others.

The Tversky loss is based on the Tversky index for measuring overlap between two segmented images (Salehi et al. 2017). The Tversky index TI_c between one image Y and the corresponding ground truth T is given by-

$$\text{TI}_c = \frac{\sum_{Mm=1} Y^{cm} T^{cm}}{\sum_{Mm=1} Y^{cm} T^{cm} + \alpha \sum_{Mm=1} Y^{cm} T^{\bar{c}m} + \beta \sum_{Mm=1} Y^{\bar{c}m} T^{cm}} \quad (17)$$

Here, c corresponds to the class and $\bar{c}$ corresponds to not being in class c. M is the number of elements along the first two dimensions of Y. α and β are weighting factors that control the contribution that false positives and false negatives for each class make to the loss. The loss L over the number of classes C is given by-

$$L = C \sum_{c=1} 1 - \text{Ti}_c \quad (18)$$

91

Equation (18) is used in the pixel classification layer of the Proposed Method: 01 which is a semantic segmentation network with the Tversky loss function.

The term "Custom Pixel Classification Layer" indicates that the pixel classification layer in this network has been customized or modified to incorporate the Tversky loss function. Instead of using a standard loss function like cross-entropy or Dice loss, the Tversky loss function has been integrated into the network's architecture to train the model for semantic segmentation.

The combination of a custom pixel classification layer with the Tversky loss function allows the network to optimize its parameters based on the specific objectives and requirements of the semantic segmentation task, potentially improving segmentation accuracy and handling class imbalance or other challenges associated with pixel-wise classification. This is the Proposed Method: 03 which is a semantic segmentation network with a custom pixel classification layer where the Tversky loss function is applied.

4.6   Custom Pixel Classification Layer with Dice Loss Function for Semantic Segmentation Network (Proposed Method: 04)

Here, to compare with the third proposed method, the dice loss function is used in the custom pixel classification layer of the semantic segmentation network. The Dice loss function is widely used in semantic segmentation to assess the resemblance between predicted and ground truth segmentation masks. It relies on the Dice coefficient, which quantifies the overlap between two sets. Specifically designed for semantic segmentation, this metric calculates the similarity between predicted and ground truth masks.

The Dice loss is derived from the Sorensen-Dice similarity coefficient, which is a measure of overlap between two segmented images. The generalized Dice loss (Crum et

al., 2006; Zhang et al., 2021) L for between one image Y and the corresponding ground truth T is given by-

$$L = 1 - \frac{2 \sum_{k=1}^{K} W_k \sum_{m=1}^{M} Y_{km} T_{km}}{\sum_{k=1}^{K} W_k \sum_{m=1}^{M} Y_{km}^2 + T_{km}^2} \tag{19}$$

Here K is the number of classes, M is the number of elements along the first two dimensions of Y, and w_k is a class-specific weighting factor that controls the contribution each class makes to the loss. w_k is typically the inverse area of the expected region:

$$w_k = \frac{1}{(\sum_{m=1}^{M} T_{km})^2} \tag{20}$$

This weighting helps counter the influence of larger regions on the Dice score and makes it easier for the network to learn how to segment smaller regions. Next, the equation (19) is used in the pixel classification layer of the Proposed Method: 01. Therefore, using this Dice Loss function in the pixel classification layer of Semantic Segmentation Network is the Proposed Method: 04. The initial goal of this proposed method is to compare the final segmentation results of other proposed methods.

4.7   Modified U-Net for Semantic Segmentation of Skin Lesion (Proposed Method: 05)

The U-Net architecture (Ronneberger et al., 2015) has been widely adopted for biomedical semantic segmentation tasks due to its effectiveness in capturing fine-grained details and handling limited training data. When applied to dermoscopic images for skin lesion segmentation, a modified U-Net can be particularly advantageous. The Modified U-Net is an extension of the original U-Net, which incorporates additional modifications

to improve segmentation performance. To compare with the proposed methods 1 to 4, this modification of U-Net is done which is the fifth proposed method for the segmentation process in this thesis. This modified U-Net model can be highly effective for semantic segmentation on dermoscopic images due to its complex and deeper architecture.

The U-Net architecture consists of an encoder-decoder network with skip connections. The encoder path captures context information by progressively downsampling the input image, while the decoder path recovers spatial information through upsampling. Skip connections are established between corresponding layers of the encoder and decoder, allowing the model to combine low-level and high-level features effectively. The Modified U-Net incorporates several modifications to enhance its performance for semantic segmentation tasks by different ablation studies:

Dilated Convolutions: Traditional convolutions have a fixed receptive field, limiting their ability to capture both local and global contexts. The Modified U-Net employs dilated convolutions in the encoder path, enabling a larger receptive field without losing resolution. This helps in capturing contextual information at multiple scales.

Residual Connections: To facilitate gradient flow and mitigate the vanishing gradient problem during training, residual connections are introduced within the encoder path. These connections alleviate the information loss that occurs during downsampling, ensuring that useful details are preserved.

Attention Mechanism: The Modified U-Net incorporates an attention mechanism to selectively focus on informative regions of the input image. By assigning importance weights to different spatial locations, the model can emphasize relevant regions and

suppress irrelevant ones. This mechanism enhances the segmentation accuracy by directing the model's attention to discriminative features.

Multi-scale Feature Fusion: To capture features at multiple scales and improve the segmentation performance, the Modified U-Net introduces a multi-scale feature fusion module. It combines features from different encoder layers, each with a different receptive field, to enrich the representation used for segmentation. The Modified U-Net offers several benefits over the original architecture:

Improved Contextual Information: By incorporating dilated convolutions, the Modified U-Net can capture contextual information at multiple scales, leading to better segmentation accuracy.

Enhanced Gradient Flow: The addition of residual connections mitigates the vanishing gradient problem, allowing more stable and effective training of deep networks.

Attention-guided Segmentation: The attention mechanism helps the model focus on relevant regions, leading to improved localization and discrimination of objects during segmentation.

Multi-scale Feature Fusion: The multi-scale feature fusion module enables the model to leverage features from different scales, enhancing the ability to handle objects of varying sizes and complexities.

The Modified U-Net architecture builds upon the strengths of the original U-Net and incorporates additional modifications to further improve semantic segmentation performance. By leveraging dilated convolutions, residual connections, attention mechanisms, and multi-scale feature fusion, the Modified U-Net achieves enhanced context modeling, gradient flow, attention-guided segmentation, and multi-scale feature

representation. These advancements make it a powerful tool for segmenting the skin lesion or region of interest in dermoscopic and non-dermoscopic images. The architecture of the modified U-Net is presented in Figure 4.6.
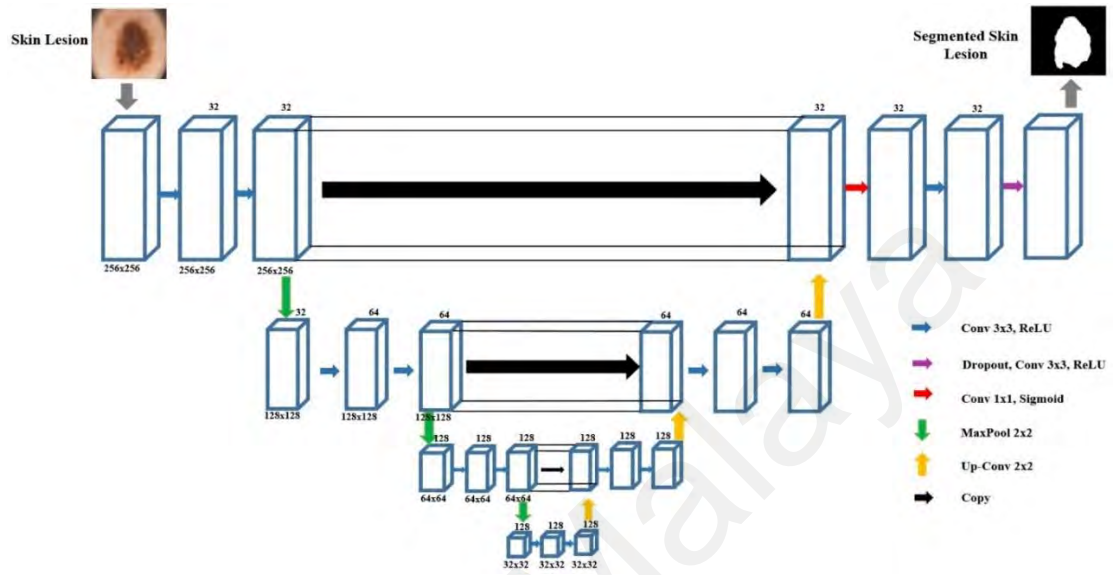
Figure 4.6: Modified U-Net architecture.

4.8   Training the Proposed Networks

All four datasets are divided into a training set (70%), validation set (20%), and test set (10%) individually for these semantic networks. The training and validation sets are shuffled during the training session. The test set is only used for testing purposes. The testing results of those test sets are reported in Section 4.10: Experimental Results. Different ablation studies show that semantic networks with Tversky and modified U-Net provide better performance when the Adam optimizer is used. Adam Optimizer is known for its effectiveness in accelerating the training process and converging to good solutions. In other semantic networks, this study found that the Stochastic Gradient Descent (SGD) optimizer is better to use. The maximum number of epochs is 250 (however, sometimes the number of epochs was increased and decreased based on the total number of training images), the size of the mini-batch is 64, and the initial learning rate (a parameter that

determines the step size at each iteration when updating the model's parameters) is 1e-3. The changing of other parameters' values such as the learning rate of drop factor and drop period are required based on the architecture of the networks. Different neural network architectures have varying complexities, depths, and internal structures. Deeper networks might require smaller learning rates to prevent vanishing gradients, while shallower networks might tolerate larger learning rates. Similarly, architectures with more parameters might benefit from stronger regularization, such as higher drop factors or longer drop periods. Finding an appropriate learning rate was achieved through techniques like learning rate schedules, adaptive learning rate methods, and using learning rate ranges during the experimentation. Experimentation, validation, and tuning were essential to determine the values that lead to optimal convergence, generalization, and performance for the proposed methods 1 to 5.

4.9 Learnable Parameters and Training Time Comparison of Proposed Methods

In this section, above mentioned proposed methods 1 to 5 are compared with the standard U-Net model based on the number of learnable parameters and layers, and the required training time on different multiracial dermoscopic datasets. The standard U-Net has the highest number of learnable parameters and layers, so the training periods are longer than other methods on different datasets. On the other hand, the first proposed method which is a simple semantic segmentation network, has the lowest number of learnable parameters and layers. As a result, in all datasets, this model requires the shortest training period. It is worth mentioning that the third and fourth proposed methods have only 11 and 10 layers. However, the third and fourth methods have a significantly large number of learnable parameters. For this reason, their training periods are also quite

long. Table 4.1 presents a comparison between all proposed methods and the standard U-Net model.

<p style="text-align:center">Table 4.1: Comparing different models with their learnable parameters and training time.</p>

Models	Learnable parameters	Number of layers	Training Data 70% of Original Datasets			
			PH2 140 images	ISIC2016 895 images	ISIC2017 1925 images	HAM10000 7011 images
Proposed Method: 01	19.6 K	13	57 sec	6 min 6 sec	13 min 7 sec	47 min 48 sec
Proposed Method: 02	28.9 K	16	1 min 4 sec	6 min 47 sec	14 min 35 sec	53 min 7 sec
Proposed Method: 03	104.5 K	11	7 min 25 sec	47 min 27 sec	102 min 5 sec	371 min 48 sec
Proposed Method: 04	104.4 K	10	6 min 41 sec	42 min 43 sec	91 min 52 sec	334 min 37 sec
Proposed Method: 05	7.6 M	46	8 min 54 sec	56 min 57 sec	122 min 30 sec	446 min 9 sec
U-Net	31.0 M	58	30 min 45 sec	196 min 37 sec	422 min 55 sec	1540 min 17 sec
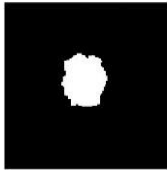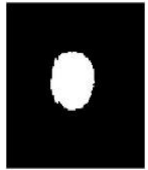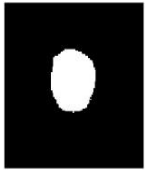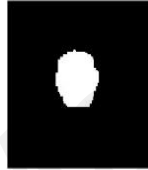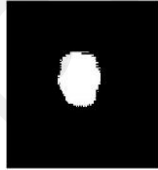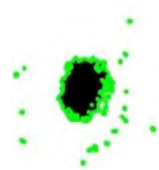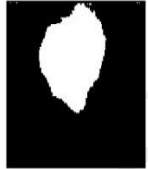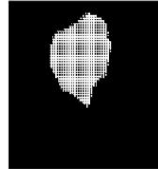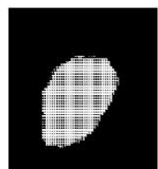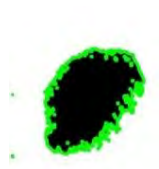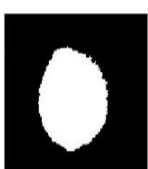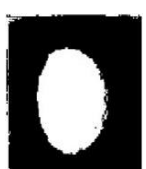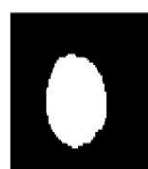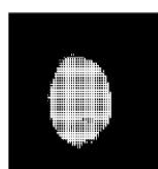
4.10   Experimental Results

After conducting the segmentation experiment with these five proposed methods on four multiracial datasets, the final outputs are analyzed and assessed in two different ways. They are-

1. Qualitative Assessment: After implementing these proposed models, four outputs from four different datasets are presented in Table 4.2. Input images from the four different datasets, their ground truth, semantic segmentation network (PM 01), dilated convolution network (PM 02), semantic segmentation network with Tversky loss

functions (PM 03), semantic segmentation network with Dice loss functions (PM 04), modified U-Net (PM 05), and the standard U-Net outputs are compared side by side in Table 4.2. Segmentation refers to the process of dividing or partitioning an image into meaningful or distinct regions or objects. It involves identifying and separating different regions of interest within an image based on certain characteristics, such as color, texture, intensity, or other visual properties. This segmentation process involves separating the skin lesion from the surrounding healthy skin tissue. Table 4.2 shows the binary images after the segmentation process where white pixels mean the region of interest (disease-infected) and black pixels mean the healthy area (non-disease-infected). The first column is a dataset, the second column is input images which were used as input for the proposed methods, the third column is the ground truth in binary images which were provided by the dataset creators, the fourth column is Proposed Method: 01, fifth column is Proposed Method: 02, the sixth column is Proposed Method: 03, the seventh column is Proposed Method: 04, the eighth column is Proposed Method: 05, and ninth column is U-Net in Table 4.2.

Table 4.2: Comparison between the outputs of different models on four different datasets.

Dataset	Input Image	Ground Truth	PM 01	PM 02	PM 03	PM 04	PM 05	U-Net
PH2								
ISIC2016								
ISIC2017								
HAM10000								

2. Quantitative Assessment: All five proposed methods including semantic segmentation, semantic with dilated convolution, semantic with Tversky, semantic with dice, and modified U-Net are trained on the PH2, ISIC2016, ISIC2017, and HAM10000 datasets. Also, U-Net and the other four models (see Chapter 2: Section 2.2.1) from the literature are trained for comparative analysis on those datasets. All of them are evaluated by using equations (1) to (6) from Chapter 3: Section 3.6 in Table 4.3 to Table 4.6 (the best results are bolded). Barın & Güraksın (2022) focused on a computer-aided automatic segmentation process by proposing a hybrid FCN-ResAlexNet architecture. Khan et al. (2021) proposed a method for multiclass skin lesion segmentation and classification which involved image enhancement, saliency estimation, and thresholding. Goyal et al. (2020), proposed fully automated deep-learning ensemble methods for accurate lesion boundary segmentation. Adegun & Viriri (2020) proposed a deep learning-based method for lesion detection and segmentation which utilized an enhanced encoder-decoder network with skip pathways and a multi-stage, multi-scale approach.

Table 4.3: Performance comparison of different models on the PH2 dataset.

Models	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
PM 01	0.98	0.97	0.91	0.96	0.83
PM 02	0.95	0.96	0.82	0.91	0.80
PM 03	**0.99**	**0.98**	**0.93**	0.97	**0.84**
PM 04	0.96	0.94	0.84	0.92	0.77
PM 05	0.96	0.96	0.85	**0.98**	0.76
U-Net	0.92	0.91	0.76	0.87	0.72
Barın & Güraksın, 2022	0.95	0.93	0.86	0.93	0.78
Khan et al., 2021	0.94	0.93	0.85	0.96	0.83
Goyal et al., 2020	0.93	0.90	0.84	0.97	0.82
Adegun & Viriri, 2020	0.87	0.86	0.76	0.85	0.80

Table 4.4: Performance comparison of different models on the ISIC2016 dataset.

Models	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
PM 01	0.97	**0.96**	0.90	0.95	**0.83**
PM 02	0.93	0.95	0.80	0.89	0.79
PM 03	**0.98**	**0.96**	**0.92**	0.96	0.82
PM 04	0.95	0.93	0.83	0.91	0.76
PM 05	0.95	0.95	0.84	**0.97**	0.75
U-Net	0.91	0.90	0.75	0.86	0.71
Barın & Güraksın, 2022	0.93	0.92	0.83	0.90	0.76
Khan et al., 2021	0.94	0.94	0.80	0.88	0.82
Goyal et al., 2020	0.94	0.91	0.79	0.89	0.82
Adegun & Viriri, 2020	0.86	0.85	0.76	0.84	0.79

Table 4.5: Performance comparison of different models on the ISIC2017 dataset.

Models	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
PM 01	0.95	0.94	0.89	0.93	0.82
PM 02	0.92	0.93	0.79	0.86	0.78
PM 03	**0.97**	**0.95**	**0.91**	0.95	**0.83**
PM 04	0.94	0.92	0.82	0.90	0.75
PM 05	0.95	0.94	0.84	**0.98**	0.74
U-Net	0.90	0.89	0.74	0.85	0.71
Barın & Güraksın, 2022	0.95	0.94	0.86	0.94	0.79
Khan et al., 2021	0.93	0.93	0.82	0.90	0.82
Goyal et al., 2020	0.92	0.91	0.80	0.90	0.81
Adegun & Viriri, 2020	0.84	0.83	0.75	0.83	0.79

Table 4.6: Performance comparison of different models on the HAM10000 dataset.

Models	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
PM 01	**0.96**	**0.94**	0.88	0.93	**0.81**
PM 02	0.91	0.93	0.78	0.87	0.77
PM 03	**0.96**	**0.94**	**0.90**	0.94	**0.81**
PM 04	0.93	0.91	0.81	0.89	0.74
PM 05	0.93	0.93	0.82	**0.96**	0.73
U-Net	0.89	0.88	0.73	0.84	0.70
Barın & Güraksın, 2022	0.94	0.91	0.85	0.92	0.77
Khan et al., 2021	0.92	0.91	0.90	0.93	0.79
Goyal et al., 2020	0.91	0.90	0.89	0.93	0.79
Adegun & Viriri, 2020	0.90	0.89	0.76	0.82	0.78

In Table 4.3 to Table 4.6, the best results are bolded and the semantic segmentation network with the Tversky loss function has the best performance on all four datasets. The Tversky loss function is a variant of the Dice coefficient that measures the similarity between two sets. It is commonly used in image segmentation tasks, where it quantifies the overlap between predicted and ground truth regions. The Tversky loss function encourages the model to accurately capture the overlap between the predicted and ground truth regions while penalizing false positives and false negatives. Its optimization helps in training models that produce more accurate segmentations for various computer vision applications. Finally, another comparison is made to make sure that proposed models are capable of handling completely unknown data from a different dataset which are not used during training sessions. To avoid overfitting, all models are tested on a test set of ISIC2018 dataset. The test set of this dataset has 1000 dermoscopic images with ground truth. The performance of all models on the test set of the ISIC2018 dataset is reported in Table 4.7 after training on the combined PH2, ISIC2016, ISIC2017, and HAM10000 datasets.

Table 4.7: Performance comparison of different models on the ISIC2018 test set when trained on combined PH2, ISIC2016, ISIC2017, and HAM10000 datasets.

Models	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
PM 01	0.93	0.91	0.86	0.90	0.79
PM 02	0.88	0.90	0.75	0.84	0.75
PM 03	**0.94**	**0.92**	**0.88**	**0.92**	**0.80**
PM 04	0.90	0.88	0.78	0.86	0.71
PM 05	0.90	0.90	0.79	**0.92**	0.71
U-Net	0.86	0.85	0.70	0.81	0.68
Barın & Güraksın, 2022	0.91	0.89	0.83	0.89	0.74
Khan et al., 2021	0.89	0.88	0.87	0.90	0.76
Goyal et al., 2020	0.88	0.87	0.86	0.90	0.76
Adegun & Viriri, 2020	0.86	0.83	0.72	0.78	0.74

All models present almost the same performance on ISIC2018 test set as other test sets. The proposed semantic segmentation network with Tversky loss function has the best performance among all models. This table demonstrate the ability of the proposed models to handle unknown data when they trained on different dataset. Additionally, a performance map is drawn based on the performance of different models on five different test sets in Table 4.8.

Table 4.8: The proposed and existing models' performance maps based on five test sets.

Datasets / Models	PH2	ISIC2016	ISIC2017	HAM10000	ISIC2018
PM 01	2nd	2nd	2nd	2nd	2nd
PM 02	6th	6th	7th	7th	5th
PM 03	1st	1st	1st	1st	1st
PM 04	9th	9th	9th	9th	9th
PM 05	8th	8th	8th	8th	8th
U-Net	10th	10th	10th	10th	10th
Barın & Güraksın, 2022	5th	5th	5th	5th	6th
Khan et al., 2021	3rd	3rd	3rd	3rd	3rd
Goyal et al., 2020	4th	4th	4th	4th	4th
Adegun & Viriri, 2020	7th	7th	6th	6th	7th

4.11   Discussion

Lesion segmentation is crucial in the diagnosis of Melanoma as it enables accurate identification and delineation of malignant skin lesions. By precisely outlining the boundaries of the lesions, it aids in quantifying their size, shape, and texture, providing essential information for effective diagnosis and treatment planning. Accurate segmentation allows dermatologists to track lesion progression, monitor changes over time, and compare images for accurate assessment of Melanoma development, leading to early detection and improved patient outcomes.

Here, a simple semantic segmentation network (PM 01) is proposed to segment the lesion automatically and train on four different datasets. Then another segmentation network (PM 02) is proposed with dilated convolution and trained on the same datasets. Two different loss functions such as Tversky (in equations 17 and 18) and Dice (in equations 19 and 20) are applied on the pixel classification layer of the first (PM 01) semantic segmentation network to implement the third (PM 03) and fourth (PM 04) proposed models. They are also trained on the same datasets. For these four models, the goal was to make them simple, efficient, and easy to implement without any complexity of models' architecture. After that, the U-Net is modified (PM 05) by an ablation study (changing different parameters and layers) and trained on those datasets. Bringing the complexity to the proposed model was the goal for this modified U-Net. These five proposed models are designed from the sketch of fundamental understanding of convolutional neural networks. The different parameters of these five proposed models are hyper-turned using "Grid Search" algorithm where number of layers, filter size, stride, padding learning rate, and batch size are determined.

Then, the standard U-Net and four other existing models from the literature are trained on those datasets for comparison. The reason for choosing U-Net for the comparison is

that U-Net is a popular architecture widely used for medical image segmentation. It consists of an encoder-decoder structure with skip connections to capture both low-level and high-level features effectively. U-Net has been successful in many medical imaging tasks due to its ability to handle limited training data. Besides, those four existing models are- hybrid fully convolutional network based (Barın & Güraksın, 2022), deep saliency-based (Khan et al., 2021), ensemble: mask RCNN and DeepLab V3+ based (Goyal et al., 2020), and enhanced encoder-decoder network based (Adegun & Viriri, 2020). To ensure a fair comparison, all models are trained under the same environment and on the same datasets.

Since all those datasets have the ground truth, the evaluation metrics (see Chapter 3: Section 3.6) are calculated by analyzing the difference between the ground truth and those models' generated output. The proposed semantic segmentation network (PM 03) with the Tversky loss function has the best performance overall on all four datasets. This segmentation process helps to understand the region of interest more precisely and the isolated lesion area is used for extracting some important features such as dermatological asymmetry and lesion size which are the key features to detect Melanoma on the dermoscopic images.

The contribution of this chapter is five different proposed semantic segmentation models which are suitable only for dermoscopic images to identify suspected lesion area. This automatic lesion segmentation process is Step 1 of Phase 1: Feature Engineering of this thesis (see Chapter 3: Figure 3.1) to detect Melanoma. Experts and non-experts can be facilitated by understanding more properly and clearly where to focus on dermoscopic images after completing this process. This segmentation process is considered the accomplishment of the first step of objectives 01 and 03 of this thesis.

4.12   Chapter Summary

Five different semantic segmentation models were implemented, tested, and described in this chapter to segment skin lesions automatically in dermoscopic images using deep convolutional neural networks (CNN) based techniques. These models were trained on four different multiracial datasets and evaluated using the corresponding test sets and a completely unknown test set. Additionally, the U-Net and the other four existing models from the literature were trained to compare their performance with the proposed models. To ensure a fair comparison, all models were trained under the same environment and on the same four datasets.

Among the proposed models, the semantic segmentation model with Tversky loss function (PM 03) demonstrated superior performance, surpassing all other models in this experiment. This achievement stands as the main contribution of this chapter.

Moreover, the remaining four proposed models can enhance the accuracy and efficiency of Melanoma diagnosis. These models can identify suspicious regions within skin lesion images, enabling dermatologists and other medical professionals to focus their examination on these areas. Furthermore, they can aid in the development of automated diagnosis systems capable of swiftly and accurately detecting Melanoma from skin lesion images.

The next chapter will discuss Step 2: Feature Extraction of Phase 1: Feature Engineering in details. The dermoscopic features such as asymmetry, pigment network, blue white veil, dots-globules, and lesion colors or color distribution will be analyzed on the dermoscopic images to detect Melanoma.

CHAPTER 5: DERMATOLOGICAL FEATURES ANALYSIS

5.1 Introduction

Diagnosing Melanoma from dermoscopic images is a challenging task requiring dermatology and image analysis expertise. One of the crucial steps in this process is feature extraction (which refers to the process of identifying and extracting relevant patterns, structures, or characteristics from images), which involves identifying and quantifying the characteristics of the skin lesion that are indicative of Melanoma. Machine learning and deep learning techniques have been used to automate feature extraction from dermoscopic images.

Automatic feature extraction from dermoscopic images plays a crucial role in dermatology and Melanoma diagnosis. Dermoscopy is a non-invasive imaging technique that allows clinicians to examine skin lesions with increased magnification and visualization of subsurface structures. Extracting relevant features automatically from dermoscopic images offers several important benefits such as efficiency, scalability, research, development, and decision support. Based on many labeled images, these strategies involve training a model to recognize patterns and features that are suggestive of Melanoma. The model can be used to assess new images and determine if a skin lesion is most likely to be Melanoma after it has been trained.

Based on the most common clinically practiced methods (Adler et al., 2018), five dermoscopic features are selected in this thesis to diagnose Melanoma: asymmetry, pigment network, blue white veil, dots-globules, and lesion colors (see Chapter 1: Introduction). Figure 5.1 shows the different research steps and green color boxes are discussed and analyzed in this chapter. These important features are extracted individually across different methods using multiracial skin lesion datasets. After the

lesion segmentation process, this feature extraction process is the next step (second step) of the first phase (feature engineering phase).
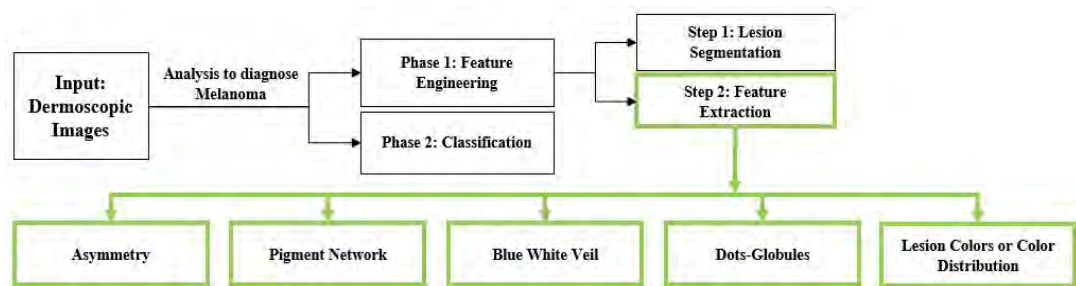
Figure 5.1: From the different steps of this thesis, different features extraction (Phase 1: Step 2) is analyzed here (green color boxes).

Individual features are meticulously extracted from dermoscopic images to provide a comprehensive description of the skin lesion's condition. This detailed analysis serves as a valuable aid for dermatologists and experts in the accurate detection of Melanoma. By reinforcing their clinical decisions, especially for young practitioners, feature analysis instills a heightened sense of confidence and precision in the diagnostic process. To avoid the complexity of understanding the process of feature extraction, each feature extraction algorithm is implemented, described, and compared (with existing approaches from the literature) individually in five different sections as shown in Figure 5.1.

5.2   Dermatological Feature: Lesion Asymmetry

Asymmetry is one of the key features used to detect Melanoma. Dermatological asymmetry refers to the unevenness or irregularity in the shape or size of a mole or lesion on the skin. To determine if a mole is asymmetrical, a doctor or dermatologist will examine it closely and compare one half of the mole to the other half. If the two halves do not match, it could be a sign of Melanoma. Figure 5.2 presents two lesions with different shapes.
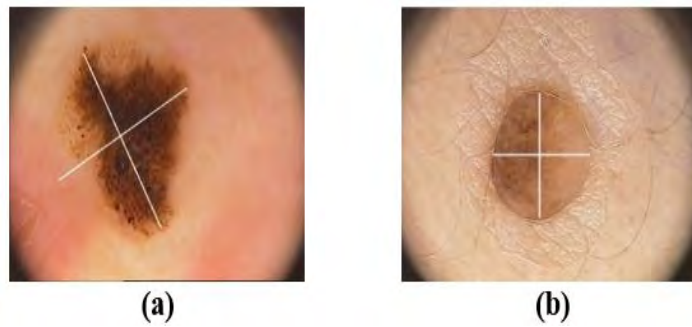
Figure 5.2: (a) asymmetrical lesion (do not match); and (b) symmetrical lesion (almost match).

Different types of lesion shapes (symmetry, half symmetry, and asymmetry) in dermoscopic images require clinical analysis and identification. Three types of patterns are usually visible in a skin lesion. However, a sharp eye is required to distinguish these patterns as usually exhibited by a skin specialist and a dermatologist. This part of the work aims to prepare a fully automated system that is incorporated with an explanation that makes this skill-required job understandable for non-experts. Therefore, two different approaches are proposed, namely Geometry-based and deep learning-based. In the first approach, skin lesions are manually analyzed based on a geometric pattern and classified as either symmetric or asymmetric (including border, structure, and color). This will describe the lesion shape so that this approach can be applied to any dermoscopic image that has no ground truth of asymmetry. In the second approach, skin lesions are automatically analyzed with several pre-trained Convolutional Neural Network (CNN) feature extractors (i.e., ResNet18, ResNet50, and ResNet101) to train a multi-layer Support Vector Machine (SVM) image classifier. This approach will provide immediate automatic decisions about lesion asymmetry without any description. The input image sources are three publicly available datasets such as PH2 (Mendonca et al., 2013), MED-NODE (Giotis et al., 2015), and ISIC2016 (Gutman et al., 2016). The experimental results of these three CNN models are recorded for a comparative analysis to identify the best-performing model. Since the second approach is deep learning-based, it requires a

significant amount of labeled data to train the deep learning model properly. In this case, finding a good number of labeled data is a difficult task. To overcome this issue, the first approach can provide labeled data (ground truth) that is determined by geometric analysis. Figure 5.3 shows the overall proposed system.
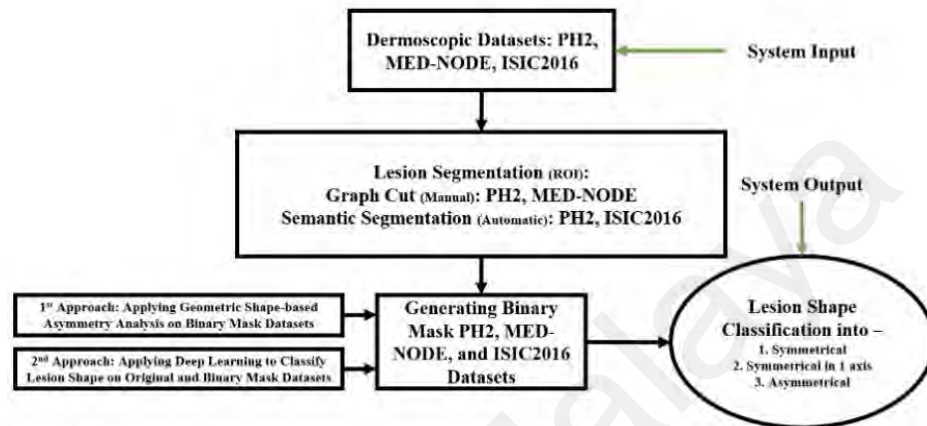
Figure 5.3: The overall proposed system to measure lesion asymmetry.

5.2.1 Image Data Acquisition for Asymmetry Analysis

The input images went through the segmentation process (to determine the region of interest) before applying the proposed approaches. The disease-affected area was kept separate from the unaffected area during this process. A binary mask (an image with white and black pixels where white pixels indicate a lesion or region of interest, while the black pixels indicate a healthy or non-disease-infected area) is created once the lesion is isolated. The input of the proposed approach is a binary mask (refer to Figure 4.1 of Chapter 4: Lesion Segmentation). The binary mask indicates a disease-affected area. Two different techniques, such as Graph Cut (manual) and Semantic Segmentation (automatic), are used to construct binary masks. Here, the manual process of Graph Cut is applied to generate binary masks. Because the MED-NODE dataset has no ground truth of binary masks (disease-affected area). Integrating both manual and automatic processes to construct binary masks can offer several advantages such as flexibility, customization,

error handling, and complex decision. Figure 5.4 presents the different steps of this research.
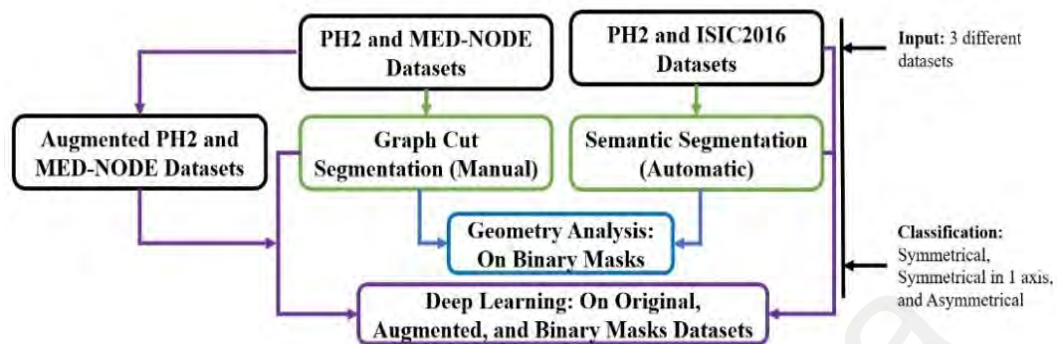
Figure 5.4: The architecture of the asymmetry analysis system.

5.2.1.1 Dataset for Lesion Segmentation

Two datasets PH2 and MED-NODE are used for the Graph Cut segmentation process. Using the ISIC2016 dataset is a time-consuming process for the manual Graph Cut segmentation process. For the semantic segmentation process, two datasets are used which are PH2 and ISIC2016. The PH2 and ISIC2016 had the ground truth about the disease-affected area in the binary mask which is required to train the semantic segmentation networks. MED-NODE is a non-annotated small dataset. Since there is no ground truth in MED-NODE, it is not possible to use it for training the semantic segmentation networks. Therefore, manually establishing the ground truth (segmented image in the binary mask) for the MED-NODE dataset is required. In this case, the Graph Cut (Chen & Bagci, 2011) method is the best choice for segmenting the lesion manually.

5.2.1.2 Dataset for Lesion Shape Classification

The PH2, MED-NODE, and ISIC2016 datasets are used for the symmetrical properties analysis of skin lesions. Only the PH2 is annotated with symmetrical information. The PH2 dataset is developed by Hispano Pedro Hospital, Spain. This is the only publicly

available annotated dataset where skin lesions' shapes are classified into three different classes such as symmetrical, partial-symmetrical, and asymmetrical. The PH2 dataset has the ground truth of the segmented area. The dataset contains 117 symmetrical, 31 partials symmetrical, and 52 asymmetrical images. These 200 images are used for the first approach of this research which is called geometrical shape-based asymmetric analysis (GSAA). For the second approach, which is deep learning-based asymmetric analysis of this research, an augmented dataset is created from the PH2 dataset. The purpose of data augmentation in deep learning is to expand the training dataset by applying various transformations (such as rotation, flipping, cropping, and scaling) to the original images, thereby improving the model's ability to generalize, reduce overfitting, and enhance its performance on real-world data. Here, different techniques such as rotation (5, 10, 90, 180, and 270-degree), adjusting brightness, and flipping are applied to make 1585 images (571 symmetrical, 496 partial-symmetrical, and 520 asymmetrical) from 200 images in the augmented dataset. This augmented PH2 dataset is used to train an image category classifier.

The MED-NODE dataset, utilized by the Department of Dermatology at the University Medical Center Groningen to develop and test the MED-NODE system for skin cancer detection from macroscopic images, comprises non-dermoscopic and non-annotated images (Giotis et al., 2015). This MED-NODE dataset consists of 70 Melanoma and 100 naevus images from the digital image archive. First, the Graph Cut method is applied to this dataset. Due to the lack of relevant lesion ground truth, semantic segmentation is not applicable in this case. Then, the proposed GSAA technique is applied to the segmented images of this dataset. Then, all images of this dataset are divided into three classes like the PH2 dataset. Again, another augmented dataset is created from the MED-NODE dataset. The same augmentation techniques that are used in the PH2 dataset are applied in the MED-NODE dataset to increase the number of images. From 170 images (58

symmetrical, 19 partial-symmetrical, and 93 asymmetrical) to 354 images (120 symmetrical, 114 partial-symmetrical, and 120 asymmetrical) are used to create this second augmented dataset. This dataset also is used for training purposes like the first augmented dataset.

Again, another dermoscopic dataset ISIC2016 is used with the ground truth of segmented area but not with the ground truth of symmetric properties. It has 1269 lesion images with the segmented lesion binary mask. This dataset is not used for the Graph Cut method. Later ISIC2016 and PH2 datasets were used to train the semantic segmentation network. The Semantic segmentation network is an unsupervised skin lesion segmentation technique. To segment the region of interest (ROI) from manual (Graph Cut) to automatic (deep learning), a newly designed semantic segmentation network is trained. As previously discussed in Chapter 4: Skin Lesion Segmentation, a new network is designed for two classes (background and lesion) because the traditional semantic segmentation network is complex to implement, and training process is time consuming due to their deep architectures. Besides, those state-of-arts models were designed for multiple classes (more than two classes), which require powerful system to run.

After the segmentation process, the GSAA is applied to the ISIC2016 dataset to classify lesion-shape properties. In the analysis process, 700 symmetrical, 300 partial-symmetrical, and 269 asymmetrical lesions are found. Furthermore, all 1269 images of this dataset are used to train an image classifier for the categorization of lesion shapes. In this dataset, an augmentation process is not required like PH2 and MED-NODE to increase the training data. For the binary classification (two classes: background and lesion), a satisfactory classification accuracy rate is found in the training process using 1269 images of ISIC2016 dataset. Therefore, this dataset does not need any augmentation process to increase training data. Table 5.1 shows the utilization of different datasets on

the geometry based and deep learning-based lesion shape analysis, and here (×) means "No", and (√) means "Yes".

Table 5.1: The utilization of different datasets on different approaches.

Approaches / Datasets		PH2	Augmented PH2	PH2 Mask	MED-NODE	Augmented MED-NODE	MED-NODE Mask	ISIC2016	ISIC2016 Mask
Geometrical Shape Analysis	Semantic Segmentation	√	x	x	x	x	x	√	x
	Graph Cut	√	x	x	√	x	x	x	x
Deep Learning Shape Analysis		x	√	√	x	√	√	√	√

5.2.2 First Approach: Geometry Shape-Based Asymmetry Analysis (GSAA) for Skin Lesion

To analysis skin lesion shape, a new novel approach is created called the GSAA. Input for the GSAA is a single skin lesion image. Like most skin lesion images, the images used in this research are also in RGB format, without any deviation. Before GSAA, the input images are segmented using the Graph Cut (Peng et al., 2010) and Semantic Segmentation Network (Guo et al., 2019). In this experiment, both dermoscopy and non-dermoscopy (macrophotography) images are used. Figure 5.5 shows the working procedure of the GSAA.
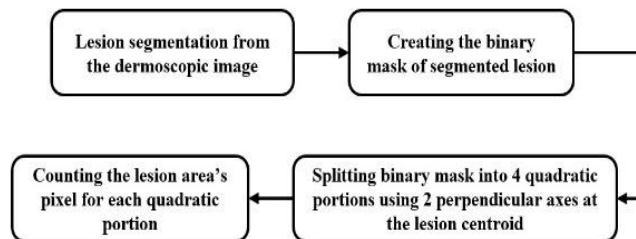
Figure 5.5: The sequential steps of the GSAA.

5.2.2.1 Creating Binary Mask with Graph Cut Segmentation

To separate foreground and background elements in an image, utilize the semiautomatic segmentation method Graph Cut (Chen & Bagci, 2011). Good initialization is not necessary for Graph Cut segmentation. Scribbles are lines painted on the image to indicate what should be in the foreground and what should be in the background. The Image Segmenter in Matlab 2022a automatically divides an image into segments based on predetermined scribbles, and it then shows the segmented image. More scribbles are added to the image to further enhance the segmentation until it produces acceptable results. For quick segmentation, the Graph Cut approach integrates graph theory to image processing. The method builds a graph of the image, with each pixel acting as a node and each edge being a weighted connection. The weight increases as the likelihood of pixel relationships increases. The program segments the items in the image by cutting along frail edges. A particular variety of the Graph cut algorithm called lazy snapping (segmenting image into foreground and background using graph-based segmentation) is used in this process. The first step is to create a background on the lesion images where skins are non-infected or healthier. The second step is to create a region of interest (foreground) on the lesion images where skins are infected. In this process, the pixels of the healthy area are selected to become the background, while the pixels of the unhealthy area are chosen to become the region of interest. By applying these two masks, the skin lesion (infected) is isolated from the background (non-infected skin) where the background pixel is 0. Then, another mask is created based on the background area and the lesion area (pixel value is 1) which is a binary image. Figure 5.6 shows the result after applying masks to the lesion image. The achieved binary image is stored to create a different dataset. Two new datasets, namely the PH2 mask dataset (200 lesion binary images) and the MED-NODE mask dataset (170 lesion binary images), are created by storing the binary mask images of the PH2 and MED-NODE datasets. Figure 5.7 shows

several binary images of both new datasets. The performance of the Graph Cut method was remarkable (weighted intersection-over-union is 91.13%) which evaluated by the equations (3) to (5) on PH2 and MED-NODE datasets in Table 5.2. Since MED-NODE did not have annotated data on ROI (lesion), the comparison result is not available. Due to the limitation of time, using graph cut was avoided on the ISIC2016 dataset (1269 images).
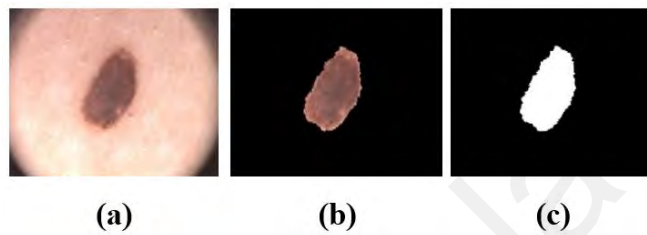
Figure 5.6: (a) Original image, (b) Color pixels represent ROI and black pixels represent the healthy area, and (c) Binary image where black pixels represent the healthy area, and white pixels represent the unhealthy area.

Figure 5.7: (a) PH2 mask dataset, and (b) MED-NODE mask dataset.

Table 5.2: Performance evaluation of Graph Cut on PH2 and MED-NODE datasets.

Dataset	*mIoU*	*wIoU*
PH2	0.84784	0.91132
MED-NODE	-	-

Note. MED-NODE has no annotated information about lesion mask or ROI.

117

5.2.2.2 Creating Binary Mask with a Simple Semantic Segmentation

A semiautomatic method is the graph cut method. By utilizing a straightforward semantic segmentation network, this semiautomatic method can become totally automatic. A semantic segmentation network divides an image into classes by classifying each pixel. Dilated convolutions, also referred to as atrous convolutions, are frequently used in semantic segmentation networks like DeepLab (Chen et al., 2018), as they increase the receptive field of the layer (the area of the input that the layers can see) without increasing the number of computation parameter.
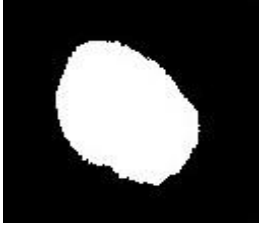
This method was used on the PH2 dataset after the 128-by-128 lesion picture dataset from ISIC2016. Ground truth data for the pixels in these datasets are also present. A pixel-label datastore and an image datastore are both used to load the training data (900 images from ISIC2016). There is a generated an image datastore for the images. The ground truth pixel labels are put into a datastore for pixel labels. In this study, we employed a straightforward semantic segmentation network with the top semantic segmentation performer, the Tversky Loss Function (see Chapter 4: Skin Lesion Segmentation). For training data, a data source was constructed, and each label's pixel count was obtained. The background has the most pixel labels. The learning process is biased in favor of the dominant class because of this class imbalance. Class weighting was employed to balance the courses to correct this. It is possible to calculate class weights using a variety of techniques. Inverse frequency weighting is a popular technique in which the class weights are equal to the inverse of the class frequencies. This approach gives underrepresented classes more weight. Inverse frequency weighting was used to determine the class weights. Using an image input layer with an input size proportional to the size of the input images, a network was developed for the classification of pixels. The next three blocks were convolution, batch normalization, and ReLU layers. 32 3-by-3 filters with progressively larger dilation factors were defined for each convolutional layer, and the

'Padding' option was set to'same' to pad the inputs to the same size as the outputs. K 1-by-1 convolutions were used in a convolutional layer with a softmax layer, a bespoke pixel classification layer, and a Tversky Loss Function layer to classify the pixels. K was the number of classes. There were designated training alternatives. Finally, the network received training. A total of 369 ISIC2016 images served as the test data. For the test Images, an image datastore was made. The ground truth pixel labels were put into a datastore for pixel labels. We produced predictions using the trained network and test data. The prediction accuracy was evaluated using equations (1) to (6) in Table 5.3 for ISIC2016 and PH2 datasets. Since the PH2 mask dataset was already created using Graph Cut, only the ISIC2016 mask dataset was created from the semantic segmentation process. Graph cut was not applied to the ISIC2016 dataset to create the ISIC2016 mask dataset. Graph cut is a manual and time-consuming process. Applying Graph cut to the 1269 images of the ISIC2016 dataset would be a time-consuming process. Finally, two predicted outputs of ISIC2016 and PH2 datasets were compared with the ground truth of their input images in Table 5.4.

Table 5.3: Performance evaluation of Semantic Segmentation on ISIC2016 and PH2 datasets.

Dataset	*gAC*	*mAC*	*mIoU*	*wIoU*	*mBF*
ISIC2016	0.90462	0.93981	0.74478	0.84686	0.71952
PH2	0.88456	0.91852	0.73045	0.83147	0.71058

Table 5.4: Visual comparison between semantic segmentation's outputs and ground truth of different datasets.

	Input	Ground Truth	Predicted Area
ISIC2016			
PH2			

5.2.2.3 Lesion Identification: Symmetrical or Asymmetrical

This section covers the process of analyzing the symmetrical properties of dermatological lesions, determining whether a lesion is completely symmetrical, symmetrical along one axis, or asymmetrical. Three mask datasets (PH2 mask, MED-NODE mask, and ISIC2016 mask datasets) were acquired from the two segmentation methods. The proposed GSAA method was applied to these mask datasets. The binary (mask) image was divided into 4 portions using two perpendicular axes at the lesion centroid. As a result, 4 different split images were found in 1 single image. Figure 5.8 shows the dividing process of the binary image. Those 4 split images are considered A (segmented area: right side bottom corner), B (segmented area: left side bottom corner), C (segmented area: left side up corner), and D (segmented area: right side up corner).
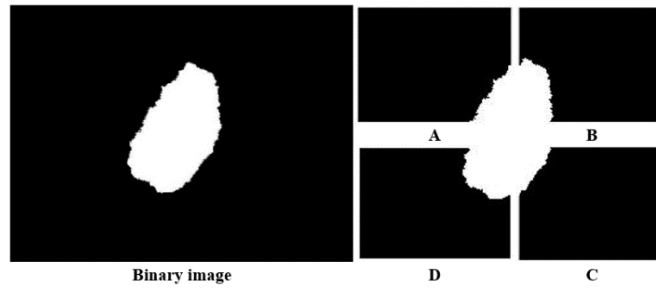
Figure 5.8: The binary image is divided into 4 split images.

For A, B, C, and D, the number of white pixels is counted separately. Each pixel is considered a unit of the white area. Then, their total white pixel numbers are divided from each other to compare the area. The following condition is followed-

Ap/Cp = 0 or 1, Bp/Dp = 0 or 1, Ap/Bp = 0 or 1, and Cp/Dp = 0 or 1.

Here, Ap, Bp, Cp, and Dp mean the number of white pixels in the split images A, B, C, and D. If a minimum of one quotient and a maximum of three quotients of these four operations become 1, then the lesion will be Symmetric in 1 axis. If all quotients of these four operations become 1, then the lesion will be Symmetric. If all quotients of these four operations become 0, then the lesion will be Asymmetric. Usually, the skin lesions do not have a smooth border and exact symmetrical shape including color and structure (Ng et al., 2005). Naturally, the lesion shapes are unsmooth and irregular. Therefore, any two split images are not expected to have the same pixel number of white areas. There might be a fraction of the quotient when we divide the number of the white pixels of 2 split images by each other. Therefore, three conditions are needed to form based on arithmetic assumption. Those are-

If, $\dfrac{any\ split\ image's\ total\ white\ pixel}{another\ split\ image's\ total\ white\ pixel} = 0.90\ to\ 1.10\ (it\ will\ be\ considered\ as\ 1)$

If, $\dfrac{any\ split\ image's\ total\ white\ pixel}{another\ split\ image's\ total\ white\ pixel} < 0.90\ (it\ will\ be\ considered\ as\ 0)$

$$\text{If, } \frac{any\ split\ image's\ total\ white\ pixel}{another\ split\ image's\ total\ white\ pixel} > 1.10\ (it\ will\ be\ considered\ as\ 0)$$

Based on the above conditions, lesion images are analyzed in PH2 dataset. The algorithm of the above conditions is presented below-

---

Step 1: *Count the white pixels in binary images A, B, C, and D.*
- *Let **Ap** be the count of white pixels in image A.*
- *Let **Bp** be the count of white pixels in image B.*
- *Let **Cp** be the count of white pixels in image C.*
- *Let **Dp** be the count of white pixels in image D.*

Step 2: *Store the white pixel counts in **Ap**, **Bp**, **Cp**, and **Dp**.*

Step 3: *Calculate the following quotients:*
- ***quotient1** = Ap / Cp*
- ***quotient2** = Bp / Dp*
- ***quotient3** = Ap / Bp*
- ***quotient4** = Cp / Dp*

Step 4: *Check if any of the quotients are close to 1:*
- *If **0.9 ≤ quotient1 ≤ 1.1**, return 1.*
- *If **0.9 ≤ quotient2 ≤ 1.1**, return 1.*
- *If **0.9 ≤ quotient3 ≤ 1.1**, return 1.*
- *If **0.9 ≤ quotient4 ≤ 1.1**, return 1.*
- *Otherwise, return 0.*

Step 5: *Calculate the sum of all quotients:*
- ***Q** = quotient1 + quotient2 + quotient3 + quotient4*

Step 6: *Determine the symmetry of the skin lesion based on the value of **Q**:*
- *If **Q == 4**, the skin lesion is symmetric.*
- *If **1 ≤ Q ≤ 3**, the skin lesion is half symmetric.*
- *If **Q == 0**, the skin lesion is asymmetric.*

Table 5.5 shows the asymmetrical analysis results for 5 samples out of 200 samples. These analytical results match with the ground truth at 99.00% as presented in Table 5.6 as a confusion matrix (CM). The CM is a graphical presentation of predicted value vs. ground truth value. The GSAA approach misclassified 2 out of 200 images due to the presence of artifacts and incompleteness (part of the lesion captured). In Table 5.5, (×) means "No", and (√) means "Yes".

Table 5.5: Geometry-based asymmetry analysis on the PH2 dataset.

Sample No.	Ap/Cp	Bp/Dp	Ap/Bp	Cp/Dp	Sym.	Sym. in 1 axis	Asym.	Ground Truth
	1	1	1	1	√	×	×	Sym.
	0	0	0	0	×	×	√	Asym.
	0	0	1	1	×	√	×	Sym. in 1 axis
	1	1	1	1	√	×	×	Sym.
	0	0	0	0	×	×	√	Asym.
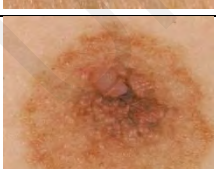
Table 5.6: The confusion matrix (CM) of the geometry-based asymmetry analysis vs. the ground truth (PH2).

		Ground Truth		
		Symmetrical	Sym. in 1 axis	Asymmetrical
GSAA's Prediction	Symmetrical	116	1	0
	Sym. in 1 axis	1	30	0
	Asymmetrical	0	0	52

The GSAA technique is applied on a non-dermoscopic (images are taken without a dermoscope device) image dataset where the ground truth on dermatological asymmetric information is absent. To demonstrate the flexibility of this algorithm on other datasets, the non-annotated MED-NODE dataset is used which contained 170 macroscopic skin lesion images with 2 classes (70 Melanoma, and 100 Naevus). After applying the GSAA techniques, 93 asymmetrical, 19 symmetrical in one axis, and 58 symmetrical images were found. Table 5.7 shows the asymmetrical analysis results for 5 samples out of 170 samples. The analytical results are applied to create three new different classes (such as symmetry, symmetry in 1 axis, and asymmetry) on the MED-NODE dataset as a new

annotated information. This new annotated information of MED-NODE dataset is the evidence that supports the success of proposed algorithm and the partial contribution of this research. However, it is important to note that even though the labeling of the MED-NODE dataset is carried out using the proposed GSAA approach, additional clinical verification is necessary to ensure the accuracy and reliability of the data labeling. In Table 5.7, the last column is "Ground Truth", and "-" means no "Ground Truth" available regarding lesion shape for MED-NODE.
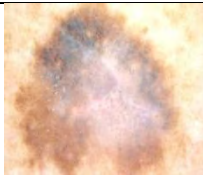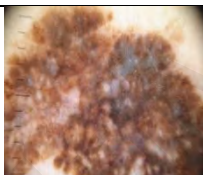
Table 5.7: Geometry-based asymmetry analysis on the MED-NODE dataset.

Sample No.	Ap/Cp	Bp/Dp	Ap/Bp	Cp/Dp	Sym.	Sym. in 1 axis	Asym.	Ground Truth
	1	1	1	1	√	×	×	-
	0	0	1	1	×	√	×	-
	0	0	0	0	×	×	√	-
	1	1	1	1	√	×	×	-
	1	1	1	1	√	×	×	-

Again, the proposed technique is applied to another dermoscopic dataset called ISIC2016. This dataset contains 1269 skin lesion images without dermatological asymmetric information like as MED-NODE dataset. Also, there is no disease

classification information except the area of the lesion (in the binary mask). Several 700 symmetrical, 300 symmetrical in one axis, and 269 asymmetrical lesion images are found after applying the GSAA on the ISIC2016 dataset. Another indication of the GSAA's accomplishment and a contribution to the ISIC2016 dataset is the division of this dataset into three classes. These contributions are not clinically approved but geometrically (shape analysis in a mathematical way) analyzed and proved. Table 5.8 shows the asymmetrical analysis results for 5 samples out of 1269 samples. Also, in this table, the last column is "Ground Truth", and "-" means no "Ground Truth" available regarding lesion shape for ISIC2016.

Table 5.8: Geometry-based asymmetry analysis on the ISIC2016 dataset.

Sample No.	Ap/Cp	Bp/Dp	Ap/Bp	Cp/Dp	Sym.	Sym. in 1 axis	Asym.	Ground Truth
	0	0	0	0	×	×	√	-
	1	1	0	0	×	√	×	-
	0	0	0	0	×	×	√	-
	0	0	1	1	×	√	×	-
	1	1	1	1	√	×	×	-

5.2.3 Second Approach: Classification of Lesion Shape

A Convolutional Neural Network (CNN) is an effective deep learning machine learning method. The training of CNNs uses extensive datasets of various image types. CNNs can learn rich feature representations for a variety of images from these sizable databases. In many cases, these feature representations perform better than manually created features like HOG, LBP, or SURF. Using a pre-trained CNN as a feature extractor is a simple approach to take advantage of CNN power without putting too much time and effort into training. CNN should be used to classify images because they can automatically learn and extract meaningful features from the skin lesion image data, leading to highly accurate and efficient image classification. This experiment employs a pre-trained CNN as a feature extractor to train an image category classifier.

5.2.3.1 Transfer Learning Approach to Classify Lesion Shape

Here, images from the PH2, MED-NODE, and ISIC2016 datasets are divided into three symmetrical groups using a multiclass linear SVM trained with CNN features derived from the images (symmetrical, symmetry in 1 axis, and asymmetrical). This method of classifying skin lesions by image category involves using features extrapolated from the lesion images to train an existing classifier. Then, features are extracted using CNN rather than picture features like HOG or SURF. These datasets were adjusted first to balance the number of images in the training set because there were different numbers of images in each category in these datasets. Pre-trained networks have become increasingly prevalent. The majority of these were developed using the ImageNet dataset, which includes 1.2 million training images and 1000 object categories (Fei-Fei et al., 2010). This experiment makes use of three such models: ResNet18, ResNet50, and ResNet101 (He et al., 2016). Aside from AlexNet, GoogLeNet, VGG16, and VGG19, other well-known networks trained on ImageNet include these. Because ResNet was

trained on a sizable dataset, it can capture rich and discriminative picture representations that can be used for a variety of image analysis applications with better performance and efficiency. The input dimensions are specified in the first layer. The required input size varies for each CNN. The image input for the ResNet series must be 224 × 224 by 3. The primary layers of CNN are the intermediate layers. There are several convolutional layers there, as well as max-pooling and rectified linear units (ReLU) layers (Krizhevsky et al., 2017). Three fully connected layers follow the layers. The classification layer is the top layer, and the features of this layer rely on the classification problem. The CNN model used in this example was initially trained to handle a 1000-way classification task. 1000 classes from the ImageNet dataset make up the classification layer as a result. It is important to note that the CNN model utilized in this study won't be applied to the initial classification task. It will instead be used to address a separate classification problem on the lesion dataset. With 70% of the images from each dataset used as training data and the remaining 30% used as validation data, the datasets are divided into training and validation data, respectively. To prevent biasing the outcomes, the split is randomly generated. After that, the CNN model processes the training and test sets.

As previously stated, the ResNet series can only analyze RGB images that are 224 by 224. The usage of an augmented image datastore allows for on-the-fly resizing and conversion of any grayscale images to RGB, avoiding the need to save all the images again in this format. When employed for network training, the enhanced picture datastore can also be used for extra data augmentation. A CNN layer each produces an activation in response to an input image. However, only a few layers of a CNN are appropriate for extracting visual features. The network's initial layers record fundamental visual characteristics like edges and blobs. This can provide a better understanding of why CNN features are so effective for image recognition tasks. It is important to note that the first layer of a network has developed filters for capturing blob and edge features. Deeper

network layers process these 'basic' characteristics after that, combining the early features to create higher-level picture features. Given that they incorporate all the elementary features into a fuller visual representation, these higher-level features are more suited for recognition tasks (Donahue et al., 2014). The activations approach makes it simple to retrieve features from one of the deeper layers. It is up to the designer to decide which deep layer to start with, however in general, it is wise to begin with the layer immediately preceding the categorization layer. This layer is known as "fc1000" in the pre-trained model. Using that layer, training features are extracted. After doing an ablation study, the 'MiniBatchSize' is set to 32 during the training process to guarantee that the CNN and image data will fit in GPU memory. Because the GPU's RAM was insufficient, the 'MiniBatchSize' required to be minimal. Also organized as columns is the activation output. This accelerated the multiclass linear SVM training that came next. Next, a multiclass SVM classifier is trained using the CNN image characteristics. The 'Learners' parameter of the fitcecoc function is set to 'Linear' for training, which employs a quick stochastic gradient descent solver. When using high-dimensional CNN feature vectors, this speeds up the training process.

5.2.3.2   Evaluation of the Lesion Shape Classifier

The procedure used for the training set is repeated to extract image features from the test set. The test features are then passed to the classifier to measure the accuracy of the trained classifier. Figure 5.9 to Figure 5.14 show the CMs for ResNet18, ResNet50, and ResNet101 on the augmented PH2, augmented MED-NODE, PH2 mask, MED-NODE mask, ISIC2016, and ISIC2016 mask datasets. From these CMs, different evaluation metrics are calculated using equations (7) to (16) in Table 5.9 to Table 5.14 for the three ResNet models.

```
confMat =                    confMat =                    confMat =

   0.6577    0.2416    0.1007      0.9329    0.0134    0.0537        0.7315    0.1611    0.1074
   0.0671    0.9060    0.0268      0.3289    0.5906    0.0805        0.0940    0.8993    0.0067
   0.0872    0.1141    0.7987      0.0738    0.0067    0.9195        0.0470    0.0201    0.9329
```

ResNet18   **ResNet 50**   **ResNet 101**

Figure 5.9: The CM for ResNet18, ResNet50, and ResNet101 on the augmented PH2 dataset.

Table 5.9: Performance evaluation (%) of the different ResNet models on the augmented PH2 dataset.

Pretrained Model	Symmetrical					Symmetrical in 1 Axis					Asymmetrical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	wSE	wSP	wPR	wAC	wF1
ResNet 18	80	84	66	83	73	72	94	**90**	85	80	86	90	80	89	83	79	89	78	86	79
ResNet 50	70	**96**	**93**	84	**80**	**97**	83	59	86	73	87	**96**	92	93	89	84	92	82	88	81
ResNet 101	84	87	73	**86**	78	83	**95**	90	90	**87**	89	96	93	94	91	85	93	85	90	85

```
confMat =                    confMat =                    confMat =

   0.7000    0.1000    0.2000      0.8667    0.0667    0.0667        0.9333    0.0667         0
   0.1333    0.8667         0           0    1.0000         0        0.1000    0.9000         0
   0.1333    0.0667    0.8000      0.2333    0.2000    0.5667        0.0333    0.0667    0.9000
```

ResNet 18   **ResNet 50**   **ResNet101**

Figure 5.10: The CM for ResNet18, ResNet50, and ResNet101 on the augmented MED-NODE dataset.

Table 5.10: Performance evaluation (%) of the different ResNet models on the augmented MED-NODE dataset.

Pretrained Model	Symmetrical					Symmetrical in 1 Axis					Asymmetrical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	wSE	wSP	wPR	wAC	wF1
ResNet 18	73	85	70	81	71	84	93	87	89	85	80	90	80	87	80	79	89	79	86	79
ResNet 50	79	93	86	88	83	79	**100**	**100**	91	**88**	89	82	57	83	70	82	92	81	87	80
ResNet 101	**88**	**96**	**93**	**93**	**90**	87	95	90	**92**	**88**	**100**	95	90	97	95	92	95	91	94	91

```
confMat =

    0.6983    0.2672    0.0345
         0    0.9831    0.0169
         0    0.0943    0.9057
```
```
confMat =

    0.8966    0.0603    0.0431
    0.1271    0.8559    0.0169
    0.0660    0.0283    0.9057
```
```
confMat =

    0.8017    0.1293    0.0690
    0.0424    0.9407    0.0169
         0    0.0566    0.9434
```

ResNet 18   **ResNet 50**   **ResNet101**
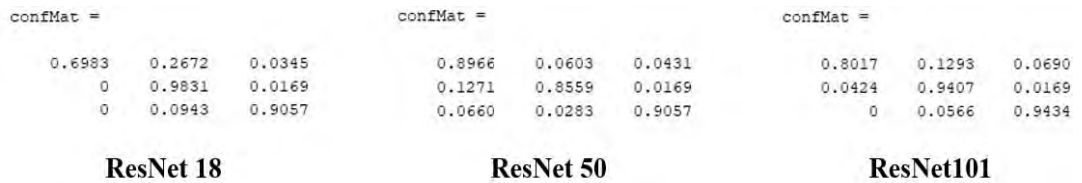
Figure 5.11: The CM for ResNet18, ResNet50, and ResNet101 on the augmented MED-NODE dataset.

Table 5.11: Performance evaluation (%) of the different ResNet models on the ISIC2016 dataset.

Pretrained Model	Symmetrical					Symmetrical in 1 Axis					Asymmetrical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	wSE	wSP	wPR	wAC	wF1
ResNet 18	**100**	87	70	90	82	73	**99**	**98**	87	84	**95**	96	91	**95**	**93**	89	94	86	91	86
ResNet 50	82	**95**	**90**	90	86	**91**	93	86	**92**	**88**	94	96	91	**95**	92	89	**95**	**89**	92	**89**
ResNet 101	95	91	80	**92**	**87**	83	97	94	**92**	**88**	91	**97**	**94**	95	93	**90**	95	89	93	89

```
confMat =

    0.8713    0.1053    0.0234
    0.1409    0.8591         0
    0.1154    0.0321    0.8526
```
```
confMat =

    0.8889    0.0702    0.0409
    0.2081    0.7852    0.0067
    0.0577    0.0321    0.9103
```
```
confMat =

    0.8129    0.1111    0.0760
    0.1074    0.8591    0.0336
    0.0449    0.0064    0.9487
```

ResNet 18   **ResNet 50**   **ResNet101**
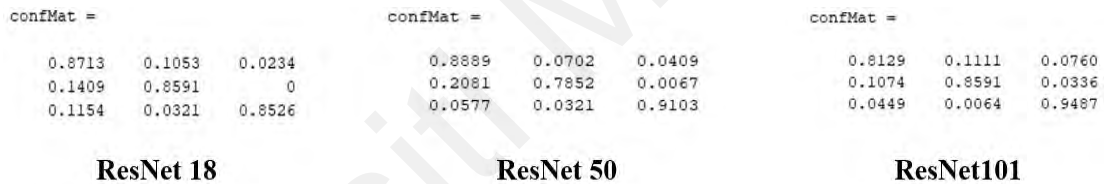
Figure 5.12: The CM for ResNet18, ResNet50, and ResNet101 on the PH2 mask dataset.

Table 5.12: Performance evaluation (%) of the different ResNet models on the PH2 mask dataset.

Pretrained Model	Symmetrical					Symmetrical in 1 Axis					Asymmetrical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	wSE	wSP	wPR	wAC	wF1
ResNet 18	77	93	87	87	82	86	**93**	86	91	86	**98**	93	85	94	91	84	93	86	90	85
ResNet 50	77	**94**	**89**	88	**83**	89	90	79	90	84	95	96	91	**95**	**93**	84	**94**	**88**	90	**86**
ResNet 101	84	91	81	**89**	**83**	88	**93**	86	91	87	90	**97**	**95**	95	92	**86**	93	86	**91**	**86**

```
confMat =                        confMat =                          confMat =

    1.0000         0         0       0.8889    0.0833    0.0278       0.9167         0    0.0833
         0    0.9706    0.0294            0    1.0000         0            0    0.9706    0.0294
    0.1667    0.0556    0.7778       0.0278    0.1389    0.8333       0.0833         0    0.9167
```

ResNet 18   **ResNet 50**   **ResNet 101**
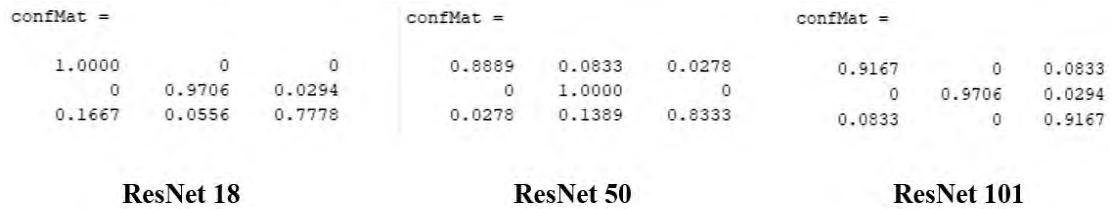
Figure 5.13: The CM for ResNet18, ResNet50, and ResNet101 on the MED-NODE mask dataset.

Table 5.13: Performance evaluation (%) of the different ResNet models on the MED-NODE mask dataset.

Pretrained Model	Symmetrical					Symmetrical in 1 Axis					Asymmetrical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	wSE	wSP	wPR	wAC	wF1
ResNet 18	85	**100**	**100**	94	92	95	98	97	97	96	96	90	78	92	86	92	94	88	93	89
ResNet 50	**97**	95	89	**95**	**93**	100	100	100	100	100	**97**	92	83	93	89	**97**	94	87	**95**	**92**
ResNet 101	92	96	92	**95**	92	100	99	97	99	98	89	**96**	**92**	**94**	**91**	91	**96**	**93**	**95**	**92**

```
confMat =                        confMat =                          confMat =

    0.8362    0.0517    0.1121       0.8718         0    0.1282       0.9744    0.0256         0
    0.0339    0.9068    0.0593       0.0256    0.9487    0.0256       0.0769    0.9231         0
         0         0    1.0000            0         0    1.0000       0.0571         0    0.9429
```

ResNet 18   **ResNet 50**   **ResNet 101**
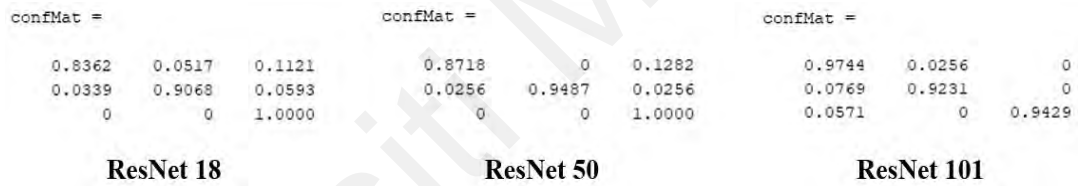
Figure 5.14: The CM for ResNet18, ResNet50, and ResNet101 on the MED-NODE mask dataset.

Table 5.14: Performance evaluation (%) of the different ResNet models on the ISIC2016 mask dataset.

Pretrained Model	Symmetrical					Symmetrical in 1 Axis					Asymmetrical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	wSE	wSP	wPR	wAC	wF1
ResNet 18	96	92	84	94	90	95	96	91	95	93	85	100	100	94	92	92	96	92	94	92
ResNet 50	97	94	87	**95**	**92**	100	98	95	98	97	86	**100**	**100**	95	93	94	**97**	**94**	**96**	**94**
ResNet 101	87	**98**	**97**	94	**92**	97	96	92	96	94	**100**	97	94	**98**	**97**	**95**	97	94	**96**	**94**

The number in the models' name, such as ResNet18, ResNet50, and ResNet101, represents the number of layers in the network. The reason ResNet101 generally has better performance on these tables (Table 5.9 to Table 5.14) compared to ResNet50 and

ResNet18 lies in the depth of the networks. Deeper networks can capture more complex patterns and representations, allowing them to learn more intricate features from the input data. This added depth enables the model to learn hierarchical representations of the data, leading to improved performance. On the other hand, ResNet50 is deeper than ResNet18, with 50 layers. The additional depth allows it to capture more intricate features and patterns, making it perform better on these tables (Table 5.9 to Table 5.14) with a wide range of tasks compared to ResNet18.

After these analytical reports are in Table 5.9 to Table 5.14, Receiver Operation Characteristic (ROC) curves are drawn from ResNet series performance (sensitivity and specificity) on the different datasets. It is a graphical representation and evaluation metric commonly used in machine learning and statistics to assess the performance of binary classification models. The ROC curve illustrates the trade-off between the true positive rate (TPR or sensitivity) and the false positive rate (FPR) for different threshold values. Figure 5.15 shows the ROC of the ResNet series on the six datasets. In Table 5.15, the area under the curve (AUC) is calculated from the ROC curves of Figure 5.15.
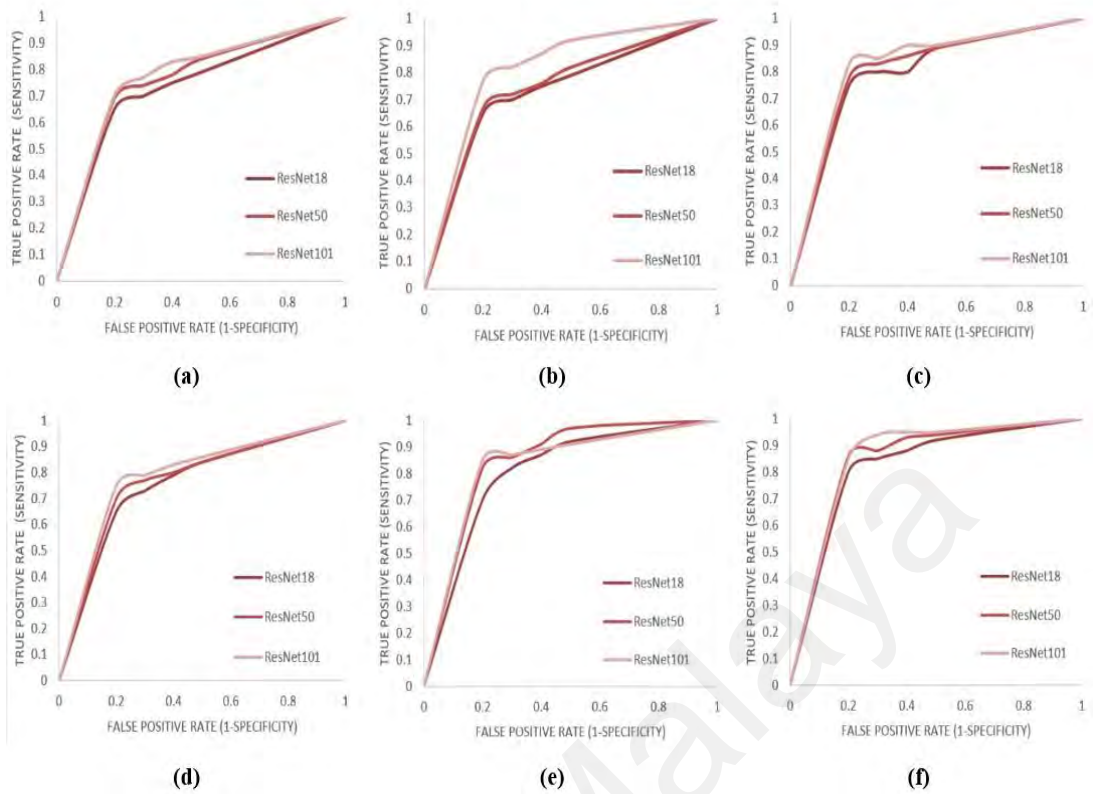
Figure 5.15: The ROC curves of ResNet series on the (a) augmented PH2, (b) augmented MED-NODE, (c) ISIC2016, (d) PH2 mask, (e) MED-NODE mask, and (f) ISIC2016 mask dataset.

Table 5.15: Measuring the AUC of the ResNet series on the different datasets.

		Datasets (AUC)					
		Augmented PH2	Augmented MED-NODE	ISIC2016	PH2 Mask	MED-NODE Mask	ISIC2016 Mask
Models	ResNet18	0.78	0.78	0.81	0.80	0.83	0.82
	ResNet50	0.82	0.80	0.83	0.83	**0.88**	0.84
	ResNet101	**0.83**	**0.85**	**0.84**	**0.85**	0.87	**0.86**

Based on Table 5.15, ResNet101 has the highest AUC of the five datasets and ResNet50 has the other. ResNet18 has a lower area comparatively on all datasets. Better results were found when ResNet101 pre-trained model was used as the extractor of dermatological asymmetric shape to feed the multiclass SVM lesion image classifier.

5.2.4   Comparative Discussion for Asymmetry Analysis for Skin Lesion

Two new methods are reported in this section. One is geometrical-based, and another is a deep learning-based dermatological shape classification. At present, to the best of our knowledge, there is no similar approach found in the literature to compare this asymmetrical analysis with the proposed methods.

However, some existing methods' achieved accuracy and experimental datasets with the number of images are observed. Table 5.16 shows the results of different approaches to compare with the proposed methods. Several random dermoscopic images from ISIC2018 (Tschandl et al., 2018) are used in (Ali et al., 2020) where asymmetry, color variegation, and diameter features are extracted. Only asymmetry features are given focus in this comparative study. Kullback-Leibler divergence technique results are reported as the proposed methods' evaluation metrics in (Chakravorty et al., 2016). Then, a feed-forward neural network with Levenberg-Marquardt Backpropagation training algorithm-based experimental results is reported as the regression coefficient (R) and mean squared error (MSE) in (Damian et al., 2021). The dermatological asymmetry measure in a hue based on the threshold binary masks algorithm presented in (Milczarski, 2017), and 33 out of 200 images of PH2 are reported as part of a lesion (not a complete lesion image).

Table 5.16: A comparative analysis of existing approaches with the proposed two algorithms.

Methods	Number of data	Sensitivity (%)	Specificity (%)	Accuracy (%)	Class
Decision Tree (Ali et al., 2020)	204 [ISIC2018]	-	-	80	2
Kullback-Leibler (Chakravorty et al., 2016)	200 [PH2]	81	91	83	3
FFN-LMBP (Damian et al., 2021)	80 [PH2]	-	-	83	2
	80 [MED-NODE]			89	2
Hue distribution (Milczarski, 2017)	167 [PH2]	-	-	83.2	3
Proposed GSAA	200 [PH2]	100	100	100	3
	170 [MED-NODE]	100	100	100	3
	1269 [ISIC2016]	100	100	100	3
Proposed pre-trained model-based	1585 [Aug PH2]	85	93	90	3
	354 [Aug MED-NODE]	92	95	94	3
	1269 [ISIC2016]	90	95	93	3
	200 [PH2 mask]	86	93	91	3
	170 [MED-NODE mask]	97	96	95	3
	1269 [ISIC2016 mask]	95	97	96	3

The proposed methods of this thesis outperform the four approaches in all evaluation metrics. In these approaches, dermoscopic asymmetry analysis was the part of a group of features analysis process toward recognizing Melanoma. For this reason, asymmetry analysis was not the priority. May be that is why these existing methods from the literature have the lower performance than the proposed methods of this thesis.

The proposed GSAA method proves to be beneficial when there is an insufficient amount of annotated data available for training a deep learning model for skin lesion shape classification. This method can be utilized to generate ground truth labels for non-annotated datasets, thereby creating an annotated dataset. Conversely, for training the proposed pre-trained model-based method, a substantial amount of annotated data is required and should be incorporated during the training process. Hence, the proposed

GSAA method can provide geometric-based ground truth that is highly valuable for training the deep learning model. By incorporating multiracial datasets into the training process of the proposed pre-trained model-based method, the research on skin lesion shape analysis becomes more robust, stronger, and adaptable to various types of skin lesion image data.

The performance of the proposed GSAA method on these three datasets is promising, as indicated by these evaluation metrics. This suggests that the output of this method is reliable and can be used as annotated information of training data for the proposed pre-trained model-based method. This research contributes to addressing the shortage of training data for skin lesion shape, thus offering a solution to the problem.

Finally, the proposed pre-trained model-based method was trained on six different datasets. Almost the same experimental results were found in all six datasets, which indicates the flexibility and reliability of this method. Improving accuracy in terms of real-time patient data is a future direction for this part of the research.

5.2.5 Applying Proposed Models on the Unfamiliar Test Set of ISIC2018

Finally, the proposed ResNet101+SVM and ResNet50+SVM models for asymmetric shape detection are trained on the PH2, MED-NODE, and ISIC2016 datasets altogether and tested on a completely unfamiliar test set, which consists of randomly selected 100 dermoscopic images from the ISIC2018 dataset. However, this test set lacked dermoscopic asymmetry ground truth. The asymmetry ground truth for this test set is established with the help of Dermatology division of Universiti Malaya, Malaysia. There are 49 images are symmetrical, 19 images are symmetrical in 1 axis, and 32 images are asymmetrical in this test set. The proposed model successfully classified 98 out of the

100 dermoscopic images based on the shape of lesions. The performance details are reported in Table 5.17.

Table 5.17: Evaluation of the proposed models on the test set of ISIC2018.

Method	Asymmetrical			Sym. in 1 axis			Symmetrical			Overall		
	PR	RC	F1	PR	RC	F1	PR	RC	F1	AC	mF1	wF1
ResNet101	100	94	97	95	100	97	98	100	99	98	98	98
ResNet50	88	88	88	79	79	79	92	92	92	88	86	88

5.3 Dermatological Feature: Pigment Network

Melanoma can be diagnosed using several techniques based on dermoscopic images. The unusual pigment network (PN) on those techniques is one of the most important criteria in Melanoma diagnosis. Finding the PN on a dermoscopic image is difficult. Drawing a distinction between typical pigment network (TPN) and atypical pigment network (APN) is even more a difficult undertaking.

The dermatological PN is a term used to describe the pattern of pigmented lines and structures in the skin, particularly in moles or nevi (especially a birthmark in the form of a raised red patch). This feature is commonly used in the diagnosis of Melanoma, as changes or irregularities in the PN can be an early sign of skin cancer. The PN in a healthy mole or nevus typically appears as a regular and symmetrical pattern of brown or black lines, with a uniform distribution and spacing. In Melanoma, the PN may become disrupted or irregular, with uneven lines, blotchy areas, or an asymmetrical pattern. Figure 5.16 illustrates the typical and atypical PN.

(a) Typical PN   (b) Atypical PN
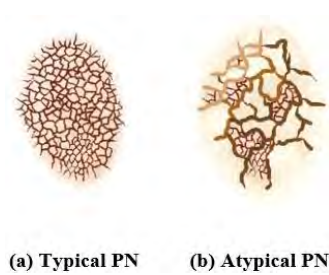
Figure 5.16: Two different types of PN.

Additionally, new or unusual colors may appear, such as red, blue, or white. Dermatologists and other healthcare professionals use various techniques to examine the PN, including dermoscopy, which involves magnifying the skin and examining it with a special light. Other diagnostic tools may include biopsy or imaging tests, depending on the individual case. Overall, the PN is an important feature to consider when evaluating moles or other pigmented lesions for Melanoma. If there is any change in the PN or other aspects of a mole, such as size, shape, or color, it is important to have it checked by a healthcare professional. However, it is not always possible going to specialists for identifying the abnormality in the PN. To overcome this issue, a directional imaging algorithm can be useful to detect the PN automatically and classify them into TPN and APN using a classifier.

5.3.1   The Proposed Pigment Network Detection Approaches

This PN detection process was constructed by following several steps as illustrated in Figure 5.17. The dermoscopic images are taken as input for this process. The source of dermoscopic images is the publicly available annotated dataset PH2 (Mendonca et al., 2013). Other datasets do not have PN labels, therefor those datasets cannot be used in this experiment. Also, it is not possible to create ground truth for PN like asymmetry in MED-NODE dataset using any mathematical or computer vision techniques. Only experts (dermatologists and doctors) can detect PN and distinguish between APN and TPN. In

addition, imaging directional filters (a combination of several directional imaging functions) are applied on the input image to isolate PN from the skin lesion to determine the condition (healthy or unhealthy) of ROI. A new dataset (output of imaging filters, as shown in Figure 5.25) is created by storing these isolated PN. Finally, the new dataset and old annotated dataset are both trained separately to see the difference in classification results between images that have been used with directional filters and those without directional filters. Different classifiers such as SVM, ANN, CNN, and Bag of Features (BoF) are engaged to measure the best performing classifiers for the PN classification. Next, to demonstrate the robustness of the proposed method, a comparison process is done with the result of existing works from the literature.
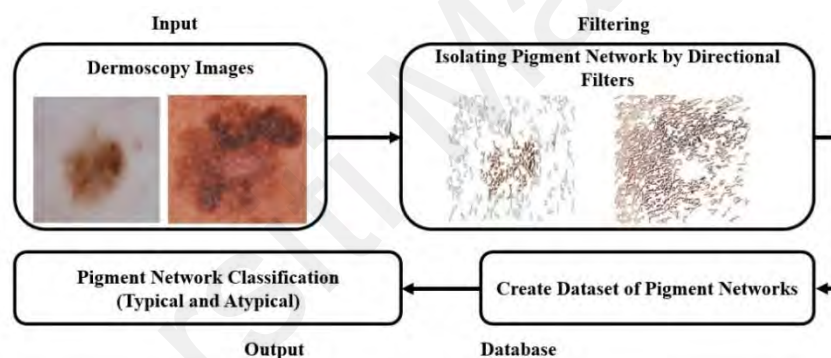
Figure 5.17: Overview of the proposed PN detection and classification system.

5.3.2 Isolating Pigment Network by Directional Filters

To detect PN from dermoscopic images, different sequential imaging techniques are applied. Figure 5.18 shows the steps of the proposed work which will be described in the following sections.
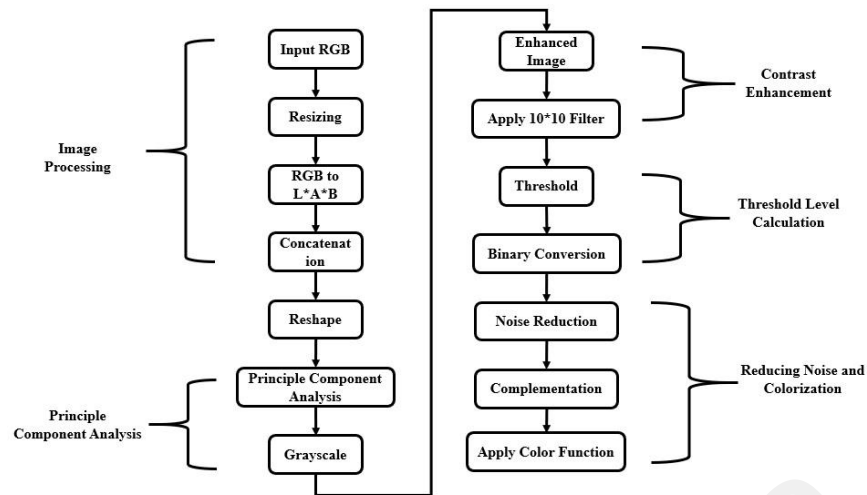
Figure 5.18: The directional imaging approaches for detecting PN.

5.3.2.1 Image Pre-Processing

Due to the different dermoscopic image sizes of the dataset, resizing is required. At first, the input images are resized into 512x512. The dimensions are in pixels as decimal dimensions are not accepted for resizing. This implies that the resizing process only accepts whole numbers for the dimensions, and no decimal values are allowed. For example, if an input image has dimensions of 800x600 pixels, it would be resized to 512x512 pixels, adjusting both the width and height to fit within the specified dimension while maintaining the original aspect ratio of the image. Figure 5.19 shows the resized input images.
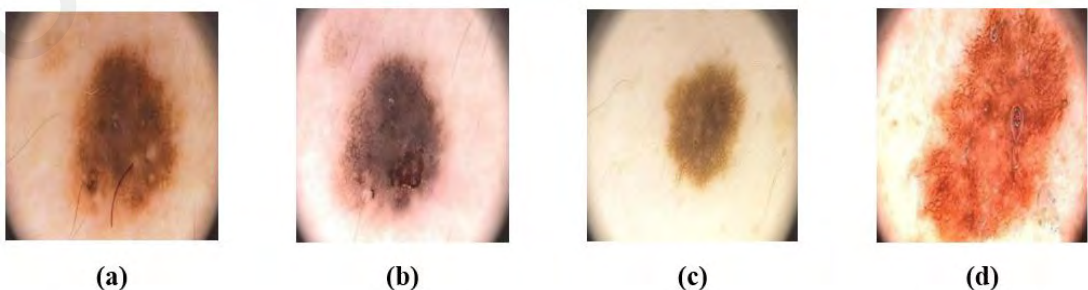
Figure 5.19: (a), (b), (c), and (d) are the outputs upon resizing input images.

Subsequently, the images are converted into double precision. The input image can be a grayscale intensity image, a true color image, or a binary image. Here, the image is an RGB image requiring further conversion of RGB values to the CIE (International Commission on Illumination) 1976 (Kuehni, 1976) L*a*b* values. According to this color space, the "L" represents the likeliness of the color. From 0 to 100, where 0 is black and 100 is white, the likelihood scales. The "a" can symbolize everything from negative to positive, from green to red. Green appears brighter the greater the negative value, and red seems brighter the higher the positive value. The letter "b" stands for blue to yellow. It might be either negative or positive. A function is used to concatenate (1, 0, 0) with the dimension (CIE has 3 channels) of the concatenate arrays. Another function is applied to perform the element-wise binary operation between that concatenated function and the last converted L*a*b images. Next, the element-wise binary operation's output image is reshaped. For this process, the new dimension number should be null since it is not required here. Instead, 3 should be used since it is the existing dimension number of the last function's output.

5.3.2.2   Principal Component Analysis

After the image preprocessing, the Principal Component Analysis (PCA) (Al-Azzawi, 2018) is done. PCA is computed by finding the components which explain the greatest amount of variance. Data standardization (single image), calculation of the covariance matrix for the features in the image, calculation of the eigenvalues and eigenvectors for the covariance matrix, sorting of the eigenvalues and corresponding eigenvectors, selection of k eigenvalues and formation of an eigenvector matrix, and transformation of the original matrix are all steps in the mathematical procedure known as PCA, which reduces the number of dimensions in data. This process returned the coefficients and the score of PCA. Then the scores are resized based on the size of the L*a*b image.

Subsequently, all the rows and columns of the first channel are counted and stored. Next, the scores of PCA are subtracted from the minimum value of scores, and then the subtracted result is divided by the difference between the maximum and minimum value of scores. This division process is an element-wise division. The final output of this analysis process is a grayscale image which is shown in Figure 5.20 for the multiple input images.
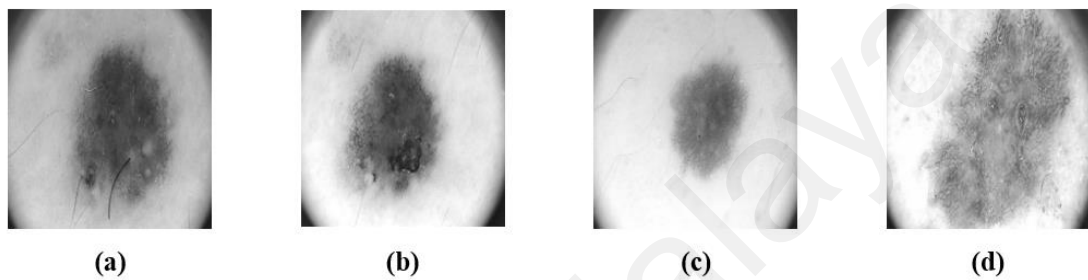
Figure 5.20: (a), (b), (c) and (d) are the outputs after PCA for different input images.

5.3.2.3 Contrast Enhancement

Contrast enhancement is requisite for the last output grayscale image. This is done by employing the adaptive histogram equalization (AHE) function. It improves the contrast of a grayscale image by applying contrast limited AHE to alter the values (Abood, 2018). The adaptive method is distinct from the conventional histogram equalization in that it computes several histograms, each one corresponding to a different area of the image, and then used them to distribute the brightness values of the image. For improving local contrast and sharpening edge definitions in various areas of an image, this is therefore perfect. This enhancement process is done block-by-block where the block size is 8x8 and the number of beams is 128. Figure 5.21 presents the enhancing process outputs.

(a)   **(b)**   **(c)**   **(d)**
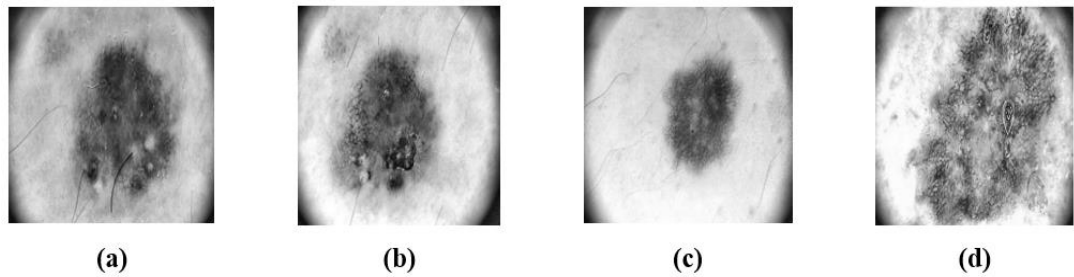
Figure 5.21: (a), (b), (c) and (d) are the outputs after applying the AHE.

Then a 10x10 filter is applied to the enhanced image. Figure 5.22 shows the output after applying the filter on the enhanced image. Later, the enhanced image is subtracted from the filtered image. The filtered image represents the desired PN in the lesion, while the enhanced image captures additional and unwanted background information. By subtracting the enhanced image from the filtered image, the directional algorithm aims to remove and suppress the unwanted area on the lesions, thus emphasizing the desired PN.

(a)   **(b)**   **(c)**   **(d)**
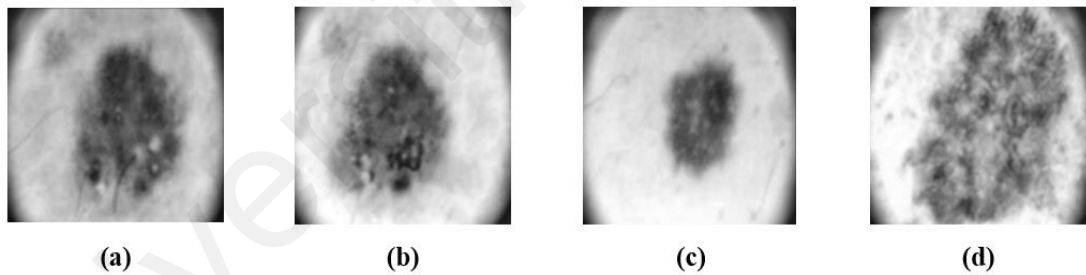
Figure 5.22: (a), (b), (c) and (d) are the outputs after applying that filter to the enhanced images.

5.3.2.4 Threshold Level Calculation

The threshold level calculation is done on the subtracted image (see section 5.3.2.3 Contrast Enhancement) as a prerequisite to segment a PN. In this calculation process, a function is created where an image is converted into an 8-bits data type; followed by counting the histogram and the number of beams. Also, the cumulative sum of the

histogram count is calculated. The mean below and above 'T' is needed to find that mathematically can be presented as

$$T(i) = \frac{\text{sum of (number of Bin} * \text{Histogram count)}}{\text{cumulative sum at the end}} \qquad (21)$$

Here, 'T' is the ratio of the sum of multiplication of bin number and histogram count to the cumulative sum indexed endmost. Furthermore, to find the mean above 'T' and the mean below 'T', the cumulative sum of the histogram counts from 1 to T(i) ('i' is an initial variable) is needed. Finally, the threshold is identified by calculating the average mean below 'T' and the mean above 'T'. Then a condition is given to make the absolute value of 'T' greater than 1. After that the threshold is normalized.

Further, the created function is used, and it applies the subtracted image (see section 5.3.2.3 Contrast Enhancement) to return the threshold level. Next, the subtracted image is converted to a binary image using the threshold level. Instead of using the threshold level as calculated, the value is reduced a little (i.e., 0.008) to make PN more distinct and precise.

5.3.2.5 Reducing Noise and Colorization

To reduce the noise from the binary image of PN, pixels smaller than 100 are removed. This process helps to separate the background from the foreground which is PN. Figure 5.23 shows the output which can be considered a clean image of PN after removing pixels which are smaller than 100.
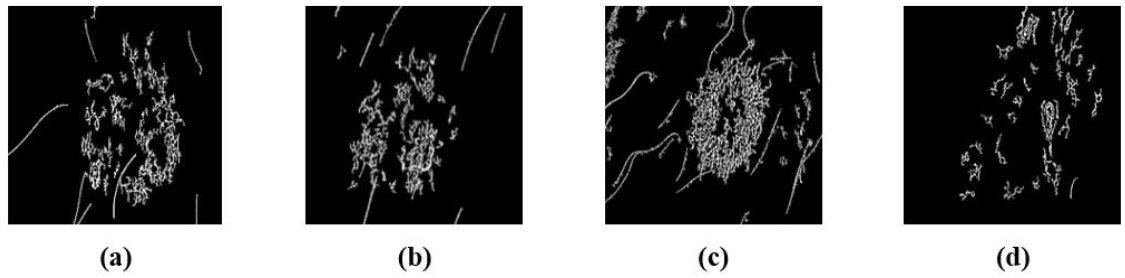
Figure 5.23: (a), (b), (c), and (d) are the noise reduction process outputs.

Since these are binary images, they are converted into color images to understand the topological pattern of PN. The complement of the binary image is taken using the complementation process which returns another complemented image. This process is applied to improve the image contrast. The complement of a binary or intensity image is calculated by the image complement procedure. The process replaces pixel values of 0 and 1 for 1 and 0, respectively, in binary images. Before producing the difference for an intensity image, the procedure subtracts each pixel value from the maximum value that the input data type can represent. Consider the input and output pixel values provided by $m(i)$ and $n(i)$, respectively, as an example. The block produces $n(i) = 1.0-m(i)$ if the input data type is a double or single-precision floating-point. When given an 8-bit unsigned integer as input, the block produces the value $n(i) = 255-m(i)$. Figure 5.24 shows this process outputs.
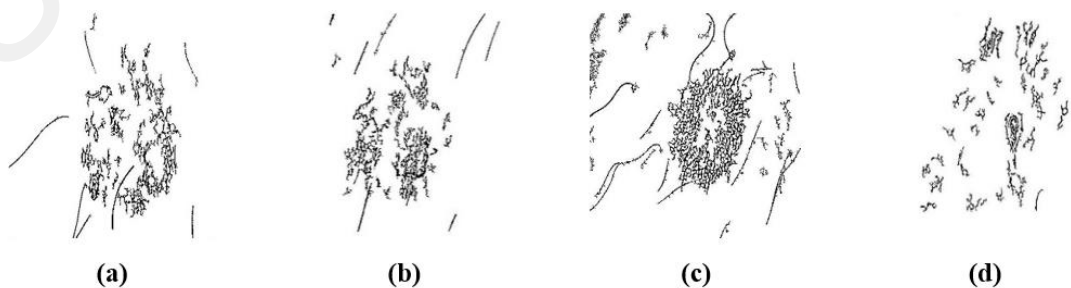
Figure 5.24: (a), (b), (c) and (d) are the complementation process outputs.

A function called colorized is created to color the last output image of PN. As a result, the distinguish color features of PN is visible. Another function is used to return the number of the input argument of the function. If this new function value is less than 3, the system would take the default color. After that, the complemented image is formed in a logical image. Then, the earlier resized image and the color space are converted into an 8-bits data type. If the dimension of the converted resized image is 2, then the image is identified as a grayscale image. The dimension of the input array is returned using another function. If this is the condition, all input channels are initialized with the same values. However, under other conditions, the channel had to be defined. Then, the colors are applied to the complemented image using RGB 8-bit color channels. Finally, the red, green, and blue channels are concatenated into 3-D arrays which produce the colored PN image. To get the desired output, the colorized function is used on the complemented image with the input resized image (512x512) in a white [1 1 1] background. Figure 5.25 shows the final outputs of the colorization process. This colorization of PN indicates the topological properties of PN.
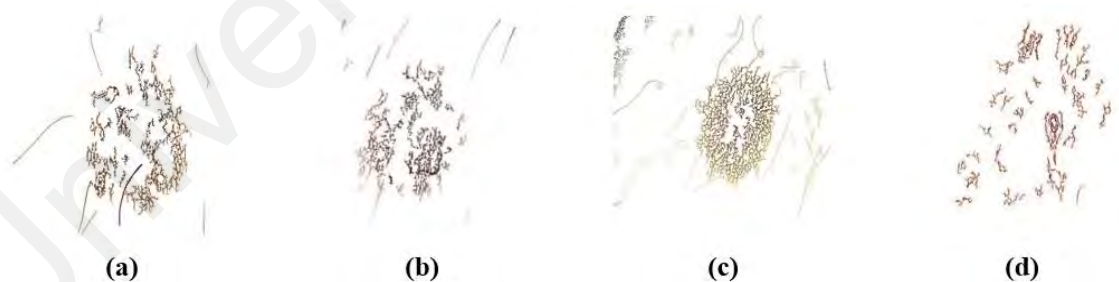
Figure 5.25: (a), (b), (c), and (d) are the outputs of the colorization process.

The algorithm of directional filters to isolate PN from the skin lesion is-

---
Step 1: Load and Resize Image
  1. *Read the dermoscopic image:* **input_image = imread('Dermoscopic Image')**
  2. *Resize the image to 512x512 pixels:* **resized_image = imresize(input_image, [512 512])**
  3. *Display the resized image:* **imshow(resized_image, [])**

Step 2: Convert Image to LAB Color Space and Perform PCA

1. *Convert the resized image to double precision:* **converted_image = im2double(resized_image)**
2. *Convert the image to LAB color space:* **lab_image = rgb2lab(converted_image)**
3. *Apply a mask to the LAB image:* **fill = cat(3, 1, 0, 0)** *and* **filled_image = bsxfun(@times, fill, lab_image)**
4. *Reshape the LAB image for PCA:* **reshaped_lab_image = reshape(filled_image, [], 3)**
5. *Perform PCA:* **[C, S] = pca(reshaped_lab_image)**
6. *Reshape the PCA output:* **S = reshape(S, size(lab_image))**
7. *Extract the first principal component and normalize it:* **S = S(:, :, 1)** *and* **gray_image = (S - min(S(:))) / (max(S(:)) - min(S(:)))**
8. *Display the grayscale image:* **imshow(gray_image)** *and* **title('Gray Image')**

Step 3: Enhance Image Contrast

1. *Apply adaptive histogram equalization:* **enhanced_image = adapthisteq(gray_image, 'numTiles', [8 8], 'nBins', 128)**
2. *Display the enhanced image:* **imshow(enhanced_image)** *and* **title('Enhanced Image')**

Step 4: Filter Image and Perform Subtraction

1. *Apply average filtering:* **avg_filter = fspecial('average', [10, 10])** *and* **filtered_image = imfilter(enhanced_image, avg_filter)**
2. *Display the filtered image:* **imshow(filtered_image)** *and* **title('Filtered Image')**
3. *Subtract the enhanced image from the filtered image:* **subtracted_image = imsubtract(filtered_image, enhanced_image)**

Step 5: Thresholding and Binary Segmentation

1. *Calculate the threshold level:* **level = Threshold_Level(subtracted_image)**
2. *Convert to binary image using the threshold:* **Binary_Image = im2bw(subtracted_image, level - 0.01)**
3. *Display the binary image:* **imshow(Binary_Image)**

Step 6: Remove Small Objects and Complement Image

1. *Remove small objects (less than 100 pixels):* **CI = bwareaopen(Binary_Image, 100)**
2. *Display the cleaned binary image:* **imshow(CI)**
3. *Complement the binary image:* **Complemented_Image = imcomplement(CI)**
4. *Display the complemented image:* **imshow(Complemented_Image)** *and* **title('Complemented Image')**

Step 7: Colorize and Display Final Result

1. *Colorize the original resized image using the complemented binary image mask:* **Final_result = Colorize_Image(resized_image, Complemented_Image, [1 1 1])**
2. *Display the final colorized result:* **imshow(Final_result)**

5.3.3 Evaluation of Proposed Pigment Network Detection Algorithm on Different Datasets

To isolate the Pigmented Network (PN) in dermoscopic images, the proposed directional imaging algorithm has been applied to various multiracial datasets, including PH2, ISIC2019, ISIC2020, and MED-NODE. While only the PH2 dataset contains ground truth annotations for PN, the proposed algorithm demonstrates the capability to detect PN in skin lesion images from the other datasets as well.

PH2: The success rate of directional imaging approaches is 96.00% for 200 images of PH2. Table 5.18 shows the success rate of the filter in two different PN classes. The proposed imaging filters cannot detect 4 out of 83 images from the TPN and 4 out of 117 images from the APN because of the different pixel intensities of these images. Figure 5.26 shows some examples of these 8 images. To detect PN using the same directional filter, a possible strategy is to reduce the threshold level of pixels (i.e., from 0.001 to 0.011). Initially, a new dataset (PN dataset) is created by storing a total of 192 images (96% of the total 200 images), containing only the features of PN. Later, to extract PN from the remaining 8 undetectable images (4 from APN and 4 from TPN), the pixel threshold level is manually and individually adjusted. Finally, the isolated PNs are stored in the PN dataset along with the other 192 images. This PN dataset, consisting of a total of 200 TPN and APN images, is used to train different classifiers.

Table 5.18: The success rate of directional imaging filters in two different classes.

Network type	Not detected	Detected
TPN	4.82%	95.18%
APN	3.41%	96.59%
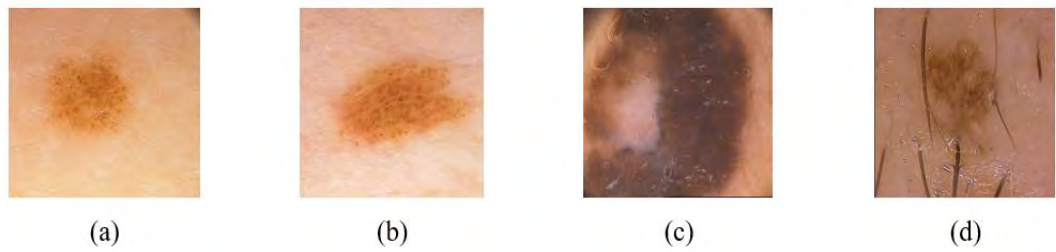
Figure 5.26: Undetectable images (a) and (b) are from the TPN, and undetectable images (c) and (d) are from the APN.

ISIC: To show the flexibility (ability to detect PN on other datasets) of this imaging algorithm, the directional imaging approaches for detecting PN are applied to several dermoscopic images of different origins such as the ISIC2019 and ISIC2020. In Figure 5.27, the input images of different sources and their corresponding output images (after applying the proposed imaging algorithm) are presented. Since ISIC derived from different years of datasets do not provide any annotated information of PN, quantitative evaluation is not available for these datasets in this research. However, detecting PN on the dermoscopic images of the ISIC datasets is a contribution of this research in the PN detection part. The quantitative evaluation for this part is the future recommendation work.
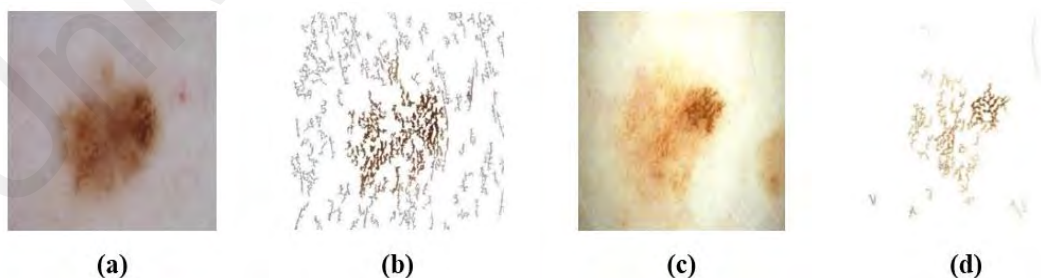
Figure 5.27: (a) is input from ISIC2019 and (b) is output; (c) is input from ISIC2020 and (d) is output.

MED-NODE: The directional imaging algorithm for detecting and isolating PN is demonstrated using a non-dermoscopy dataset called MED-NODE. This dataset

149

comprises macroscopic images, which are photographs or digital images taken through digital cameras or similar devices to magnify the object. All images in the dataset contain Melanoma and Naevus lesions. Figure 5.28 showcases the input images from the non-dermoscopic source alongside their corresponding output images obtained after applying the proposed imaging algorithm. However, it is important to note that the MED-NODE dataset, similar to the ISIC datasets, does not provide any annotated information on pigmented nevi. Therefore, a quantitative evaluation of these datasets is also not available in this research. It is recommended that future work includes a quantitative evaluation of this aspect as well.

(a) **(b)** **(c)** **(d)**
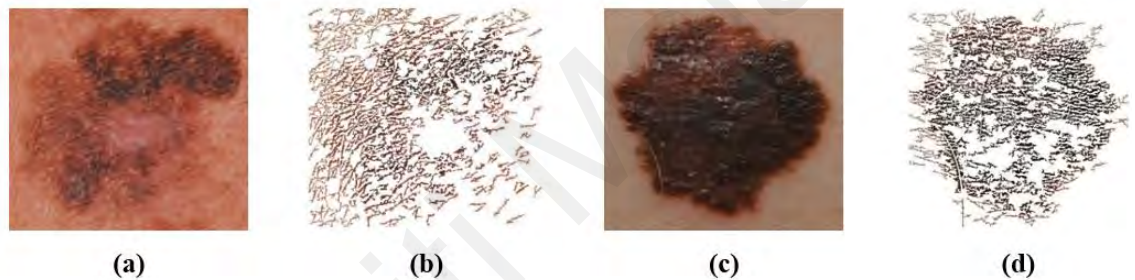
Figure 5.28: (a) and (c) are inputs from MED-NODE where (b) and (d) are their outputs.

5.3.4 Pigment Network Classification

The appearance, pattern and color intensity are the primary differences between the APN and TPN (refer to Figure 5.16). APN displays an irregular and asymmetric pattern with variations in color, thickness, spacing, or distribution of the lines. The lines may be thicker, more uneven, or have abrupt changes in color or thickness. TPN exhibits a regular and symmetric pattern of interconnected brown lines, forming a mesh-like network. The lines are usually thin, evenly spaced, and evenly colored. The line formations of the APN are irregular dark-brown, gray, or black color (Vance & Goding, 2004). In contrast, the line of TPN forms with a uniformly distributed light brown tint (Barhoumi & Baâzaoui, 2014). Consequently, several machine learning-based classifiers are implemented to

categorize them into two classes. The SVM (Cortes & Vapnik, 1995), ANN (Janney et al., 2018), CNN (Valueva et al., 2020), and BoF (Yu et al., 2013) are trained on the PN datasets. Among these four, the best two are reported, namely, CNN and BoF. To demonstrate the effectiveness and accuracy of the imaging mask in identifying different types of PN, two datasets are employed during the training process. Initially, CNN is trained on the PH2 dataset, which contains lesion images including PN. Additionally, a new dataset consisting solely of PN images is utilized for training the CNN. Subsequently, the BoF method is trained using both datasets. It is important to note that neither the ISIC dataset nor the MED-NODE dataset provides annotated information specifically for PN. Therefore, CNN is not trained on these datasets due to the absence of the necessary annotations for PN.

5.3.4.1 Classifier: Convolutional Neural Network

Defining Network Structure: With 5 convolutional layers and 4 max-pooling layers, a simple Convolutional Neural Network (CNN) inspired by LeNet (LeCun et al., 1989) is created and trained. CNN is a crucial technique for deep learning and works particularly well for identifying medical images. Each category in the training set had 80% of the 200 images for each label, while the validation set contained 20% of the total image data. The image input layer specifies the image size, which in this case is 256 by 256-3. These figures represent the channel size, height, and width. The RGB images that make up the PN data. Consequently, the color channel's channel size is 3. The convolutional layer specifies the filter size, which is the height and width of the filters used by the training function when it scans the images. In this case, the number 3 denotes a 3-by-3 filter size. The number of neurons that connect to the same region of the input is known as the number of filters. This variable controls how many feature maps there will be. To add padding to the input feature map, use the 'Padding' name-value pair. The 'same' padding

makes sure that the spatial output size is the same as the input size for a convolutional layer with a default stride of 1. Network training becomes a simpler optimization issue due to batch normalization layers, which normalize the activations and gradients propagating through a network. To speed up network training and lower the sensitivity to network initialization, batch normalization layers are utilized between convolutional layers and nonlinearities, such as clipped rectified linear unit (Clipped ReLU) layers. A nonlinear activation function comes after the batch normalization layer. Clipped ReLU is one of the utilized activation functions.

The performance of various activation functions, such as ReLU, ELU, Clipped ReLU, Leaky ReLU, PReLU, and Tanh, in the hidden layers was compared by an ablation study. The outcomes showed that for the few image datasets, utilizing the Clipped ReLU activation function produced the best results. Next, the max-pooling layer determines the 3-by-3 size of the pool and returns the maximum values of rectangle input regions. The rectangle in this case measures 2 by 2. The final fully connected layer combines the attributes to categorize the images. As a result, there are exactly as many classes in the target data as there are output dimensions in the final fully connected layer. The output size in this case is 2, representing the two classes (i.e., APN and TPN). The softmax adjusted the fully connected layer's output to a normal value. Positive numbers that add up to one make up the output of the softmax layer, which the classification layer can utilize to determine classification probabilities. The categorization layer is the top layer. This layer calculated the loss by classifying each input into one of the mutually exclusive classes using the probabilities supplied by the softmax for each input. The training options are defined following the network structure definition.

Training the Network: With an initial learning rate of 0.01 and stochastic gradient descent with momentum, the network is trained. There can be a maximum of 250 epochs.

A full training cycle on the entire training dataset constitutes an epoch. Specifying validation data and periodicity allows for the monitoring of network accuracy during training. Every epoch involves 25 validation frequencies for data shuffling. When training CNN, overfitting is a common problem where the model does well on the training data but fails to generalize to new data. In this instance, the CNN training is terminated early to reduce overfitting during the training phase. The Matlab2021a program calculates the accuracy of the validation data at regular intervals while training the network using the training data. Updating the network weights does not use the validation data.

The architecture determined by layers is used to train the CNN. A CNN's layers each produce an activation in response to an input image. However, only a few layers of a CNN are appropriate for extracting visual features. The network's initial layers record fundamental visual characteristics like edges and blobs. Figure 5.29 displays the CMs for two separate datasets together with the final validation accuracy calculation after this trained network has predicted the labels of the validation data. The percentage of labels that the network properly predicts is known as accuracy.
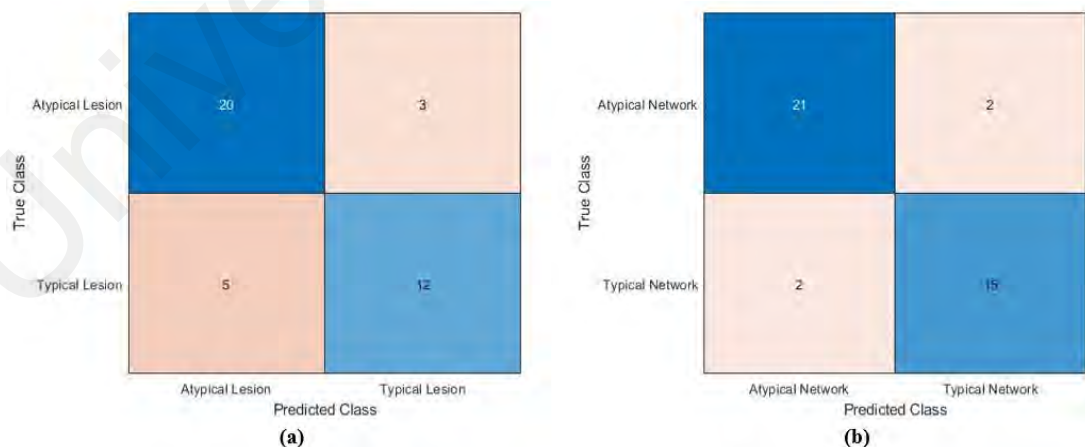
Figure 5.29: The CM, (a) is on PH2, and (b) is on the PN dataset using CNN.

The CNN classifier's result is evaluated using different evaluation metrics such as sensitivity, specificity, precision, and accuracy for the CM (a) and (b). The PH2 dataset contains skin lesion images, and the new dataset contains only PN. The performance of CNN trained on lesion images is not as good as when the CNN is trained exclusively on PN (New) dataset. Typically, lesion images encompass numerous dermatological features. However, the PN (New) dataset, which solely focuses on PN, consists of a limited set of features that prove valuable in classifying PN into two distinct classes. The results of the evaluation are shown in Table 5.19.

Table 5.19: The result of evaluation metrics on two different datasets for the CNN classifier.

Datasets	SE	SP	PR	AC
PH2	0.80	0.78	0.80	0.80
PN	0.90	0.89	0.90	0.90

From Table 5.19, it is found that CNN classifier performs better on the PN-based (new) dataset than the lesion-based (PH2) dataset.

5.3.4.2 Classifier: Bag of Features (BoF)

Defining the BoF: The BoF approach is implemented to classify PN. This approach is also known as a bag of words. The process of providing a category name to an image under the test is known as visual image categorization. Images portraying just about any form of thing can be found in categories. For example, images of animals can be found in categories such as cats, dogs, birds, mammals, or reptiles. Similarly, images of vehicles can be categorized into cars, trucks, motorcycles, bicycles, or boats. The same goes for various other objects, such as buildings, landscapes, food, people, sports, and so on. These categories help organize and classify images based on their content or subject matter. The data is divided into two categories: training and validation. 80% of the images are chosen

from each batch for training and the remaining 20% for validation. To avoid biases in the results, a split random is applied. BoF is a natural language processing technique that has been extended to computer vision. A "vocabulary" of speeded-up robust features (SURF) (Bay et al., 2006) representative of each image category is created because images do not contain discrete words. The SURF is a fast and robust algorithm for local, similarity-invariant representation and comparison of images. This is done with a single call to the specific function, which extracts the SURF features from each image through all categories and creates the visual vocabulary by lowering the number of features using K-means clustering to quantize feature space.

The BoF process also has an encode function that counts the number of visual features (color and thickness of PN) occurrences in an image. It resulted in a histogram, which is a new and simplified representation of an image. Figure 5.30 shows us the example of histogram images from the PH2 and PN datasets.
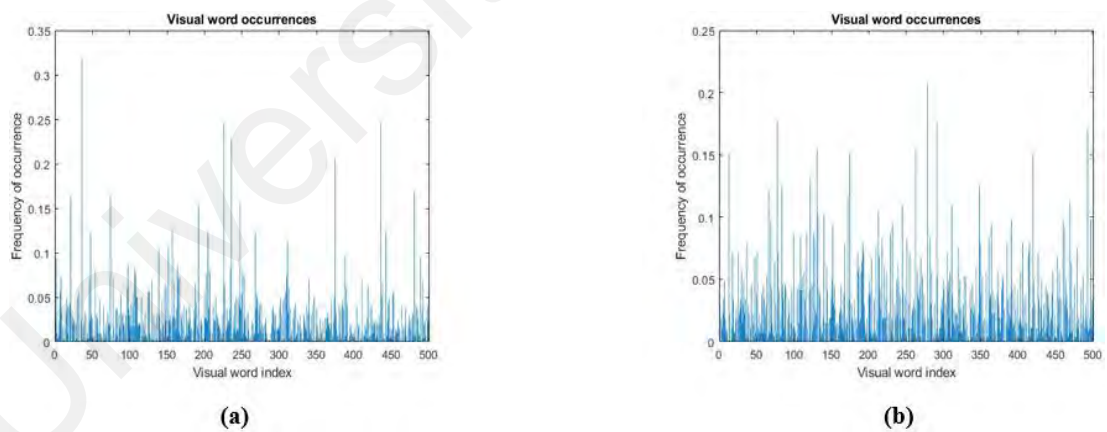
Figure 5.30: One of the histogram images from PH2 is (a), and from the PN dataset is (b).

Training the BoF: This histogram is used to train a classifier as well as for image categorization. It converts an image to a feature vector. The category classifier function calls for the encoded training images from each category to be input for a classifier

training procedure. The encode process uses the input of BoF to generate feature vectors for each image category in the training set. An evaluation is performed upon completing the training process. This is followed by a sanity check with the training set, which yields a near-perfect CM. The classifier is then evaluated on the validation set, which uses datasets not utilized during the training process. The CM is returned by default by the evaluate function, which is a useful first indicator of how well the classifier is functioning. In Figure 5.31, the CMs for two different datasets are shown after predicting the labels of the validation set using this trained classifier. The BoF exhibits superior performance when trained on the PN dataset, which exclusively consists of PN-related images. However, the BoF's performance is slightly lower when trained on the PH2 dataset, which encompasses skin lesion images. The evaluation result of BoF classifier is presented in Table 5.20.
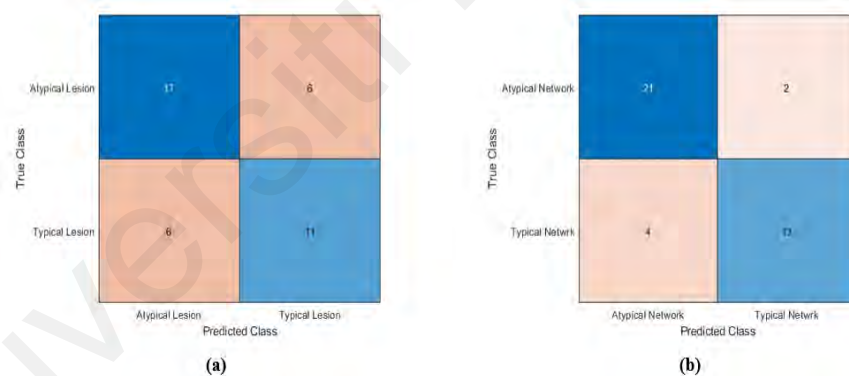
Figure 5.31: The CM, (a) is on PH2 and (b) is on the PN dataset using BoF.

Table 5.20: The result of evaluation metrics (%) on two different datasets for the BoF classifier.

Datasets	SE	SP	PR	AC
PH2	0.70	0.69	0.70	0.70
PN	0.85	0.82	0.85	0.85

In Table 5.20 and Table 5.19, both the BoF and CNN exhibit better performance on the PN dataset compared to the PH2 dataset. This indicates using the PN dataset helps to achieve better performance to classify PN.

By comparing Table 5.19 and Table 5.20, CNN is a better classifier than BoF for the PN categorization based on the results of different evaluation metrics. In Figure 5.32, the ROC (receiver operating characteristic) curves of these two classifiers on two datasets are shown.
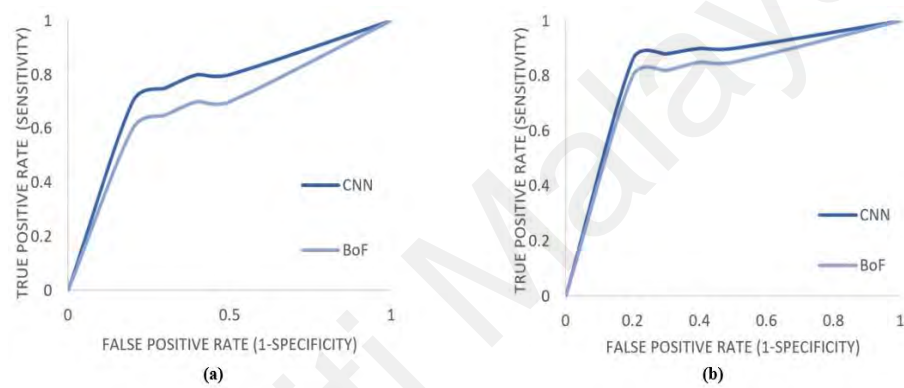
Figure 5.32: ROC curves for (a) PH2 dataset, and (b) PN dataset for the CNN and BoF respectively.

The area under the ROC curve (AUC) for CNN and BoF are 0.76 and 0.69 (on PH2); 0.84 and 0.80 (on PN dataset). The AUC value indicates that the CNN performs better than BoF as a classifier on the PN dataset compared to the PH2 dataset, showcasing its superior performance.

5.3.5 A Comparative Study for Pigment Network Detection Method

To demonstrate the usefulness and effectiveness of the proposed PN identification approach compared to state-of-the-art methods, three existing methods are analyzed. A

157

significant number of research is done to detect PN. However, only three works are done on classification of PN.

In the comparison process, the reported result of two layers of feed-forward ANN is included which was proposed by Pathan et al. (2018). Then, the outcome of the Adaboost classifier reported by Barata et al. (2012) is considered. In both cases, experimental approaches were done on 200 dermoscopic images of PH2. Without implementing their approaches, a direct comparison between their results and proposed method is performed.

Finally, the proposed method is analyzed against the naïve Bayes classifier by Kropidlowski et al. (2015). They implemented their algorithm (based on region histogram, co-occurrence matrix, run-length matrix, gradient matrix, and autoregression model) on their dataset that contains only 44 PN images (19 APN and 25 TPN). For the sake of fair comparison, their algorithm on PH2 is implemented and the output is evaluated. The results obtained indicate the successfulness of the proposed method over the other three methods as shown in Table 5.21.

Table 5.21: Performance (%) comparison between the state of art methods and the proposed method for identifying PN on PH2.

Methods	SE	SP	PR	AC
ANN (Pathan et al., 2018)	84.6	88.7	86.7	86.7
AdaBoost (Barata et al., 2012)	**91.1**	82.1	86.2	86.2
Naïve Bayes (Kropidlowski et al., 2015)	85.0	88.5	85.0	87.0
CNN (Proposed method)	90.0	**89.0**	**90.0**	**90.0**

5.3.6   Applying Proposed CNN Model on the Unfamiliar Test Set of ISIC2018

Finally, the proposed CNN model for pigment network detection is trained on both PH2 and PN datasets combined and tested on a completely unfamiliar test set, which consists of randomly selected 100 dermoscopic images from the ISIC2018 dataset. There are 73 images with atypical and 27 images with typical pigment networks in this test set.

The proposed model successfully classified 90 out of the 100 dermoscopic images based on the pigment network. The performance details are reported in Table 5.22.

Table 5.22: Evaluation of the proposed models on the test set of ISIC2018.

Method	Atypical Pigment Network			Typical Pigment Network			Overall		
	PR	RC	F1	PR	RC	F1	AC	mF1	wF1
Proposed CNN	94	92	93	79	85	82	90	88	90
Bag of Features	90	82	86	61	74	67	80	76	81

5.4   Dermatological Feature: Blue White Veil

The Blue White Veil (BWV) is a visual feature that can be observed in some cases of Melanoma. This feature refers to an opaque, whitish-blue coloration that is visible on the surface of the lesion when it is examined with a dermatoscope, a handheld device that allows for a magnified examination of the skin. The BWV is caused by the scattering of light within the skin tissue and is thought to be due to the presence of melanin in the upper layers of the skin. Figure 5.33 shows the example of BWV in the lesion images.

(a)   (b)
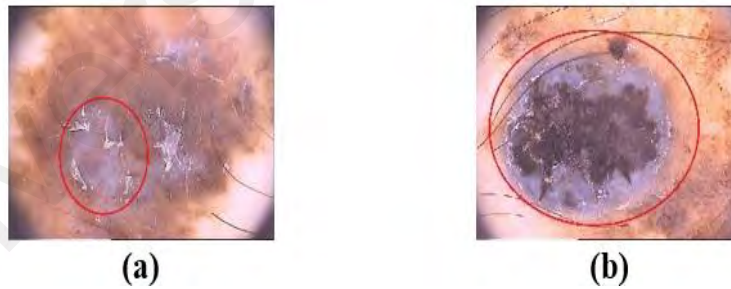
Figure 5.33: The BWV (red circle) in the dermoscopic images (a) and (b) of PH2 dataset.

When the BWV is observed in combination with other features such as asymmetry, pigment network, and variegated coloration, it can be suggestive of Melanoma. The BWV is a dermatological feature that can be helpful in the detection of Melanoma, but it is difficult to notice the lesion surface by the experts and dermatologists via the naked eye.

Therefore, this manual detection process can be fully automated by the proposed algorithm described below. Three different datasets are used in this section such as PH2, MED-NODE, and Derm7pt. PH2 and Derm7pt have ground truth about BWV, and MED-NODE has not any information about BWV. Since MED-NODE is without ground truth, an imaging algorithm is applied in the preprocessing section to detect BWV on the images of this dataset. Figure 5.34 presents the overall process of BWV detection on the skin lesions.
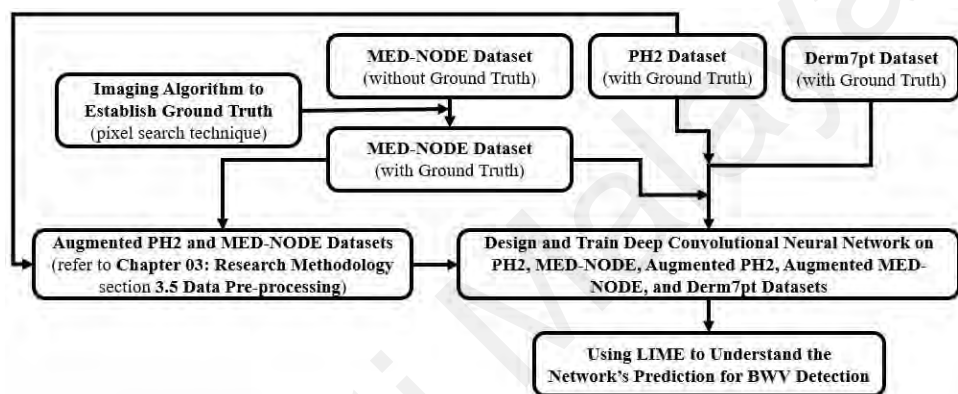
Figure 5.34: The overall process of BWV detection on multiracial datasets.

5.4.1 5.4.1 Establishing Ground Truth of BWV on MED-NODE

The BWV detection is done on original dataset. There is no need of preprocessing as color element will be affected. Since the MED-NODE dataset has no annotated information about BWV, a proposed color threshold-based analysis is done to determine the presence of a veil by matching the color palette. An algorithm is developed to find the pixel value on lesions according to the pixel value of the color palette. If the pixel value of the lesion matches the pixel value of the color palette, then the presence of BWV is confirmed on the lesion. The color palette is a combination of 80 color patches based on the most frequent colors identified as BWV (Madooei et al., 2013; Landa & Fairchild, 2005). Figure 5.35 shows those 80 color patches.

Figure 5.35: The 80 color patches based on the most frequent colors identified as BWV (Madooei et al., 2013).

Each input image is processed through checking pixel (based on a color threshold) to find the similar pixel (red, green, and blue components) value as those 80 patches on the lesion. At the end of this process, the MED-NODE dataset is divided into two classes containing 44 BWV and 126 Non-BWV lesions. The algorithm to detect BWV on the dermoscopic images is-

---

Step 1: Read and Resize Image
*Read the input dermoscopic image: **input_image = imread('Dermoscopic Image');***

Step 2: Create an Empty Mask
*Create an empty mask of the same size as the input image: **mask = zeros(size(input_image, 1), size(input_image, 2));***

Step 3: Set Thresholds for BWV
Define the thresholds for BWV's colors in RGB values: A range of RGB channels values for the BWV's colors according to 80 color patches.

Step 4: Iterate Over Each Pixel in the Image
Iterate over each pixel in the image:
- *Retrieve the RGB color values of the current pixel.*
- *Check if the pixel's RGB values fall within the BWV's threshold ranges.*
- *If the condition is met, set the corresponding pixel in the mask to 1 (or True); otherwise, set it to 0 (or False).*

Step 5: Apply Morphological Operations
Apply morphological operations (e.g., erosion, dilation) on the mask to refine the detected regions if needed.

Step 6: Find Contours or Connected Components

Find contours or connected components in the mask to identify regions of BWV's colors resembling a veil.

Step 7: Additional Analysis or Post-Processing (Optional)
Optionally, perform additional analysis or post-processing on the detected regions.
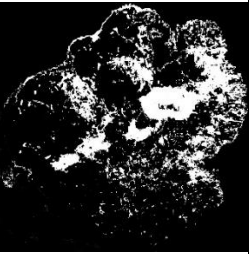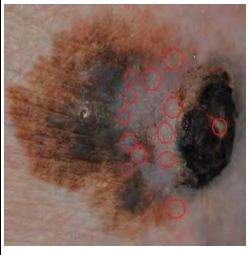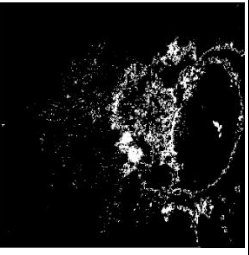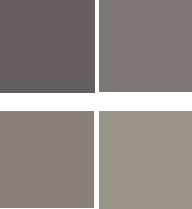
Step 8: Display the Resulting Mask
Display the resulting mask or overlay it on the original image to visualize the detected veil-like regions.

Table 5.23 describes the imaging algorithm (pixel search technique) by visualizing the presence of BWV on the input image. Finding those specific pixels in the lesion images involves locating the pixel's coordinates (row and column) within the image's grid of pixels. This is typically done using image's (x, y) coordinates, where 'x' represents the column and 'y' represents the row.

Table 5.23: Description of imaging algorithm (pixel search technique) on MED-NODE dataset.

Original	Presence of BWV	Binary Mask	Color palette	RGB Value	
				Red	Min: 85
					Max: 142
				Green	Min: 76
					Max: 132
				Blue	Min: 67
					Max: 116
				Red	Min: 86
					Max: 135
				Green	Min: 76
					Max: 123
				Blue	Min: 74
					Max: 123

5.4.2   Deep Convolutional Neural Network for BWV

The initial stage in designing and training a Deep Convolutional Neural Network (DCNN) for distinguishing two classes is to precisely identify the regions of interest in lesion images that are especially associated with BWV. Due to the difficulties presented by the two classes concerned, conventional state-of-the-art deep learning models may not be sufficiently effective for this task. It is possible to train the DCNN to concentrate on these important features by carefully identifying and isolating the BWV regions within the images. This could result in a classification model that is more reliable and specifically designed for dermoscopic image analysis and has higher accuracy. The proposed DCNN model (31 layers including input and output layers) is inspired by LeNet (LeCun et al., 1989) and the third research objective of this thesis (see Chapter 6: Classification Based Melanoma Detection). Different neural network parameters such as number of layers, activation function, kernel or filter size, number of filters or neurons, dilation factor, padding, and stride went through the different ablation studies to design an effective deep learning model for the classification of BWV lesions. Table 5.24 describes the proposed CNN layer by layer with different parameters. These parameters are responsible for estimating the network performance. The performance of DCNN is influenced by various parameters, including number of layers, number of filters (feature maps), kernel size, stride, padding, activation function, pooling, dropout, learning rate, batch size, optimizer, weight initialization, and data augmentation. Overall, finding the optimal combination of these parameters often involves experimentation and hyperparameter tuning, as different tasks and datasets might require specific settings for optimal DCNN performance.

Table 5.24: Step-by-step description of the proposed DCNN model.

No.	Layer Type	Kernel Size	Number of Filters / Neurons	Dilation Factor	Padding	Stride
1	Image/Input	-	-	-	-	-
2	Convolution	5x5	8	2x2	same	2x2
3	Normalization	-	-	-	-	-
4	Custom	-	8	-	-	-
5	Max Pooling	5x5	-	-	same	2x2
6	Convolution	3x3	16	3x3	same	3x3
7	Normalization	-	-	-	-	-
8	Custom	-	16	-	-	-
9	Max Pooling	3x3	-	-	same	3x3
10	Convolution	5x5	32	2x2	same	2x2
11	Normalization	-	-	-	-	-
12	Custom	-	32	-	-	-
13	Max Pooling	5x5	-	-	same	2x2
14	Convolution	3x3	64	1x1	same	1x1
15	Normalization	-	-	-	-	-
16	Custom	-	64	-	-	-
17	Max Pooling	3x3	-	-	same	1x1
18	Convolution	5x5	128	2x2	same	2x2
19	Normalization	-	-	-	-	-
20	Custom	-	128	-	-	-
21	Max Pooling	5x5	-	-	same	2x2
22	Convolution	3x3	256	1x1	same	1x1
23	Normalization	-	-	-	-	-
24	Custom	-	256	-	-	-
25	Max Pooling	3x3	-	-	same	1x1
26	Convolution	5x5	512	3x3	same	3x3
27	Normalization	-	-	-	-	-
28	Custom	-	512			
29	Fully connected	-	-	-	-	-
30	Softmax	-	-	-	-	-
31	Classification / Output	-	-	-	-	-

The input image should be in RGB format with dimensions 256-by-256. Height and width of the filters or kernels, expressed as a vector *[h w]* of two positive integers, where *h* represents height and *w* represents width. The size of the local regions that the neurons in the input connect to depends on the filter or kernel size. It specifies a positive integer as the number of filters or kernels. According to this figure, there are exactly as many

neurons in the layer as there are input regions in the input. This parameter controls how many channels (feature maps) will be delivered as a layer. The term "dilation factor" in this context refers to the factor for dilated convolution, also called atrous convolution, and specified as a vector *[h w]* of two positive integers, where *h* is the vertical dilation and *w* is the horizontal dilation. Without adding more parameters or computations, using dilated convolutions expands the layer's receptive field (the portion of the input it can view). Based on Table 5.24, in 'same' padding, the amount of padding is dynamically calculated which is required to maintain the output size equal to the input size (width or height) when the stride is 1. If the stride is greater than 1, the output size is determined by dividing the input size by the stride and rounding up to the nearest integer (ceiling function). Padding is equally added to the top, bottom, left, and right sides of the input. When the required vertical padding is an odd value, extra padding is added to the bottom to ensure an equal amount on both sides. Similarly, when the required horizontal padding is odd, extra padding is added to the right to maintain symmetry. The 'same' padding mode is commonly used in CNNs to ensure that the output feature map has the same spatial dimensions as the input, which can be beneficial in various applications like image segmentation, where preserving spatial information is crucial for accurate results.

5.4.3   Defining Custom Deep Learning Layer

Often a custom deep learning layer is required to be built for the classification or regression problem to get better performance. Custom deep learning layers are required in specific scenarios where standard or pre-defined layers provided by popular deep learning frameworks may not fulfill the specific needs of a particular task or model architecture. There are several reasons why custom layers may be necessary: novel architectures, specialized operations, complex operations, efficiency and optimization, domain-specific requirements, model interpretability, transfer learning, pre-training,

experimentation, and research. A special deep learning layer called the PReLU layer, which has a learnable parameter, is created and used in a convolutional neural network (Crnjanski et al., 2021). Any input value that is less than zero for each channel is multiplied by a scalar that was learnt during training as part of the PReLU layer's threshold operation. The PReLU layer applies scaling coefficients a_i to each channel of the input for values less than zero. During training, the layer picks up on these coefficients to create a learnable parameter. Figure 5.36, from (He et al., 2015) compares the ReLU and PReLU activation functions.
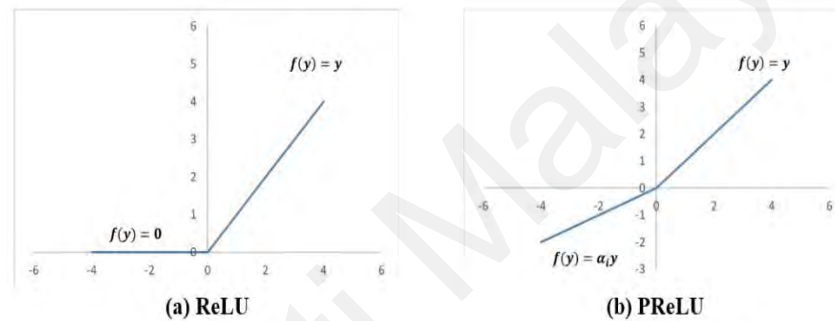
Figure 5.36: ReLU vs. PReLU. For PReLU, the coefficient of the negative part is not constant and is adaptively learned.

The mathematic formula of the PReLU function is-

$$f(y) = \begin{cases} y & if\ y > 0 \\ a_i y & if\ y \leq 0 \end{cases} \qquad (22)$$

Here y is the input of the nonlinear activation f on channel i, and a_i is the coefficient controlling the slope of the negative part. The subscript i in a_i indicates that the nonlinear activation can vary on different channels.

5.4.4 Training the Network

The PH2, MED-NODE, and Derm7pt datasets are involved separately to train the proposed network. The image data of each dataset is divided into three sets such as training set (70%), validation set (25%), and test set (5%). The test set is separated at the beginning of training and used for the evaluation of the proposed deep convolutional neural network. A total of 95% including training and validation sets involves directly in the training process. The network learnable parameters are updated in a custom training loop using the stochastic gradient descent with momentum (SGDM) algorithm. The training options are for SGDM, including learning rate (0.01), the maximum number of epochs (60 – 300, small – larger datasets), shuffling the training data before each training epoch and shuffling the validation data before each network validation, frequency of network validation in the number of iterations (30), and displaying plots during network training. The network validation results (using a test set) will be shown as a CM to understand the efficiency of the trained network (see Figure 5.42).

5.4.5 Understanding Network Predictions Using LIME

To understand why a deep convolutional neural network chooses a particular classification, this section describes how to apply locally interpretable model-agnostic explanations (LIME). Because deep neural networks are complex, it might be challenging to understand their judgments. Due to LIME's model-agnostic nature and capacity to produce human-understandable explanations for individual predictions, LIME is preferred over SHAP (SHapley Additive exPlanation) for interpreting image classifiers. LIME excels in scenarios where the underlying model is intricate or opaque, offering insightful information without being constrained by the model architecture. The LIME approach recreates the categorization behavior of a deep neural network using a simpler, easier to comprehend model, such as a regression tree. Understanding the choices made

by this more basic model can help in understanding the choices made by the neural network (Ribeiro et al., 2016). The fundamental model is applied to evaluate the relevance of input data features as a proxy for the significance of the characteristics to the deep neural network. When a deep network decides to classify anything, the removal of a particular feature has a major impact on the classification score. Therefore, the simple model also values that attribute. The LIME technique can be used to compute maps of feature importance for determining important features. The algorithm of LIME for images is-

---

Step 1: Segment the input image into features
def segment_image(image):
*   # Implement image segmentation algorithm*
*   return segmented_image*

Step 2: Generate synthetic images with random feature inclusion/exclusion
def generate_synthetic_images(image, num_samples):
*   synthetic_images = []*
*   for i in range(num_samples):*
*   synthetic_image = randomly_include_exclude_features(image)*
*   synthetic_images.append(synthetic_image)*
*   return synthetic_images*

Step 3: Classify synthetic images with the deep network
def classify_synthetic_images(synthetic_images, deep_network):
*   predictions = []*
*   for image in synthetic_images:*
*   prediction = deep_network.predict(image)*
*   predictions.append(prediction)*
*   return predictions*

Step 4: Fit a simpler regression model using binary features for the scores
def fit_regression_model(synthetic_images, predictions):
*   regression_model = SimpleRegressionModel()*
*   regression_model.fit(synthetic_images, predictions)*
*   return regression_model*

Step 5: Compute feature importance and create an importance map
def compute_feature_importance(regression_model, original_image):
*   feature_importance = regression_model.compute_importance(original_image)*
*   importance_map = convert_importance_to_map(feature_importance)*
*   return importance_map*

Main LIME algorithm:

def LIME_algorithm(original_image, deep_network, num_samples):

   Step 1: *Segment the input image*

   segmented_image = segment_image(original_image)

   Step 2: *Generate synthetic images*

   synthetic_images = generate_synthetic_images(segmented_image, num_samples)

   Step 3: *Classify synthetic images with the deep network*

   predictions = classify_synthetic_images(synthetic_images, deep_network)

   Step 4: *Fit a regression model*

   regression_model = fit_regression_model(synthetic_images, predictions)

   Step 5: *Compute feature importance and create an importance map*

   importance_map=compute_feature_importance(regression_model, original_image)

   return importance_map

5.4.5.1   Loading Pretrained Network and Image

The proposed DCNN model is trained on different datasets. After that the trained model predicted BWV on the test sets. To understand the predicting process of the trained model, LIME is applied. At first, the pre-trained proposed deep network is loaded. The image input size and the output classes of the network are extracted. The image is loaded individually. The images are of two skin lesions called BMB00079 (from PH2) and BNK00100 (from MED-NODE). The image is resized to the network input size. The image is classified, and the two classes are displayed with the highest classification score in the image title. DCNN classified both skin lesions as BWV which is presented in Figure 5.37. Understandably, the network also assigned a high probability to the non-BWV class (because there are some common features in both BWV and non-BWV classes). The *imageLIME* function is used to understand which parts of the image the network is used to derive these classification decisions. In BWV present and absent images, the lesion backgrounds are same. In the interested area of skin lesion images, several features are found usually which are common for both BWV and non-BWV lesions. However, the distinguish features make them separate from each other. Since there are only two classes such as BWV and non-BWV for the lesion classification, that is why the probability of both classes are high.

169

BWV (0.55); Non-BWV (0.45)   BWV (0.60); Non-BWV (0.40)

(a)   (b)
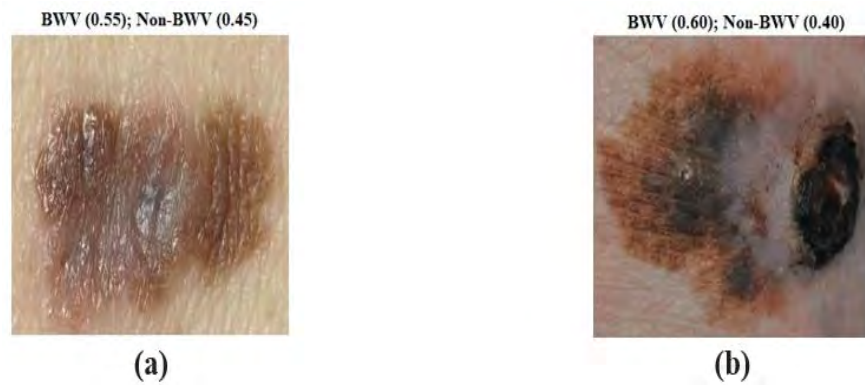
Figure 5.37: The classification decision of DCNN.

5.4.5.2 Identifying Areas of an Image the Network Uses for Classification

Using LIME, it is possible to determine which portions of an image are crucial for a particular class. The BWV predicted class is first observed. The graphic makes this class suggestion in several places. LIME function segments the input image into superpixels by default to identify features in the image. The image is segmented into square features using a particular segmentation method using a grid function. To map the significance of various superpixel properties, we employ the LIME function. Figure 5.38 shows both BWV photos with the LIME map overlayed. It is important to remember that Figure 5.38 represents the order of Figure 5.37. The map displayed the portions of the image that are crucial for BWV classification. The probability score for the BWV class dramatically decreased when the red parts of the map were eliminated since they had a higher importance. To construct its BWV forecast, the network concentrated on the red and yellow regions of the lesions.
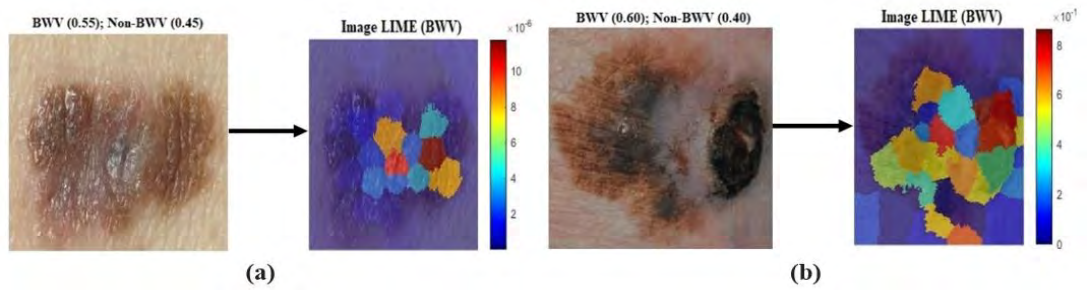
Figure 5.38: Displaying important parts of images for BWV class.

5.4.5.3 Comparing to Results of Non-BWV Class

For the two input images (BMB00079 and BNK00100), the proposed model predicted to have the veil with a score of 55% and 60% for the BWV class, and 45% and 40% for the non-BWV class. The highest probability is in the BWV class and the second highest probability is in the non-BWV class for having veil. These courses share a lot in common. Comparing the LIME maps calculated for each class reveals which areas of the lesion are more significant for both classes. The non-BWV class's LIME map is generated using the same settings. The red, brown, and dark brown hues of the lesion receive more attention from the network than the other hues in the non-BWV class. A visual representation is shown in Figure 5.39. While both maps highlighted the lesions' other colors, the network identifies that the lesions' bluish and whitish area indicates the BWV class, while the lesions' red, brown, and dark-brown colors (based on quantitative assessment) indicate the non-BWV class.

Figure 5.39: Comparing the focus points of Non-BWV class.

5.4.5.4 Comparing LIME with Grad-Cam

There are two alternative methods for explaining the predictions of machine learning models: Grad-CAM and LIME. Both approaches strive to increase a model's interpretability and transparency by offering insights into how a model makes decisions. They have different approaches and application cases, though. The generated map is upsampled by Grad-CAM to create a smooth heat map of the critical regions of the image. By determining the significance of square or rectangular features and upsampling the resulting map, LIME is used to create the similar-looking maps. In Figure 5.40, LIME and Grad-Cam produced images are compared side by side. Use of the segmentation with grid function allows for the division of the picture into a grid of square features as opposed to irregular superpixels. By setting bicubic with output upsampling, the generated map is upsampled to correspond to the image resolution. The original computed map's resolution is improved by increasing the number of features to 100. This results in a 10-by-10 grid of features because the image is square. By selecting specific characteristics at random, replacing all their pixels with the average picture pixel, and then essentially eliminating those features, the LIME technique creates synthetic images based on the original observation. There are now 6000 random samples. More samples produce better results when there are more features. The LIME function's default simple model is a regression tree. An alternative approach is to fit lasso regression to a linear regression model. The LIME approach also firmly identifies the fuzzy area of the lesions as significant to the prediction of BWV, much like the gradient map generated by Grad-CAM.
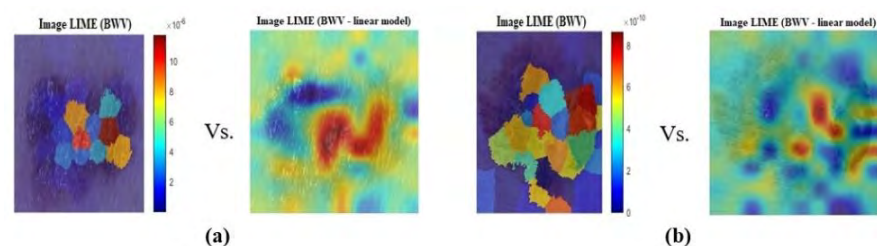
Figure 5.40: Grad-CAM produced the smooth heatmaps.

5.4.5.5 Displaying Only the Most Important Features

Plots of LIME data frequently only display the most crucial elements. When a more interpretable and fine-grained understanding of how specific features or superpixels affect the model's predictions is needed, LIME is preferred to Grad-CAM. While Grad-CAM focuses on the overall activation regions of the last convolutional layer, providing a more comprehensive view of significant regions but potentially omitting finer details, LIME generates localized explanations by approximating the complex model with a simpler regression model. This method offers more detailed insights into the significance of each region in the input image. When the LIME function is applied to the BWV, a map of the features is produced, and this map is used to compute each feature's relevance. The most crucial superpixel features are identified using these findings, and the four most crucial features are chosen to be displayed in the image. The LIME map is calculated, and the feature map and estimated relevance of each feature are obtained. The top four feature indices have been located. The LIME map is then used to mask out the images, leaving only the pixels in the four most crucial superpixels exposed. Then the masked images are generated. They are presented in Figure 5.41, which is the output of Figure 5.37.
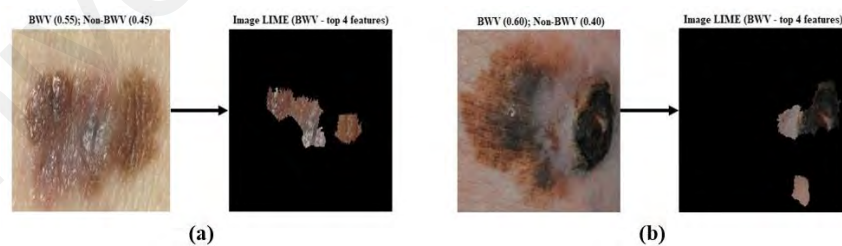
Figure 5.41: Displaying the most important four features.

5.4.6   Experimental Results of the Proposed CNN vs. Other Methods

In this experiment, a total of 5 datasets (PH2, Augmented PH2, MED-NODE, Augmented MED-NODE, and Derm7pt) including 2 augmented datasets are used (Other datasets were not used due to the lack of ground truth). The proposed DCNN is trained

on these datasets. Then, the trained model is used for classifying test sets of these 5 datasets. To evaluate the model's performance, a CM is generated each time for different test sets. The five CMs (with first row and column for BWV class, and second row and column for non-BWV class) for five different datasets are presented in Figure 5.42.

(a) PH2   **(b) Augmented PH2**   **(c) MED-NODE**

(d) Augmented MED-NODE   **(e) Derm7pt**
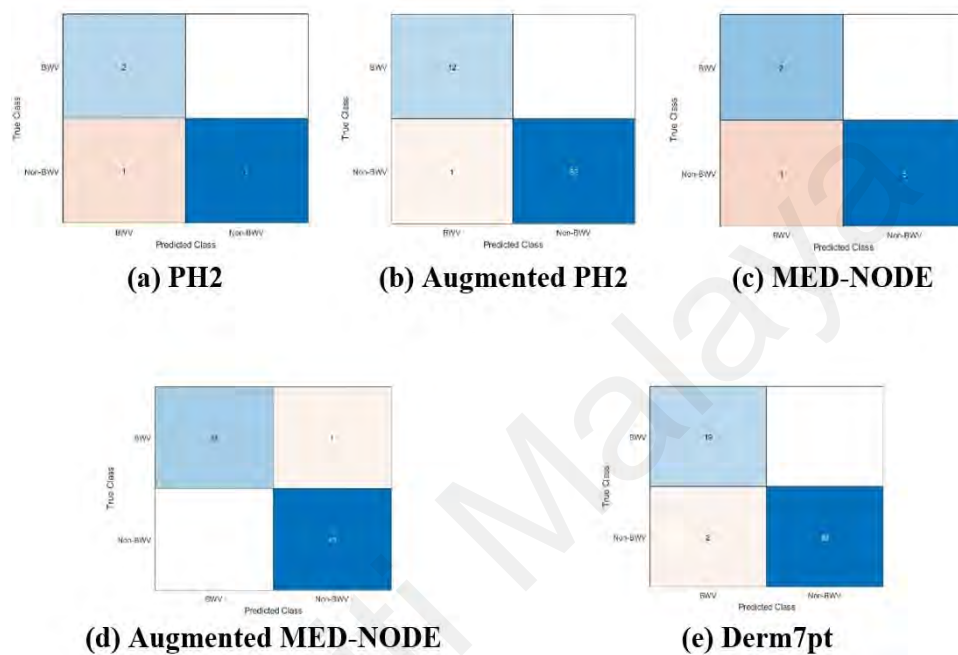
Figure 5.42: The CMs of the test sets of different datasets.

The evaluation metrics such as accuracy (*AC*), precision (*PR*), Sensitivity (*SE*), F1-score (*F1*), and specificity (*SP*) are applied to these CMs. After applying the formula of these evaluation metrics on those CMs, the result is reported in Table 5.25. From the literature, three different approaches (Celebi et al., 2008; Kropidlowski et. al. 2016; Madooei et al., 2019) are implemented for detecting BWV and they are applied to the same 5 datasets. The best effort is provided to avoid biases in the classification results. In Table 5.25, also, the results of the existing three methods are compared with the achieved results of the proposed DCNN. The area under the receiver operating characteristic curve (AUC) is also reported for all four methods. To draw the receiver operating characteristic curve, those CMs are used, and the AUC value is calculated.

Table 5.25: Classification results vs. other methods on test sets of different datasets.

Dataset	Method	AC	PR	SE	F1	SP	AUC
PH2	Celebi et al., 2008	70.00	100.00	40.00	57.14	100.00	65.78
	Kropidlowski et. al. 2016	70.00	50.00	33.33	40.00	85.71	66.50
	Madooei et al., 2019	80.00	100.00	50.00	66.67	100.00	77.00
	Proposed CNN	**90.00**	**100.00**	**66.67**	**80.00**	**100.00**	**86.40**
Augmented PH2	Celebi et al., 2008	90.77	91.67	68.75	78.57	97.96	89.00
	Kropidlowski et. al. 2016	87.69	75.00	64.29	69.23	94.12	84.30
	Madooei et al., 2019	92.31	83.33	76.92	80.00	96.15	90.70
	Proposed CNN	**98.46**	**100.00**	**92.31**	**96.00**	**100.00**	**94.50**
MED-NODE	Celebi et al., 2008	62.50	100.00	40.00	57.14	100.00	59.90
	Kropidlowski et. al. 2016	62.50	50.00	33.33	40.00	80.00	60.00
	Madooei et al., 2019	75.00	100.00	50.00	66.67	100.00	73.20
	Proposed CNN	**87.50**	**100.00**	**66.67**	**80.00**	**100.00**	**85.50**
Augmented MED-NODE	Celebi et al., 2008	90.91	85.71	80.00	82.76	95.00	88.00
	Kropidlowski et. al. 2016	87.27	71.43	76.92	74.07	90.48	85.80
	Madooei et al., 2019	92.73	78.57	91.67	84.62	93.02	90.10
	Proposed CNN	**98.18**	**92.86**	**100.00**	**96.30**	**97.62**	**96.45**
Derm7pt	Celebi et al., 2008	92.08	84.21	76.19	80.00	96.25	90.00
	Kropidlowski et. al. 2016	90.10	78.95	71.43	75.00	95.00	87.50
	Madooei et al., 2019	94.06	89.47	80.95	85.00	97.50	92.20
	Proposed CNN	**98.02**	**100.00**	**90.48**	**95.00**	**100.00**	**96.40**

Based on the information in Table 5.25, it is demonstrated that the proposed DCNN outperforms these existing three methods on the test sets of five different datasets. The multiple instance learning method which was proposed by Madooei et al. is in the second position for all five datasets. For the proposed DCNN, the average accuracy, precision, sensitivity, F1-score, specificity and AUC are respectively 94.43, 98.52, 83.13, 89.46, 99.52, and 91.85 in those datasets.

5.4.7   Applying Proposed DCNN Model on the Unfamiliar Test Set of ISIC2018

At the end, the proposed DCNN model for BWV detection is trained on the PH2, Med-Node, and Derm7pt datasets altogether and tested on a completely unfamiliar test set, which consists of randomly selected 100 dermoscopic images from the ISIC2018 dataset. However, this test set lacked dermoscopic BWV ground truth. The BWV ground truth for this test set is established with the help of Dermatology division of Universiti Malaya, Malaysia. There are 42 images with BWV and 58 images without BWV in this test set. The proposed model successfully classified 97 out of the 100 dermoscopic images based on the presence of BWV. The performance details are reported in Table 5.26.

Table 5.26: Evaluation of the proposed models on the test set of ISIC2018.

Method	BWV			Non-BWV			Overall		
	PR	RC	F1	PR	RC	F1	AC	mF1	wF1
Proposed Model (PReLU)	95	98	96	98	97	97	97	97	97
Proposed Model (ReLU)	78	86	82	89	83	86	84	84	84

5.5   Dermatological Feature: Dots-Globules

One of the features that dermatologists look for when examining skin lesions is the presence of dots-globules. Dots-globules are small, pigmented structures that can be seen under a dermatoscope. They are often seen in melanocytic lesions, which are lesions that arise from melanocytes, the cells that produce pigment in the skin. Dots are small, round, and uniform in size, while globules are larger and more irregular in shape. Their unpredictable pattern makes the traditional image processing techniques less useful to classify them. The presence of dots-globules in a skin lesion can indicate that it is a Melanoma, particularly if they are asymmetrically distributed or arranged in a radial pattern. Figure 5.43 shows the dots-globules on the dermoscopic images in terms of benign (absent), melanocytic (atypical), and non-melanocytic lesions (typical).

(a) in benign lesion　　(b) in melanocytic lesion　　(c) in non-melanocytic lesion
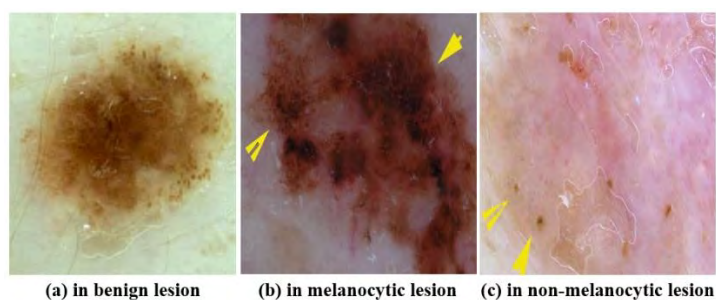
Figure 5.43: The presence of dots-globules in different skin lesions.

Here, a deep learning model with nested layers is designed to classify dots-globules into three classes by using three different datasets- PH2, augmented PH2, and Derm7pt (see Chapter 3: Research Methodology). Due to the lack of ground truth, other datasets cannot be used to train the network. Then, the experimental results are reported and compared with the existing methods from the literature. Figure 5.44 presents the overall process of dots-globules classification on dermoscopic images.
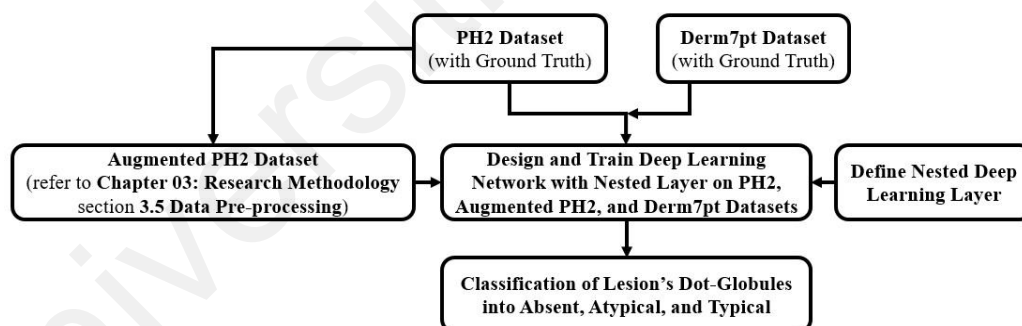
Figure 5.44: The overall classification process of dots-globules on the lesion images.

5.5.1　Defining Nested Deep Learning (DL) Layer

Nested deep learning (DL) layers, also known as hierarchical or nested architectures, refer to the integration of multiple levels of deep learning models within a single system. These layers simplify complex data understanding by progressively learning more abstract features, enhancing pattern recognition of dots-globules. This hierarchy captures

local and global details, aiding accurate predictions. Hierarchical structures compress data information as they ascend, saving computation and memory. They also support transfer learning, speeding up training when data is limited (for PH2 dataset). Hierarchical layers enhance model interpretability and robustness to reduce noise that is ideal for variable inputs. This architecture's flexibility and stacking capabilities empower tackling diverse tasks and input types effectively.

A DL model object is defined in the properties part of the layer declaration to create a custom layer that defines a layer graph, known as network composition. This technique can create a single custom layer that represents a block of learnable layers (a residual block), a network with control flow, or a network with loops.

Declare the network in the properties (Learnable, State) portion of the layer definition for nested networks that include both learnable and state parameters (i.e., networks with batch normalization or Long Short-Term Memory layers). A residual block is represented by a custom layer. The convolution, group normalization, ReLU, and addition layers make up the learnable block of layers in the custom layer residual block layer, which also includes a skip connection, an optional convolution layer, and a group normalization layer in the skip connection. A single input from the layer serves as both the input and output for each branch. Figure 5.45 highlights the residual block structure.
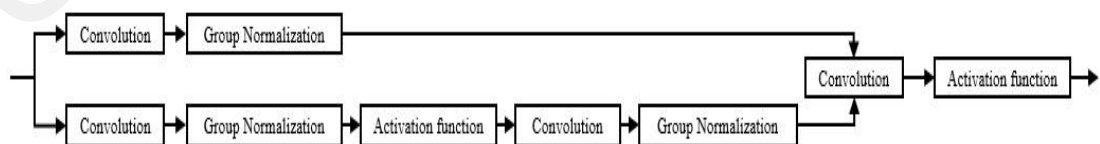
Figure 5.45: A residual block structure.

To create a custom deep learning layer, these steps are followed: naming the layer (e.g., DotsGlobulesLayer) for easy integration into the proposed network; declaring its properties, including learnable and state parameters; constructing a function outlining layer creation and property initialization; developing an initialize function to set parameter values during network initialization; defining forward functions for data propagation during prediction and training; implementing a reset state function for state parameter resetting; and designing a backward function specifying loss derivatives concerning input data and learnable parameters for backward propagation, or ensuring forward functions support deep learning array objects if no backward function is specified. Based on these steps, the pseudo code of custom deep learning layer is-

```
class DotsGlobulesLayer:
    Name = "DotsGlobulesLayer"
    Description = null
    Type = null
    Inputs = 1
    Outputs = 1
    LearnableParams = null
    StateParams = null

    constructor():
        # Initialize layer properties
        Name = "DotsGlobulesLayer"
        Description = null
        Type = null
        Inputs = 1
        Outputs = 1
        LearnableParams = null
        StateParams = null

    initialize():
        # Initialize learnable and state parameters
        # This function can be overridden to specify custom initialization

    forward(input_data):
        # Perform forward propagation
        # This function must be implemented to define how data passes forward through the layer
        # at prediction time and at training time

    reset_state():
```

```
    # Reset state parameters
    # This function can be overridden to specify custom reset behavior

backward(input_data, loss_derivatives):
    # Perform backward propagation
    # This function can be implemented to calculate the derivatives of the loss with respect
to the input data and the learnable parameters

# Usage:
layer = DotsGlobulesLayer()
layer.initialize()
output = layer.forward(input_data)
layer.backward(input_data, loss_derivatives)
```

5.5.2   Train Deep Learning Model with Nested Layers

Residual connections are a feature that is present in many convolutional neural network models. There are links in a neural network type known as a residual network that bypass the main neural network layers. The training of deeper networks is made possible by residual connections, which improve gradient flow inside the network. The deeper network might improve accuracy for more difficult tasks.

For residual connections to function, a neural network must have skip connections that directly link previous layers to later levels. These connections enable the network to learn residual mappings, which indicate the discrepancies between the desired output and the current output of a layer, rather than just depending on the output of each layer. The custom layers for residual blocks, which each have multiple convolutions, group normalization, and activation function layers with a skip connection, are used to train a network. Using the custom layer DotsGlobulesLayer, a residual network with six residual blocks is defined. The network consists of 15 layers and 32 filters. In Table 5.27, the different layers of the residual network are presented briefly which is inspired by a novel deep learning network: nested sparse network (Kim et al., 2018).

Table 5.27: The proposed Nested Deep Learning Network.

Layer	Name	Description (Different Parameters)
1st	Image Input	224×224×3 images with 'zero centers' normalization
2nd	2-D Convolution	32 7×7 convolutions with stride [2 2] and padding 'same'
3rd	Group Normalization	Group normalization
4th	Activation Function	Leaky ReLU
5th	2-D Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
6th	DotsGlobulesLayer	Residual block with 32 filters, stride 1
7th	DotsGlobulesLayer	Residual block with 32 filters, stride 1
8th	DotsGlobulesLayer	Residual block with 64 filters, stride 2, and skip convolution
9th	DotsGlobulesLayer	Residual block with 64 filters, stride 1
10th	DotsGlobulesLayer	Residual block with 128 filters, stride 2, and skip convolution
11th	DotsGlobulesLayer	Residual block with 128 filters, stride 1
12th	2-D Global Average Pooling	2-D global average pooling
13th	Fully Connected	3 fully connected layer
14th	Softmax	softmax
15th	Classification Output	crossentropyex

The determination of the choice of this proposed Nested Deep Learning Network Architecture was based on several factors. Firstly, the complexity and intricacy of the 'dots-globules' dermoscopic feature was considered. Given its nuanced characteristics and potential variations, a deep learning network with multiple layers and residual blocks was deemed necessary to capture and learn the intricate patterns effectively. Additionally, an extensive literature review was conducted, and existing deep learning architectures and techniques were analyzed and applied to similar image analysis tasks. This analysis helped identify the strengths and limitations of previous approaches (Li et al., 2022; Chen et al., 2022) and provided insights into the most effective design choices.

Moreover, preliminary experiments and empirical evaluations were performed using different network architectures and variations. These experiments involved training and testing the models on a representative dataset of dermoscopic images with "dots-globules"

to assess their performance in terms of accuracy, precision, recall, and other relevant metrics. The results of these experiments guided the selection of the nested deep learning network with six residual blocks as the most suitable choice for achieving superior detection and classification performance.

Overall, the determination of this proposed deep learning network design involved a combination of domain expertise, thorough research, comparative analysis of existing methods, and empirical evaluations to ensure an optimal solution for the specific task of detecting and classifying 'dots-globules' in dermoscopic images.

Next, the training options are specified. They are (a) training the network with a mini-batch size of 32, (b) shuffling the data every epoch, (c) validating the network once per epoch using the validation data, (d) producing output the network with the lowest validation loss, and (e) displaying the training progress in a plot and disable the verbose output. The Stochastic Gradient Descent (SGD) is used as a network training optimizer. Because its generalization performance is significantly better than Adam optimizer on image classification problems. To train the model on PH2, augmented PH2, and Derm7pt datasets, the datasets are divided into training sets (70%), validation sets (20%), and test sets (10%). The test performance is reported in the following section.

5.5.3 Outputs and Comparison for Dots-globules Detection Algorithm

After training the proposed model on training sets and validation sets of different datasets, the test sets are applied. Those test sets' results are reported in Table 5.28, Table 5.29, and Table 5.30. As well as other existing models' experimental results on the same test sets are reported for comparison purposes in these tables. A very limited amount of research work is involved directly to classify the dots-globules into different classes. However, for the sake of fair analysis and comparison, those existing works are

implemented in the same environment as the proposed model. The proposed model and existing methods such as Joanna (Jaworek-Korjakowska & Tadeusiewicz, 2013) and Ilias (Maglogiannis & Delibasis, 2015) are compared using different evaluation metrics such as sensitivity, specificity, precision, accuracy, and F1 score in Table 5.28, Table 5.29, and Table 5.30 using equation (7) to (11). In all tables the performances of the proposed Nested method (bolded) outperformed the performances of existing methods Joanna and Ilias. Joanna's performances are in the second position on all three tables. Ilias's proposed method was for two classes- atypical and typical dots-globules. Since Ilias's method did not work for absent class, in the comparison process absent class is not available.

Table 5.28: The proposed model and other models' experimental results on the PH2 dataset.

Model	Absent					Atypical					Typical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1
Joanna	60	73	58	61	60	59	65	61	60	60	70	83	78	75	72	63.00	73.67	65.67	65.33	64.00
Ilias	-	-	-	-	-	60	73	71	70	69	65	74	72	73	70	62.50	73.50	71.50	71.50	69.50
Nested	88	83	78	85	82	63	92	83	80	71	100	94	80	95	89	83.67	89.67	80.33	86.67	80.67

Table 5.29: The proposed model and other models' experimental results on the augmented PH2 dataset.

Model	Absent					Atypical					Typical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1
Joanna	62	75	60	65	63	61	67	63	61	62	71	83	77	76	73	64.67	75.00	66.67	67.33	66.00
Ilias	-	-	-	-	-	75	74	76	75	72	72	71	74	73	70	73.50	72.50	75.00	74.00	71.00
Nested	88	87	83	88	86	83	93	83	90	83	82	97	90	93	86	84.33	92.33	85.33	90.33	85.00

Table 5.30: The proposed model and other models' experimental results on the Derm7pt dataset.

Model	Absent					Atypical					Typical					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1
Joanna	63	76	62	68	66	62	69	66	63	64	73	85	79	78	76	66.00	76.67	72.00	69.67	68.67
Ilias	-	-	-	-	-	78	75	77	77	77	76	75	78	76	76	77	75	77.50	76.50	76.50
Nested	87	96	85	94	86	86	90	88	88	87	91	95	90	94	90	88	94	88.25	92.00	87.67

The proposed deep learning model with nested layers outperforms (the best results are bolded in all three tables) the existing approaches in all three datasets (PH2, augmented PH2, and Derm7pt datasets). The difference is significant between the proposed model's performance and others' performance.

5.5.4 Applying Proposed Nested Model on the Unfamiliar Test Set of ISIC2018

Finally, the proposed Nested model for dermoscopic feature dots-globules detection is trained on both PH2 and Derm7pt datasets combined and tested on a completely unfamiliar test set, which consists of randomly selected 100 dermoscopic images from the ISIC2018 dataset. There are 20 images with atypical dots-globules, 15 images with typical dots-globules, and 65 images without dots-globules in this test set. The proposed model successfully classified 95 out of the 100 dermoscopic images based on the dots-globules. The performance details are reported in Table 5.31.

Table 5.31: Evaluation of the proposed model on the test set of ISIC2018.

Method	Absent			Atypical			Typical			Overall		
	PR	RC	F1	PR	RC	F1	PR	RC	F1	AC	mF1	wF1
Proposed Nested	100	95	98	95	95	95	78	93	85	95	93	95
Jonna et al.	95	65	77	80	60	69	24	67	36	64	60	69

5.6   Dermatological Feature: Lesion Colors

Color variation is an important dermatological feature to diagnose Melanoma. Melanoma is a type of skin cancer that arises from melanocytes, which are the cells that produce the pigment that gives color to the skin. Melanomas can have a range of colors, including brown, black, red, white, blue, or a combination of colors. If the region of interest has multiple colors or shades within the same lesion is often red flagged by dermatologists. These colors may appear irregular or patchy, with different areas of the mole having different colors. Figure 5.46 shows a skin lesion with various colors from PH2 dataset.
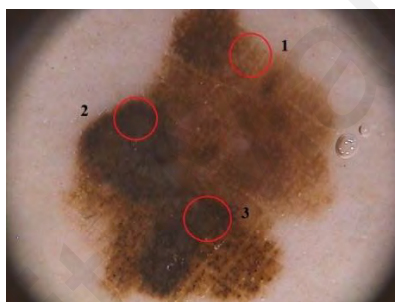
Figure 5.46: A skin lesion of PH2 dataset with multiple colors in the region of interest (1 is light brown, 2 is dark brown, and 3 is black).

Additionally, Melanomas may have a color that is different from the surrounding skin, or they may have a color that changes over time. It is important to note that color variation is one of several features dermatologists use to detect Melanoma. Detecting the presence of different colors in a skin lesion often requires the expertise of a trained observer. This process is manual and relies on the trained eyes of medical professionals to identify subtle color variations that could indicate Melanoma. Here, different colors in a lesion are detected and counted automatically to assist dermatologists and experts in a better understanding of skin conditions. This color detection process in dermoscopic images is difficult to identify by the naked eyes of experts and non-experts. This section of this

thesis is reported to fully automate this process using an annotated (PH2) dataset and a non-annotated dataset (MED-NODE). Since there is no color information provided by the MED-NODE dataset, pre-processing is required to annotate this dataset. Moreover, while working on this thesis, no existing works have been found that directly address the classification of skin lesion colors. Figure 5.47 shows the different steps of skin lesion color detection and classification process.
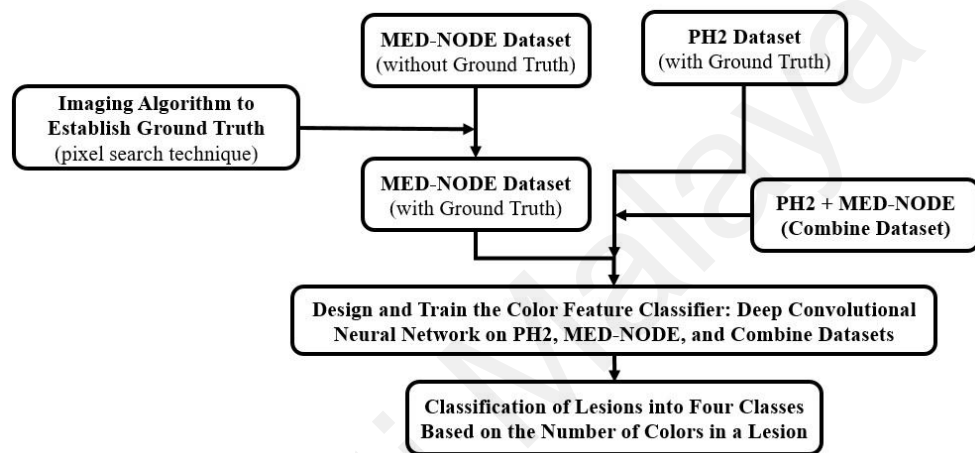
Figure 5.47: The overall process of dermoscopic color feature detection and classification process.

5.6.1 Image Acquisition for Color Feature

MED-NODE has 170 macroscopic images without information of lesions' color (the common colors of skin lesion to diagnose Melanoma based on the annotated information of PH2 dataset). Hence, the goal of this section is to provide color (RGB color code for light brown, dark brown, red, blue-gray, white, and black) information on those lesions. There is no color palette exist for different lesion colors as the BWV (see Figure 5.35). As a reference PH2 dataset is used to classify the lesion of MED-NODE dataset based on the number of colors. After the final performance, it is found that light brown is 113, dark brown is 93, red is 56, blue gray is 25, white is 36, and black is 45 times in 170

images of lesions. These dermoscopic and macroscopic images are categorized based on the number of containing colors in Table 5.32.

Table 5.32: The lesion image categorization on PH2 and MED-NODE.

Datasets	Total Number	Any One Color	Any Two Colors	Any Three Colors	Any Four Colors	Any Five Colors	Any Six Colors
PH2	200	54	107	23	13	3	-
MED-NODE	170	64	53	30	13	4	6
Combine	370	118	160	53	26	7	6

5.6.2 Establishing the Ground Truth for MED-NODE

Since PH2 has the ground truth of different colors on the lesions that annotated information is used to establishing the ground truth for MED-NODE. The lesion's colors were identified as light brown, dark brown, red, blue-gray, white, and black on PH2 dataset. The RGB value for those colors was absent in PH2. However, the RGB color channel value is found in the color chart (Lamb et al., 2022). The six common dermoscopic image colors (based on PH2) are taken to find the RGB value (Fang et al., 2017). Table 5.33 shows the minimum (min) and maximum (max) values for those six common colors' RGB channels.

Table 5.33: The minimum and maximum value of RGB color channel based on the color chart.

Name of Channel	Light Brown		Dark Brown		Red		Blue Gray		White		Black	
	Min	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
Red	210	222	139	210	128	255	112	176	245	255	0	104
Green	180	184	69	133	0	99	128	196	245	255	0	104
Blue	135	140	19	63	0	71	144	222	245	255	0	104

From Table 5.33, the minimum and maximum RGB color channel values are taken to justify whether demoscopic image colors fit in that range or not. For this reason, several

images are taken from both PH2 and MED-NODE datasets. In those images, several random pixels are selected by visual prediction to fit into Table 5.33 provided RGB channel range for the six colors. Table 5.34 presents three demonstrations of random color selection and fitting into the RGB channel range of Table 5.33. For each image, randomly 10 points are selected to see those 10 points' pixel value whether fit into predetermined min max RGB range or not.

Table 5.34: Demonstration of fitting into RGB channel range (based on Table 5.33) of three lesion images.

Input Image	Name	Dataset	Name of Channel	Light Brown	Dark Brown	Red	Blue Gray	White	Black
	IMD417.bmp	PH2	Red	213	173	-	120	249	62
			Green	144	95	-	130	249	49
			Blue	103	59	-	150	250	41
			RGB image (color patch)						
	1685446.JPG	MED-NODE	Red	-	190	-	161	-	34
			Green	-	115	-	163	-	30
			Blue	-	22	-	162	-	27
			RGB image (color patch)						
	1761976-h.JPG	MED-NODE	Red	184	140	129	168	-	43
			Green	180	87	55	186	-	32
			Blue	105	63	43	164	-	30
			RGB image (color patch)						

Based on the RGB channels' range, the value of six colors for the dermoscopic and macroscopic images are found. The pixel values of IMD417.bmp, 1685446.JPG, and 1761976-h.JPG lesion images of Table 5.34 are matched with the provided information (the color range) in Table 5.33. It seems the predetermined min-max RBG range is capable to cover the pixel values of all images of PH2 and MED-NODE datasets.

By using the provided six colors of RGB channels' range in Table 5.33, separately those six colors are detected on all MED-NODE dataset's images (see Table 5.32). All images of the MED-NODE dataset are classified into six classes based on the containing total number of colors. The algorithm for detecting six colors on the lesions using the RGB values is-

---

Step 1: Define color thresholds for each color
red_threshold = [lower_red_value, upper_red_value]
white_threshold = [lower_white_value, upper_white_value]
black_threshold = [lower_black_value, upper_black_value]
blue_gray_threshold = [lower_blue_gray_value, upper_blue_gray_value]
light_brown_threshold = [lower_light_brown_value, upper_light_brown_value]
dark_brown_threshold = [lower_dark_brown_value, upper_dark_brown_value]

Step 2: Loop through each pixel in the dermoscopic image
for each pixel in the image:

- *Get the color values of the pixel*
  red = pixel.red
  green = pixel.green
  blue = pixel.blue

- *Check if the pixel falls within the red color range*
  if red >= red_threshold[0] and red <= red_threshold[1]:
  Pixel is within the red color range

- *Check if the pixel falls within the white color range*
  if red >= white_threshold[0] and red <= white_threshold[1] and
  green >= white_threshold[0] and green <= white_threshold[1] and
  blue >= white_threshold[0] and blue <= white_threshold[1]:
  Pixel is within the white color range

- *Check if the pixel falls within the black color range*
  if red >= black_threshold[0] and red <= black_threshold[1] and
  green >= black_threshold[0] and green <= black_threshold[1] and
  blue >= black_threshold[0] and blue <= black_threshold[1]:
  Pixel is within the black color range

- *Check if the pixel falls within the blue-gray color range*
  if red >= blue_gray_threshold[0] and red <= blue_gray_threshold[1] and
  green >= blue_gray_threshold[0] and green <= blue_gray_threshold[1] and
  blue >= blue_gray_threshold[0] and blue <= blue_gray_threshold[1]:
  Pixel is within the blue-gray color range

- *Check if the pixel falls within the light brown color range*
  if red >= light_brown_threshold[0] and red <= light_brown_threshold[1] and
  green >= light_brown_threshold[0] and green <= light_brown_threshold[1] and
  blue >= light_brown_threshold[0] and blue <= light_brown_threshold[1]:
  Pixel is within the light brown color range
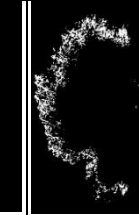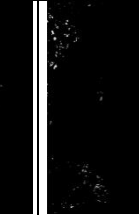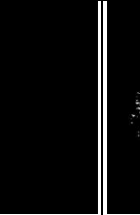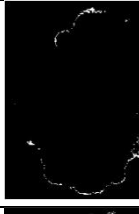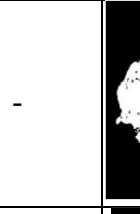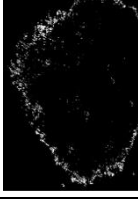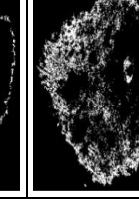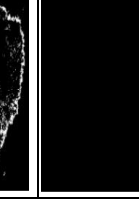
- *Check if the pixel falls within the dark brown color range*
  if red >= dark_brown_threshold[0] and red <= dark_brown_threshold[1] and
  green >= dark_brown_threshold[0] and green <= dark_brown_threshold[1] and
  blue >= dark_brown_threshold[0] and blue <= dark_brown_threshold[1]:
  Pixel is within the dark brown color range

Step 3: *After processing all pixels, the detected colors are gathered and performed further analysis or actions.*

Finally, Table 5.35 presents a demonstration to illustrate the color detection process of the MED-NODE dataset using a computer program. Above mentioned the pseudo code belongs to this computer program. This is how a non-annotated dataset becomes annotated dataset. In this table, the column 2 to 7 present the mask version of lesion images for the corresponding color. The white pixels represent the specific color for each column.

Table 5.35: Detecting six colors on macroscopic images of MED-NODE dataset using RGB value.

Images	Light Brown	Dark Brown	Red	Blue Gray	White	Black
	-		-		-	
					-	

5.6.3   Designing and Training the Color Feature Classifier

The same deep learning model with a custom deep learning layer was used in this section as the 5.4.2 section (which was designed for BWV detection). This deep learning model was trained to detect the color of BWV. Since this model gives a satisfactory performance that is why the same model is trained for detecting different colors in the lesions. The only difference is the number of classes for prediction. It is for two classes in the BWV detection algorithm, and it is for four classes in this color detection algorithm. Hence, instead of two, four is assigned in the number of classes inside the classification layer of the proposed deep learning model. The limited amount of training data is the reason for using four classes instead of six classes for six different colors. All lesion images which have four, five, and six colors are combined into one class because of their small number. Table 5.36 represents the class of dataset based on the number of skin lesion colors.

Table 5.36: The classification process for skin lesion color.

Existing number of colors in a lesion	The class name of dataset
If there is one color in a lesion (example: red)	1
If there are two colors in a lesion (example: red, dark)	2
If there are three colors in a lesion (example: red, dark brown, light brown)	3
If there is four or more than four colors in a lesion	4

In PH2 and MED-NODE datasets, the total numbers of lesion images with 4 colors, 5 colors, and 6 colors together are respectively 16 and 23 only. The small amount of image data for the different classes is not sufficient to train a deep-learning model. This is a reason for their combination in a single class. The PH2, MED-NODE, and combine (PH2 + MED-NODE) datasets are involved to train the proposed network. The image data of each dataset is divided into three sets such as training set (70%), validation set (25%), and

test set (5%). A total of 95% including training and validation sets involves directly in the training process. The network training option remained the same as previously applied for detecting BWV using deep convolutional neural network.

5.6.4   Experimentation Results and Comparison for Color Feature Classifier

Since there is no similar analytical technique for skin lesion color in the existing literature, experimental results are compared with another (from ablation study) proposed model's experimental results. Different activation functions have varying effects depending on the data and the specific problem being addressed. The ablation study was involved with other activation functions such as: Clip ReLU, Leaky ReLU, ELU, and Tanh. Only the remarkable two functions' performances were reported here.

Instead of ReLU activation function layers, custom PReLU layers are used in the same proposed model to do ablation study. The proposed model (PReLU) and proposed model (ReLU) are trained on PH2, MED-NODE, and combined (PH2 + MED-NODE) datasets. Their experimental results are compared in Table 5.37, Table 5.38, and Table 5.39. These tables are reported based on evaluation metrics such as- Sensitivity (SE), Specificity (SP), Precision (PR), Accuracy (AC), and F1-score (F1) using equations (7) to (11). It is important to note that the performance improvement observed with PReLU might be specific to these lesion color datasets and the complexity of the dermoscopic image classification task. The analytical results indicate that the proposed model with custom PReLU layers has better performance in three datasets. The difference between the two models' performance is significant across all comparisons. The performance improvement was observed when using a custom PReLU layer instead of a ReLU activation function in proposed CNN model for dermoscopic image classification can be attributed to several factors.

Table 5.37: The experimental result on PH2 Dataset.

Models	One Color					Two Colors					Three Colors					Four Colors					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1
Proposed (ReLU)	80	93	80	90	80	91	89	91	90	91	100	100	100	100	100	100	100	100	100	100	92.75	95.5	92.75	95	92.75
Proposed (PReLU)	83	100	100	95	91	100	90	91	95	95	100	100	100	100	100	100	100	100	100	100	95.75	97.5	97.75	97.5	96.5

Table 5.38: The experimental result on MED-NODE Dataset.

Models	One Color					Two Colors					Three Colors					Four Colors					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1
Proposed (ReLU)	83	90	83	88	83	80	91	80	88	80	100	100	100	100	100	100	100	100	100	100	90.75	95.25	90.75	94	90.75
Proposed (PReLU)	100	91	83	94	91	83	100	100	94	91	100	100	100	100	100	100	100	100	100	100	95.75	97.75	95.75	97	95.5

Table 5.39: The experimental result on Combine Dataset.

Models	One Color					Two Colors					Three Colors					Four Colors					Average				
	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1	SE	SP	PR	AC	F1
Proposed (ReLU)	90	89	75	89	82	80	82	75	81	77	80	97	80	95	80	57	100	100	92	73	76.75	92	82.5	89.25	78
Proposed (PReLU)	100	93	83	95	91	93	91	88	92	90	71	100	100	95	83	80	100	100	97	89	86	96	92.75	94.75	88.25

The ReLU activation function has gained popularity due to its effectiveness in mitigating the vanishing gradient problem. However, a limitation known as the "dying ReLU" (Lu, 2020) phenomenon emerges, wherein certain neurons output zero for inputs less than or equal to zero, rendering them inactive. In response to this concern, the PReLU introduces an adaptive negative slope parameter, thereby facilitating intricate representation acquisition and circumventing the issue of dormant neurons. By incorporating additional learnable parameters, PReLU augments the model's capacity to capture intricate nonlinear relationships intrinsic to the data. The acquisition of a learnable negative slope not only allows PReLU to discern finer nuances present in dermoscopic images but also contributes to heightened performance, particularly in intricate tasks such as image analysis. Addressing the gradient saturation predicament encountered by ReLU, PReLU rectifies this by affording a non-zero gradient for negative inputs. This adaptive adjustment enables a more informative gradient propagation throughout the network, consequently expediting the learning process and fostering enhanced convergence. Because of this expedited learning process and refined gradient flow, PReLU demonstrates superior potential in augmenting classification performance.

5.6.5 Applying Proposed Model on the Unfamiliar Test Set of ISIC2018

Finally, the proposed model is tested on a completely unfamiliar test set which is the random 100 selected dermoscopic images from ISIC2018 dataset. There are 5 images in "One Color", 28 images in "Two Colors", 39 images in "Three Colors", and 28 images in "Four Colors" classes. Since the proposed model is trained in four classes, lesion images are classified into four color classes. If any lesion has four or more colors, that lesion is classified as "Four Colors" class. However, this test set lacked dermoscopic color ground truth. These test set's color ground truth is established with the help of the Cancer Research Lab of Bangladesh Cancer Research Institute, Bangladesh. These 100

dermoscopic images are classified by proposed model and the output is evaluated by different evaluation metrics. The proposed model successfully detected 97 out of 100 dermoscopic images' the number of colors. The performance details are reported in Table 5.40.

Table 5.40: Evaluation of the proposed models on the test set of ISIC2018.

Method	One Color			Two Colors			Three Colors			Four Colors			Overall		
	PR	RC	F1	PR	RC	F1	PR	RC	F1	PR	RC	F1	AC	mF1	wF1
Proposed Model (PReLU)	71	100	83	100	89	94	98	100	99	100	100	100	97	94	97
Proposed Model (ReLU)	71	100	83	92	86	89	90	90	90	93	93	93	90	89	90

5.7 Fusion of Five Dermoscopic Features to Diagnose Melanoma

All five dermoscopic features are extracted from a completely unfamiliar test set of the ISIC2018 dataset. This test set comprises 100 skin lesion images, each accompanied by clinical diagnosis ground truth. The three features - asymmetry, pigment network, and BWV - reported in Tables 5.41 to 5.44 are subjected to the clinically practiced 3-point checklist method for melanoma detection. The rules-based 3-point checklist method successfully detects 13 out of 20 melanomas, resulting in a 65% success rate, which increases to 100% when the other two important features are considered for Melanoma diagnosis.

When the fusion of five features (Phase 1) is applied to these 100 test images to implement clinically practiced early Melanoma detection method, the overall accuracy reaches 96%, the weighted-F1 score is 96%, and the recall for Melanoma is 100%. Evaluation metrics on the performance of Phase 1 (Feature Engineering) are presented in Table 5.45. Furthermore, Phase 1 is compared with Phase 2 (Classification) in the Chapter 6.

Table 5.41: Applying the proposed methods on the test set (1 to 25 of 100 dermoscopic images) to compare Phase 1 against the clinical Ground Truth (GT).

ISIC Archive	Clinical Method: Three-point-checklist						Additional Two Features				Diagnose	
	Asymmetry		Pigment Network		BWV		Dots-Globules		Lesion Color			
	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Phase: 1	Clinical GT
ISIC_0012255	S	S	*Ty*	At	*Yes*	No	*Ty*	Ab	1	1	Benign	Ty
ISIC_0012346	S	S	Ty	Ty	No	No	Ab	Ab	1	1	Benign	Ty
ISIC_0012576	H	H	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0012585	A	A	*Ty*	At	Yes	Yes	Ty	Ty	3	3	Benign	Ty
ISIC_0012623	*S*	A	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0012627	S	S	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0012633	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3 +**	**4**	**Mel***	**Mel**
ISIC_0012643	S	S	At	At	Yes	Yes	Ab	Ab	3	3	Benign	Ty
ISIC_0015294	A	A	At	At	Yes	Yes	Ab	Ab	3	3	Benign	Ty
ISIC_0015351	S	S	At	At	No	No	Ty	Ty	3	3	Benign	Ty
ISIC_0015370	A	A	At	At	Yes	Yes	Ab	Ab	3 +	5	Benign	Ty
ISIC_0015462	**A**	**A**	***At***	**Ty**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel****	**Mel**
ISIC_0015480	H	H	At	At	Yes	Yes	Ab	Ab	2	2	Benign	Ty
ISIC_0015492	H	H	At	At	Yes	Yes	Ab	Ab	3	3	Benign	Ty
ISIC_0015518	**A**	**A**	***At***	**Ty**	**Yes**	**Yes**	**At**	**At**	**3 +**	**5**	**Mel****	**Mel**
ISIC_0015552	S	S	Ty	Ty	No	No	Ab	Ab	2	2	Benign	At
ISIC_0015590	S	S	At	At	Yes	Yes	Ab	Ab	3	3	Benign	Ty
ISIC_0015634	A	A	At	At	No	No	Ty	Ty	2	2	Benign	Ty
ISIC_0016351	S	S	At	At	No	No	Ty	Ty	*1*	2	Benign	Ty
ISIC_0016714	A	A	At	At	No	No	Ty	Ty	2	2	Benign	Ty
ISIC_0016804	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0017341	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0017398	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3 +**	**4**	**Mel***	**Mel**
ISIC_0017399	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel***	**Mel**
ISIC_0017460	**A**	**A**	**Ty**	**Ty**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel****	**Mel**

Note: S = Symmetry, H = Half Symmetry, A = Asymmetry, Ty = Typical, At = Atypical, Mel = Melanoma, "Yes" = BWV, No = non-BWV, Ab = Absent, and "3 +" = more than three colors. In Phase: 1, clinically practiced "three-point-checklist" method (asymmetry, pigment network, and BWV) is followed at first, and additionally two more features (dots-globules, and lesion color) are analyzed to identify suspicious "Melanoma" only. "*" means Melanoma is detected by the "three-point-checklist" method, and "**" means Melanoma is detected by "three-point-checklist" + two other features.

Table 5.42: Applying the proposed methods on the test set (26 to 50 of 100 dermoscopic images) to compare Phase 1 against the clinical Ground Truth (GT).

ISIC Archive	Clinical Method: Three-point-checklist						Additional Two Features				Diagnose	
	Asymmetry		Pigment Network		BWV		Dots-Globules		Lesion Color			
	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Phase: 1	Clinical GT
ISIC_0017474	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel***	**Mel**
ISIC_0017702	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0017755	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0018111	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0018179	S	S	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0018248	S	S	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0018375	S	S	Ty	Ty	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0018472	H	H	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0018521	H	H	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0018556	S	S	Ty	Ty	*No*	Yes	Ab	Ab	1	1	Benign	Ty
ISIC_0018611	S	S	At	At	No	No	Ab	Ab	1	1	Benign	Ty
ISIC_0018680	S	S	Ty	Ty	No	No	Ab	Ab	1	1	Benign	Ty
ISIC_0019049	S	S	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0019309	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0019334	**A**	**A**	**At**	**At**	**Yes**	**Yes**	***Ty***	**At**	**3 +**	**6**	**Mel***	**Mel**
ISIC_0019723	S	S	At	At	Yes	Yes	Ab	Ab	**3** +	4	Benign	Ty
ISIC_0019794	S	S	Ty	Ty	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0019883	H	H	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0020233	H	H	Ty	Ty	No	No	Ty	Ty	3	3	Benign	Ty
ISIC_0020418	S	S	Y	At	No	No	Ab	Ab	3 +	4	Benign	Ty
ISIC_0020861	S	S	Y	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0020893	S	S	Y	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0020953	S	S	Ty	Ty	No	No	Ty	Ty	2	2	Benign	Ty
ISIC_0020999	H	H	*Ty*	At	No	No	Ty	Ty	3 +	4	Benign	At
ISIC_0021037	S	S	Ty	Ty	No	No	Ab	Ab	2	2	Benign	At

Note: S = Symmetry, H = Half Symmetry, A = Asymmetry, Ty = Typical, At = Atypical, Mel = Melanoma, "Yes" = BWV, No = non-BWV, Ab = Absent, and "3 +" = more than three colors. In Phase: 1, clinically practiced "three-point-checklist" method (asymmetry, pigment network, and BWV) is followed at first, and additionally two more features (dots-globules, and lesion color) are analyzed to identify suspicious "Melanoma" only. "*" means Melanoma is detected by the "three-point-checklist" method, and "**" means Melanoma is detected by "three-point-checklist" + two other features.

Table 5.43: Applying the proposed methods on the test set (51 to 75 of 100 dermoscopic images) to compare Phase 1 against the clinical Ground Truth (GT).

ISIC Archive	Asymmetry Pred.	Asymmetry GT	Pigment Network Pred.	Pigment Network GT	BWV Pred.	BWV GT	Dots-Globules Pred.	Dots-Globules GT	Lesion Color Pred.	Lesion Color GT	Diagnose Phase: 1	Diagnose Clinical GT
ISIC_0021041	*H*	A	*Ty*	At	No	No	*Ty*	Ab	3 +	4	Benign	At
ISIC_0021152	H	H	At	At	Yes	Yes	Ty	Ty	3 +	5	Benign	Ty
ISIC_0021158	H	H	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0021202	S	S	Ty	Ty	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0021251	S	S	At	At	No	No	Ab	Ab	3 +	4	Benign	Ty
ISIC_0021448	A	A	At	At	Yes	Yes	Ab	Ab	3 +	4	*Mel**	Ty
ISIC_0021449	A	A	At	At	Yes	Yes	Ab	Ab	3 +	4	*Mel**	Ty
ISIC_0021504	S	S	At	At	Yes	Yes	Ab	Ab	3	3	Benign	Ty
ISIC_0021714	A	A	Ty	Ty	Yes	Yes	Ab	Ab	3 +	5	Benign	At
ISIC_0021762	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0021816	**A**	**A**	*At*	**Ty**	**Yes**	**Yes**	**At**	**At**	**3 +**	**5**	**Mel****	**Mel**
ISIC_0021904	H	H	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0021914	S	S	Ty	Ty	No	No	Ab	Ab	**3 +**	6	Benign	At
ISIC_0021990	S	S	Ty	Ty	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0022029	H	H	At	At	Yes	Yes	Ab	Ab	**3 +**	5	Benign	Ty
ISIC_0022039	S	S	Ty	Ty	No	No	*At*	Ty	3	3	Benign	Ty
ISIC_0022147	H	H	Ty	Ty	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0022192	H	H	Ty	Ty	No	No	Ab	Ab	3 +	5	Benign	At
ISIC_0022221	S	S	*At*	Ty	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0022657	H	H	At	At	No	No	Ab	Ab	3 +	4	Benign	Ty
ISIC_0022738	S	S	At	At	Yes	Yes	Ab	Ab	3	3	Benign	Ty
ISIC_0023371	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3 +**	**4**	**Mel***	**Mel**
ISIC_0023508	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3 +**	**6**	**Mel***	**Mel**
ISIC_0023628	H	H	At	At	No	No	Ab	Ab	3 +	4	Benign	Ty
ISIC_0023678	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty

Note: S = Symmetry, H = Half Symmetry, A = Asymmetry, Ty = Typical, At = Atypical, Mel = Melanoma, "Yes" = BWV, No = non-BWV, Ab = Absent, and "3 +" = more than three colors. In Phase: 1, clinically practiced "three-point-checklist" method (asymmetry, pigment network, and BWV) is followed at first, and additionally two more features (dots-globules, and lesion color) are analyzed to identify suspicious "Melanoma" only. "*" means Melanoma is detected by the "three-point-checklist" method, and "**" means Melanoma is detected by "three-point-checklist" + two other features.

Table 5.44: Applying the proposed methods on the test set (76 to 100 of 100 dermoscopic images) to compare Phase 1 against the clinical Ground Truth (GT).

| ISIC Archive | Clinical Method: Three-point-checklist | | | | | | Additional Two Features | | | | Diagnose | |
| | Asymmetry | | Pigment Network | | BWV | | Dots-Globules | | Lesion Color | | | |
	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Phase: 1	Clinical GT
ISIC_0023755	**A**	**A**	**Ty**	**Ty**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel****	**Mel**
ISIC_0023831	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0023900	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	***3***	**2**	**Mel***	**Mel**
ISIC_0023904	S	S	*Ty*	At	No	No	Ty	Ty	*1*	2	Benign	Ty
ISIC_0023924	**A**	**A**	**Ty**	**Ty**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel****	**Mel**
ISIC_0023936	S	S	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0024135	S	S	At	At	Yes	Yes	Ty	Ty	3	3	Benign	Ty
ISIC_0036073	**A**	**A**	**Ty**	**Ty**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel****	**Mel**
ISIC_0036085	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3 +**	**4**	**Mel***	**Mel**
ISIC_0036098	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3 +**	**5**	**Mel***	**Mel**
ISIC_0036101	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel***	**Mel**
ISIC_0036121	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel***	**Mel**
ISIC_0036147	S	S	At	At	Yes	Yes	Ab	Ab	3 +	4	Benign	Ty
ISIC_0036174	S	S	At	At	Yes	Yes	Ty	Ty	3	3	Benign	Ty
ISIC_0036206	H	H	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0036236	H	H	Ty	Ty	No	No	Ty	Ty	3	3	Benign	Ty
ISIC_0036237	S	S	Ty	Ty	Yes	Yes	Ab	Ab	3 +	4	Benign	Ty
ISIC_0036240	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0036247	S	S	Ty	Ty	No	No	Ab	Ab	2	2	Benign	At
ISIC_0036281	**A**	**A**	**At**	**At**	**Yes**	**Yes**	**At**	**At**	**3**	**3**	**Mel***	**Mel**
ISIC_0036291	H	H	At	At	No	No	Ab	Ab	3	3	Benign	Ty
ISIC_0036306	A	A	At	At	Yes	Yes	Ab	Ab	3 +	6	*Mel**	Ty
ISIC_0036321	A	A	At	At	Yes	Yes	Ty	Ty	3 +	4	*Mel**	Ty
ISIC_0036328	S	S	At	At	No	No	Ab	Ab	2	2	Benign	Ty
ISIC_0036333	S	S	*Ty*	At	*Yes*	No	*Ty*	Ab	3	3	Benign	Ty

Note: S = Symmetry, H = Half Symmetry, A = Asymmetry, Ty = Typical, At = Atypical, Mel = Melanoma, "Yes" = BWV, No = non-BWV, Ab = Absent, and "3 +" = more than three colors. In Phase: 1, clinically practiced "three-point-checklist" method (asymmetry, pigment network, and BWV) is followed at first, and additionally two more features (dots-globules, and lesion color) are analyzed to identify suspicious "Melanoma" only. "*" means Melanoma is detected by the "three-point-checklist" method, and "**" means Melanoma is detected by "three-point-checklist" + two other features.

Table 5.45: Evaluation of Phase 1 (Feature Engineering) on the test set of ISIC2018.

Method	Melanoma			Non-Melanoma			Overall		
	PR	RC	F1	PR	RC	F1	AC	mF1	wF1
Phase 1 (3-point checklist)	76	65	70	92	95	93	89	82	89
Phase 1 (All five features)	83	100	91	100	95	97	96	94	96

5.8   Chapter Summary

This chapter identifies, examines, and categorizes each of the five key dermatological features: asymmetry, pigment network (PN), blue white veil (BWV), dots-globules, and lesion colors or color distribution for Melanoma diagnosis. Dermatological features are extracted and analyzed using image processing and pattern recognition techniques, and several deep learning models are employed to categorize the data into distinct classes. Each feature detection technique (except lesion colors feature) is evaluated against previously published works in the literature.

Two methods were proposed for dermatological asymmetry analysis. The first method involved lesion segmentations (Graph Cut and Semantic segmentation network) and geometry-based analysis. The second method was a deep learning-based lesion shape analysis to determine dermatological asymmetric properties. PH2, MED-NODE, and ISIC2016 datasets were used to analysis dermatological asymmetry feature. MED-NODE and ISIC2016 became annotated datasets after the analysis process.

In the PN identification, a two-step process was performed. The first step involved the proposed directional imaging algorithm to detect the PN on dermoscopic images and isolate it from the background. The second step involved the classification of the isolated PN into atypical PN (APN) and typical PN (TPN) using Bag of Features (BoF) and the proposed Convolutional Neural Network (CNN). PH2, ISIC2019, ISIC2020 and MED-NODE datasets were applied here to isolate the PN from the background.

A novel imaging technique was proposed to measure the color of the BWV on the skin lesions, transforming non-annotated MED-NODE dataset into annotated dataset. Another proposed CNN was used to detect BWV on the dermoscopic images, and the explainable artificial intelligence (XAI) LIME was involved in explaining the decisions made by the proposed CNN. PH2, MED-NODE, and Derm7pt datasets were engaged in this feature analysis process.

A nested deep learning layer was introduced, and a deep learning network with nested layers was proposed to detect and classify dots-globules into atypical, typical, and absent categories on the skin lesions. PH2 and Derm7pt datasets were involved in this classification process.

Another novel imaging algorithm based on RGB color range was proposed to detect different colors on the dermoscopic images. The proposed algorithm helped in transforming non-annotated MED-NODE dataset into the annotated one. The proposed CNN for BWV detection was used to detect multiple colors on lesions and classify lesion images into four different classes based on the number of colors present in the lesion. PH2 and MED-NODE were utilized in this dermatological color analysis process.

The fusion of these five features is considered Step 2: Feature Extraction of Phase 1: Feature Engineering in this thesis (refer to Figure 5.1). The primary objective of analyzing dermoscopic features is to boost the diagnostic decisions made by dermatologists and skin specialists, enhancing the accuracy and robustness of their conclusions. The outcomes of feature analysis presented in this chapter do not directly diagnose Melanoma; rather, they provide insights into the skin condition of the region of interest for both experts and non-experts.

In the following chapter, a method founded on a Classification approach and utilizing deep learning models will be introduced (Phase 2: Classification). This method will be capable of detecting Melanoma from dermoscopic images. A convolutional neural network has been meticulously designed and trained using multiple datasets to effectively detect Melanoma (refer to Chapter 6: Classification Based Melanoma Detection).

CHAPTER 6: CLASSIFICATION BASED MELANOMA DETECTION

6.1   Background

This research has two phases. The first phase is "Feature Engineering" (including lesion segmentation and feature analysis), and the second phase is "Classification" (deep learning-based approach). Chapter 4 and Chapter 5 discussed lesion segmentation and feature analysis (extraction and classification). This first phase is more into explaining the skin condition to recognize Melanoma on the dermoscopic images. Individual features are detected and classified in their domain such as absent, present, typical, and atypical which are helpful to support the diagnosing decision for the experts and non-experts. However, this process requires following several steps which is time-consuming. Employing a machine learning approach yields instant results suitable for a screening test. Also, given the critical nature of early Melanoma detection requiring continuous observation. To solve this issue, a method is introduced to automatically recognize Melanoma on the dermoscopic images immediately by analyzing lesion shape, color, texture, and size. A trained deep-learning model can be useful in this case. Using deep learning-based classification methods for recognizing Melanoma in dermoscopic images offers several advantages such as automated analysis, accuracy and consistency, improved detection, potential for early detection, and complementary tools. Here, a classification-based detection method is presented that utilizes advanced deep learning techniques. The Classification phase is designed to assist in detecting Melanoma rapidly. Training the proposed deep learning-based model requires a minimum of one annotated dataset. The purpose of annotating data is to convey precise information to deep learning models. Once trained, the model acts as a classifier (machine learning models that make decisions without explanation), providing instant detection results for Melanoma. This classification-based dermoscopic image analysis holds great potential for improving Melanoma screening tests. As part of this thesis, the deep learning-based screening test

represents Phase 2, aiming to enhance the reliability and acceptability of the detection method. Figure 6.1 shows different research steps, and a green color box is discussed and analyzed in this chapter.
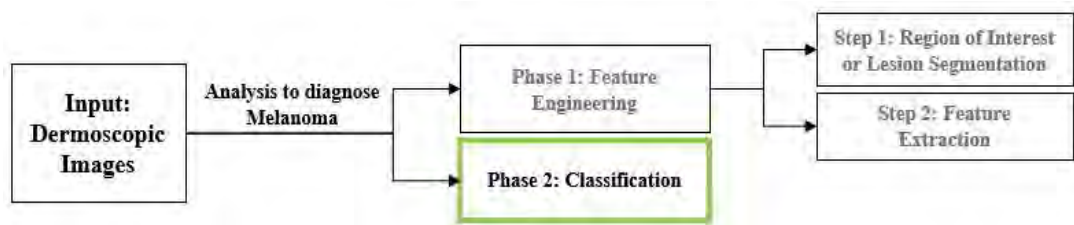
Figure 6.1: From the different steps of this thesis, Classification-based Melanoma detection (Phase 2) is analyzed here (green color box).

To achieve this, the model is trained on diverse datasets, enabling it to handle variations in Melanoma cases effectively. Moreover, using multiple datasets during deep learning model training improves generalization, adaptation, and performance in healthcare applications. The proposed model is compared to several conventional methods. This comparison serves as an evaluation of its effectiveness, contributing to the accomplishment of the second and third objectives of this research.

6.2 The Implementation of Classification Phase

Neural networks with multiple layers are used in deep learning to identify complex patterns and representations in data. Activation functions are critical in enabling neural networks to capture intricate relationships in data and can aid in normalization, decrease gradient-related issues, and enhance the overall performance of the network, especially in Melanoma detection. A CNN model is trained using six different activation functions (ReLU, ELU, Clipped ReLU, Leaky ReLU, PReLU, and Tanh) to classify skin lesion images in the PH2 dataset, augmented PH2 dataset, and ISIC archive. Before designing the CNN model, specific datasets need to be acquired for model training. Augmenting

image data operation (see Chapter 3: Research Methodology) is necessary to achieve the expected accuracy. This is because it is usually difficult to achieve an acceptable accuracy rate with a limited amount of data applied during the training process. Figure 6.2 displays the proposed approaches sequentially.
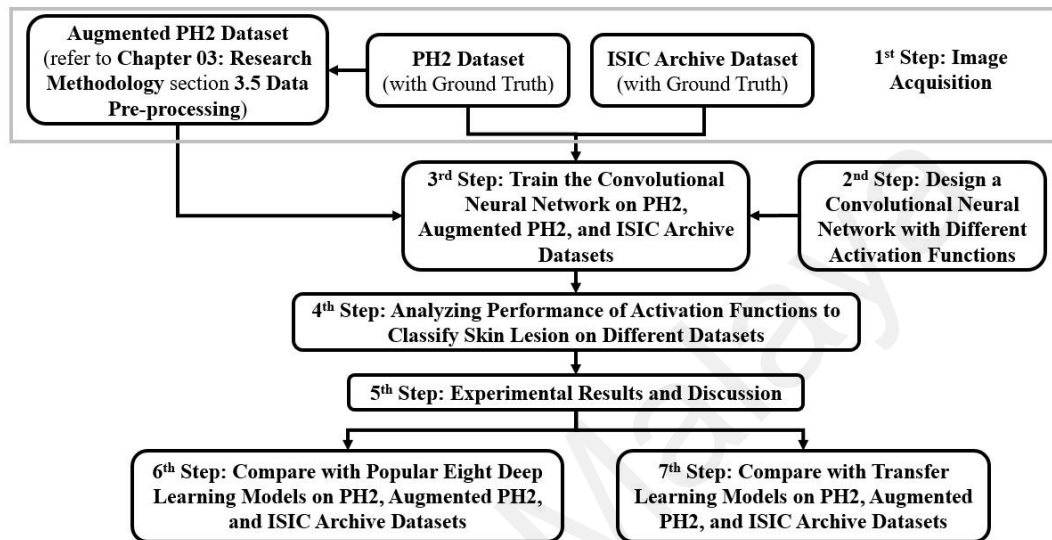
Figure 6.2: Different steps of the proposed Classification-based Melanoma detection method by CNN.

6.2.1   Image Acquisition

This chapter divides the lesion images into three classes, including atypical nevus, typical nevus, and Melanoma, utilizing the PH2 and ISIC archive datasets (see Chapter 3: Research Methodology section 3.4 The Source of Data). Since PH2 has only 200 data including 80 atypical nevus, 80 typical nevus, and 40 Melanoma, the augmentation process is done to increase the amount of data from 200 to 400. The rotation techniques including 90, 180, and 270 degrees are applied in the augmentation process (refer to Chapter 3: Research Methodology section 3.5 Data Pre-processing).

Thousands of dermoscopic images of different skin conditions, including the three classes of Melanoma, benign, and nevus, are available for research in the ISIC archive. A

random selection of 100 Melanoma images from 5598, 100 benign images from 47,684, and 100 nevus images from a selected number of dermoscopic images are used to train the proposed CNN model. The total number of datasets used is three. For the classification-based approach, they are PH2 (200 images), augmented PH2 (400 images), and ISIC archive (300 images).

6.2.2   Designing A CNN Model

Proposing a new CNN model is essential for detecting Melanoma in dermoscopic images, as it enhances feature extraction for Melanoma-specific patterns and increases accuracy. The design is adapted to solve the difficulties presented by dermoscopic images and promotes improvements in early and precise Melanoma diagnosis. Existing CNN models are complex, difficult to implement, time-consuming to train, and power-consuming processes. Also, the existing methods from literature were not built for detecting and analyzing the critical features of dermoscopic images. The traditional CNN models were trained on a large amount of data, making them less learning efficient. To overcome these issues, a deep CNN model is proposed here which can learn faster from the small training dataset. The proposed CNN design includes 2D convolution layers, batch normalization layers, activation function layers, and max-pooling layers in its basic concept. LeNet (Lecun et al., 1998) was originally designed for hand-written digit recognition and several of its fundamental ideas have been borrowed. The proposed CNN model's design includes 3 initial convolutional layers, 2 pooling layers, and 3 fully connected layers, followed by 7 more convolutional layers, 6 pooling layers, and a final 3 fully connected layers. The best possible combination is tested for each of the model's parameters, including convolutional filter, max-pooling filter, stride, dilation factor, and max epochs. Hyperparameter tuning is done through several "Grid search" testing and training processes to find the optimal combination for a given machine learning model.

This ensures that the hyperparameters remain consistent when different activation functions are tested during the experiment stage and yield the best performance on a validation set (Nugraha & Sasongko, 2022). In Table 6.1, the values of several hyperparameters are presented which are determined by doing several experiments.

Table 6.1: The values of several hyperparameters for the proposed CNN model.

Name of Hyperparameter	Value of Hyperparameter
Pooling size	3
Stride	2
Dilation Factor	2
Padding	same
Optimizer	Stochastic Gradient Descent with Momentum (SGDM)
Filter size	3x3
Learning rate	0.01
Number of epochs	250
Validation frequency	25
Batch size	32

The best combination of hyperparameters listed in Table 6.1 yields optimal performance for detecting Melanoma in dermoscopic images. Additionally, to improve efficiency, the proposed model is made deeper by increasing the number of layers (from 11 to 31). To determine the ideal number of layers, an ablation study (Grid search) is carried out, which involves testing the effects of removing the final convolutional layer and retraining the network while monitoring the performance. Figure 6.3 shows the modifications made during the design process.
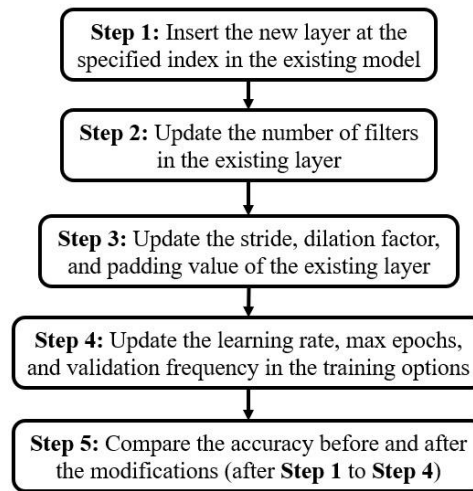
Figure 6.3: Different modification steps of the proposed CNN model.

Figure 6.4 shows the proposed CNN model for classifying dermoscopic images into three groups: Melanoma, common nevus, and atypical nevus. It has 31 layers, including 3 fully connected layers for lesion classification, 6 pooling layers for feature extraction, and 7 convolutional layers. The section on experimental result analysis contains comprehensive information.
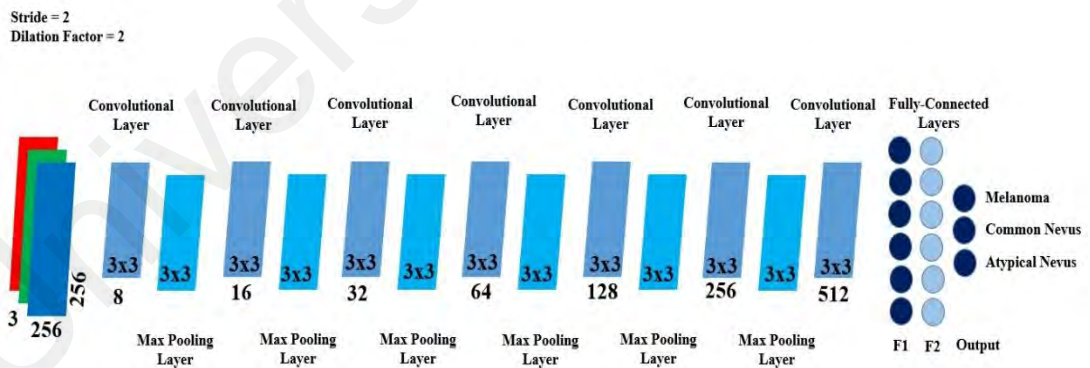
Figure 6.4: The proposed Convolutional Neural Networks for classification.

6.2.3 Training the Proposed CNN Model

To train the proposed CNN model, the datasets are randomly split into 80% for training and 20% for validation. Since the training datasets are relatively small - PH2 has 200,

Augmented PH2 has 400, and ISIC archive has 300 images—the model is trained for 250 epochs with an initial learning rate of 0.01 (selected after studying existing models from the literature) and a cross-validation frequency of 25 to produce 10 validation results. A 5-fold cross-validation procedure is applied to all three datasets.

It is important to note that an iteration refers to passing a batch through the CNN, while an epoch refers to scanning the entire dataset. In this case, the PH2 (160 training images) and ISIC archive (240 training images) datasets have 1 iteration per epoch, while the augmented PH2 (320 training images) dataset has 2 iterations per epoch.

Except for the Clipped ReLU, all activation functions in the PH2 dataset continue to maintain constant training accuracy after approximately 150 to 180 epochs. On the augmented PH2 dataset, the training accuracy remains constant after roughly 200 to 220 epochs. In the ISIC archive dataset, the training accuracy remained constant after 180 epochs. After 220 epochs (when the validation error is lower), the learning rate for most activation functions nearly remains constant, suggesting that the training process should be stopped to prevent overfitting. Using the PH2 dataset, augmented PH2 dataset, and ISIC archive dataset, respectively, Figure 6.5, Figure 6.6, and Figure 6.7 shows the training accuracy curve for various nonlinear activation functions.
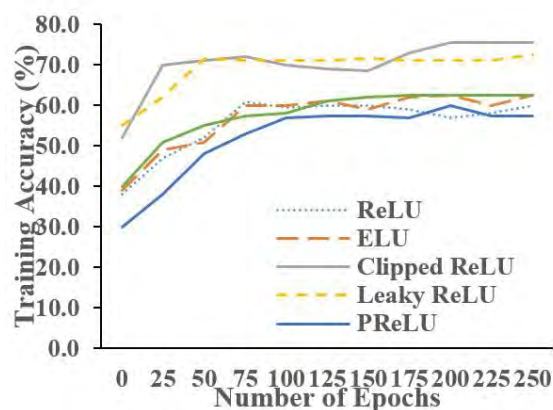
Figure 6.5: The proposed CNN training curve for different activation functions using the PH2 dataset.
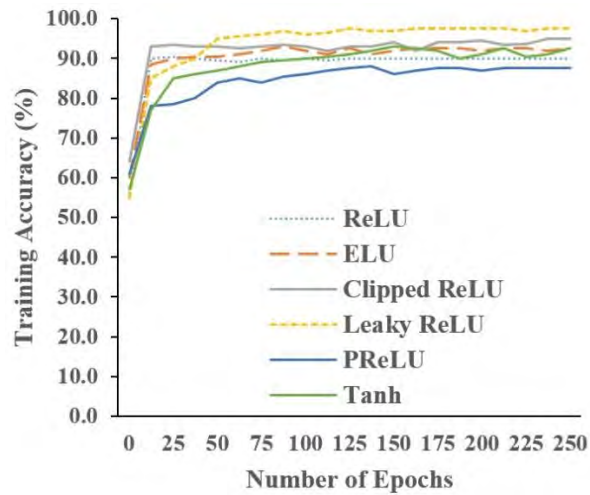
Figure 6.6: The proposed CNN training curve for various activation functions using the augmented PH2 dataset.
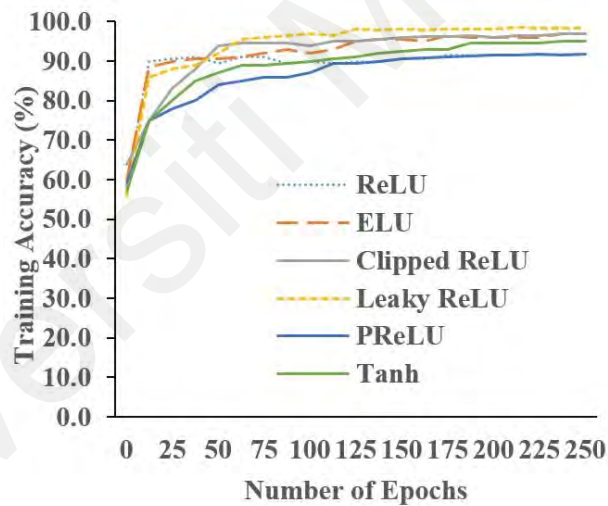
Figure 6.7: The proposed CNN training curve for different activation functions using the ISIC archive dataset.

6.2.4 Performance of Various Activation Functions on the Proposed CNN

Activation functions in CNN introduce nonlinearity, enabling the network to learn complex relationships in data. They model complex patterns, differentiate between classes, and enhance the network's capacity to capture subtle features needed for tasks such as image classification and object detection. The appropriate activation functions

affect the network's convergence speed, gradient flow, and accuracy of learning and generalization from data.

ReLU, ELU, Clipped ReLU, Leaky ReLU, PReLU, and Tanh functions are applied to the PH2, augmented PH2, and ISIC archive datasets while maintaining other hyper-parameters constant to compare the performance of activation functions in the proposed CNN model for Melanoma classification. Numerous factors, including model adaptability, non-linearity, avoiding vanishing, performance, and generalization, drive the testing of CNNs with multiple activation functions.

After the CNN model has been trained on the 200 high-resolution skin lesion images in the PH2 dataset, it is first tested on a test set of 40 images. To improve accuracy, the CNN model is retrained using the 400 skin lesion images from the augmented PH2 dataset and evaluated using a test set of 80 images. The ISIC archive dataset, which contains 300 high-resolution dermoscopic images, is used to train the CNN model, and a test set of 60 images is used to evaluate it so that the result variation can be seen once more. The inclusion of multiple datasets in CNN serves several important purposes, including increasing variability and addressing dataset bias. Table 6.2 compares the training times (TT) and validation accuracy (VA) for hidden layers with different activation functions. Leaky ReLU outperforms other activation functions, producing the fourth-shortest training time on the augmented PH2 and ISIC archive datasets, while Clipped ReLU outperforms other activation functions, producing the second-shortest training time on the PH2 dataset.

Table 6.2: Comparing the proposed model's accuracy (%) and training duration on different datasets.

Activation Functions	PH2 Dataset (160 Training Images)		Augmented PH2 Dataset (320 Training Images)		ISIC archive Dataset (240 Training Images)	
	VA (%)	TT (seconds)	VA (%)	TT (seconds)	VA (%)	TT (seconds)
ReLU	60.00	**55**	90.00	111	91.67	**59**
ELU	62.50	74	92.50	148	96.67	77
Clipped ReLU	**75.50**	56	95.00	110	96.67	58
Leaky ReLU	72.50	56	**97.50**	113	**98.33**	61
PReLU	57.50	118	87.50	233	91.67	121
Tanh	62.50	56	92.50	**109**	95.00	**59**

This experiment shows that Clipped ReLU performs better with PH2 dataset while Leaky ReLU performs better with Augmented PH2 and ISIC archive datasets. The comparison results showed that the proposed model uses less processing power and picks up features from the standard dataset more quickly when the Leaky ReLU function is added to the hidden layers. Additionally, a change in the output is seen when the filter values in the 2D convolutional and max-pooling layers are changed. The filter size typically ranges from 1 x 1 to 5 x 5, with 3 x 3 being the ideal size, for both 2D convolutional and max-pooling layers.

Figure 6.8, Figure 6.9, and Figure 6.10, on the other hand, graphically represent the confusion matrix (CM) which is a summary of prediction results on this classification problem, for each training process using three distinct datasets and a variety of activation functions. These CMs count and categorize the number of disease predictions that are correct and incorrect. When compared to other activation function types, it has been found that using Clipped ReLU and Leaky ReLU activation functions in the activation function layers enhances the performance of the proposed CNN model. However, more details information is reported in the next section.
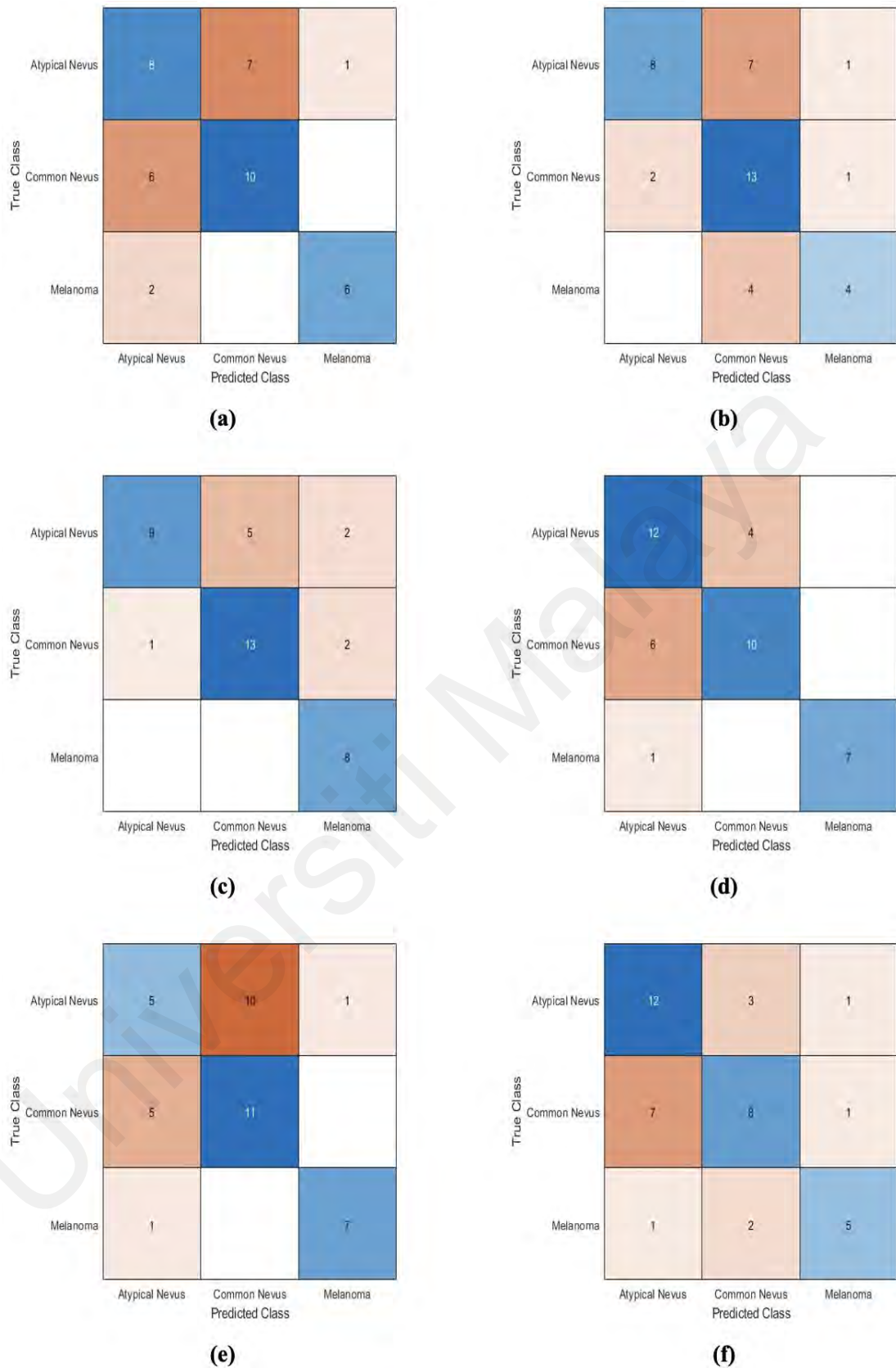
Figure 6.8: The CMs for the ReLU, ELU, Clipped ReLU, Leaky ReLU, PReLU, and Tanh functions are (a), (b), (c), (d), (e), and (f) on the PH2 dataset, respectively.
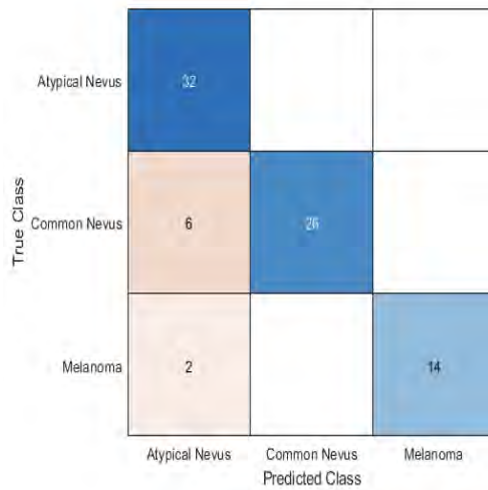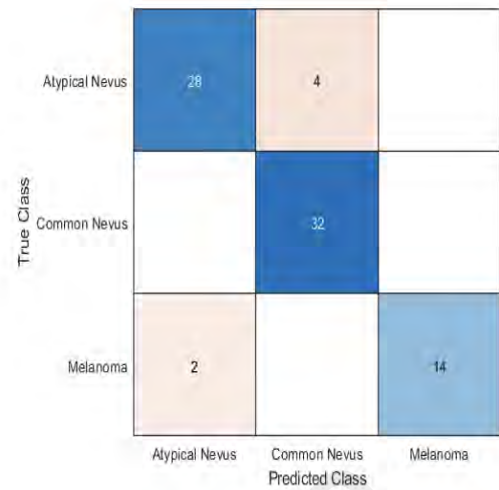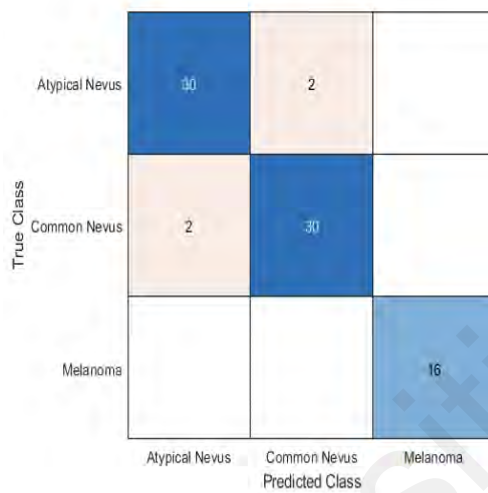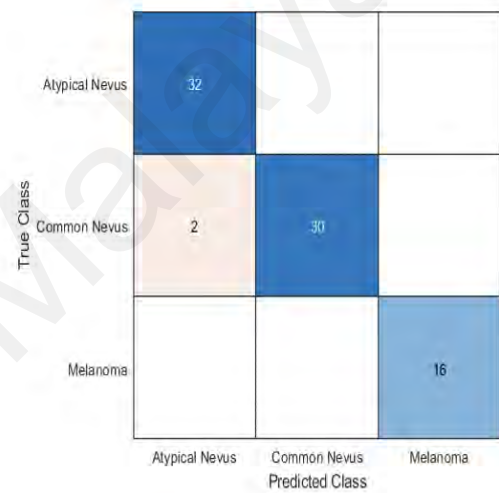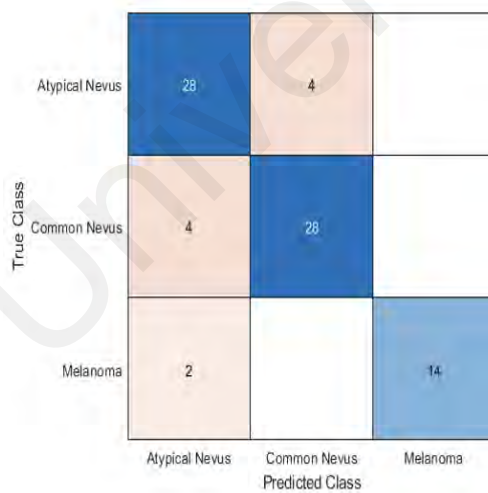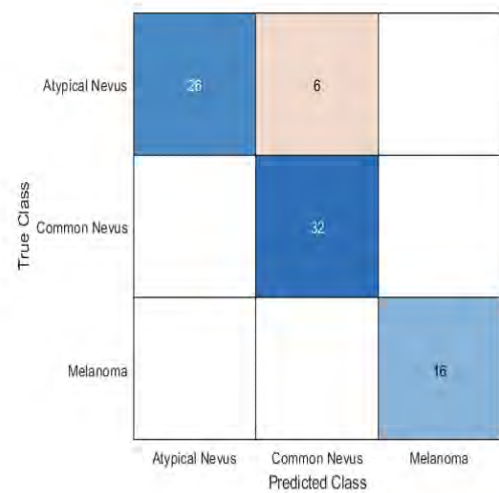
Figure 6.9: The CMs for the ReLU, ELU, Clipped ReLU, Leaky ReLU, PReLU, and Tanh functions are (a), (b), (c), (d), (e), and (f) on the augmented PH2 dataset, respectively.
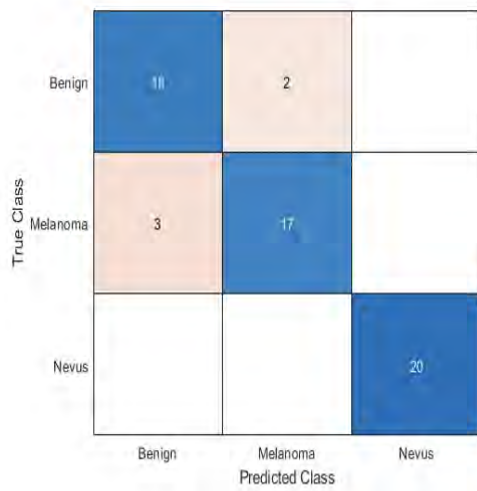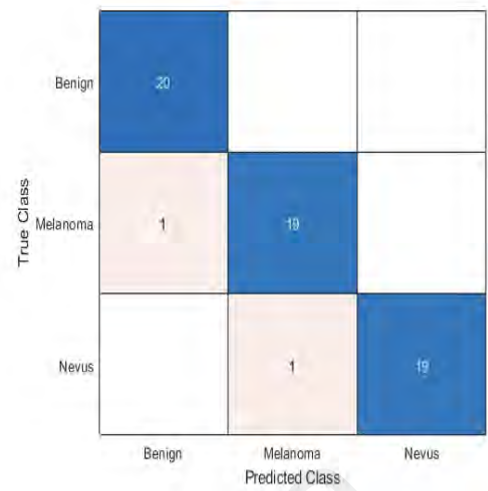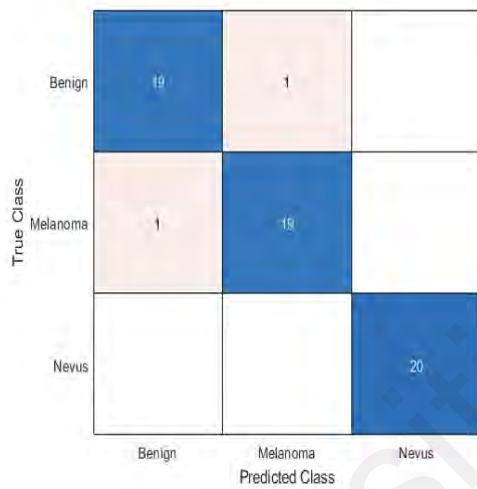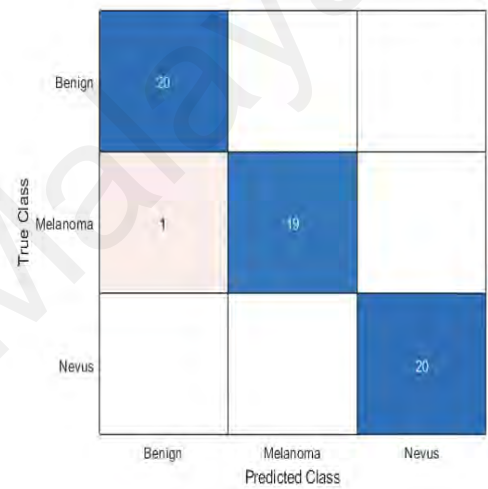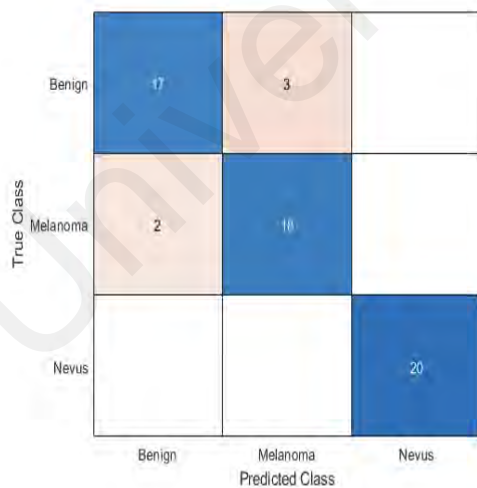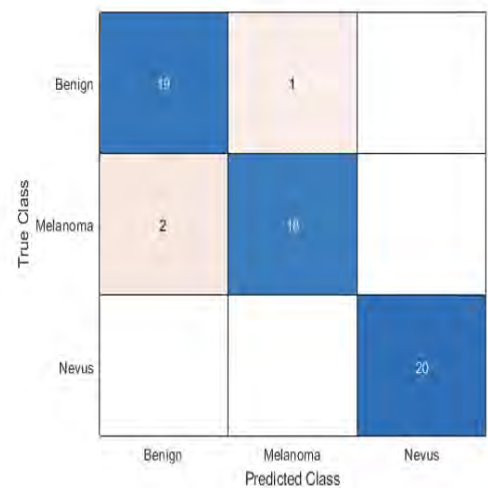
Figure 6.10: The CMs for the ReLU, ELU, Clipped ReLU, Leaky ReLU, PReLU, and Tanh functions are (a), (b), (c), (d), (e), and (f) on the ISIC archive dataset, respectively.

6.3 The Proposed CNN (Classification-based): Experimental Results and Discussion

The performance evaluation values for each class and all models, including sensitivity, specificity, precision, F1 score, accuracy, macro-precision, macro-recall, and macro-F1 score, are calculated after applying equations (7) to (14) to these CMs. The number of reported positive cases that were correctly predicted to be positive, reported negative cases that were correctly predicted to be negative and reported positive cases that were correctly predicted to be positive are listed first. The reported other matrices are based on the sensitivity, specificity, and precision of each class. Table 6.3 on the PH2 dataset, Table 6.4 on the augmented PH2 dataset, and Table 6.5 on the ISIC archive dataset separately present the evaluation values in percentages for each class and the overall model with the ReLU, ELU, Clipped ReLU, Leaky ReLU, PReLU, and Tanh functions. On the augmented PH2 and ISIC archive datasets, the proposed CNN model is trained using six nonlinear activation functions. Except for the activation functions, all the proposed CNN model's parameters remain the same during the training phase. Each of the six nonlinear activation functions mentioned above is applied to a different layer of the network, and the results are reported. At the beginning of the analysis process, it seems these activation functions are greatly responsible for the network performance. There is a wide discussion about Table 6.3, Table 6.4, and Table 6.5 in the following subsections 6.3.1, 6.3.2, and 6.3.3.

Table 6.3: Performance (%) of proposed model with various nonlinear activation functions on PH2.

Activation Function	Atypical Nevus					Common Nevus					Melanoma					wPR	wRC	wF1
	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC			
ReLU	50	67	50	50	60	63	70	59	61	67	75	97	86	80	93	61	60	60
ELU	50	92	80	62	75	**81**	54	54	65	65	50	94	67	57	85	67	62	62
Clipped ReLU	56	**96**	**90**	**69**	80	**81**	86	**72**	76	**85**	**100**	88	67	80	90	**78**	**75**	**74**
Leaky ReLU	75	71	63	**69**	73	63	83	71	67	75	88	**100**	**100**	**93**	**98**	74	73	73
PReLU	31	75	45	37	58	69	58	52	59	63	88	97	88	88	95	56	58	56
Tanh	**75**	67	60	67	70	50	79	62	55	68	63	94	71	67	88	63	63	62

Table 6.4: Performance (%) of proposed model with various nonlinear activation functions on augmented PH2.

Activation Function	Atypical Nevus					Common Nevus					Melanoma					wPR	wRC	wF1
	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC	SE	SP	PR	F1	*AC*			
ReLU	**100**	83	80	89	90	81	**100**	**100**	90	93	88	**100**	**100**	93	98	92	90	90
ELU	88	96	93	90	93	**100**	92	89	94	95	88	**100**	**100**	93	98	93	93	92
Clipped ReLU	94	96	94	94	95	94	96	94	94	95	**100**	**100**	**100**	**100**	**100**	95	95	95
Leaky ReLU	**100**	96	94	**97**	**98**	94	**100**	**100**	**97**	**98**	**100**	**100**	**100**	**100**	**100**	**98**	**98**	**98**
PReLU	88	88	82	85	88	88	92	88	88	90	88	**100**	**100**	93	98	88	88	88
Tanh	81	**100**	**100**	90	93	**100**	88	84	91	93	**100**	**100**	**100**	**100**	**100**	94	92	92

Table 6.5: Performance (%) of proposed model with various nonlinear activation functions on ISIC archive.

Activation Function	Benign					Nevus					Melanoma					wPR	wRC	wF1
	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC	SE	SP	PR	F1	*AC*			
ReLU	90	93	86	88	92	100	100	100	100	100	85	95	89	87	92	91	91	91
ELU	100	98	95	98	98	95	100	100	97	98	95	98	95	95	97	96	96	96
Clipped ReLU	95	98	95	95	97	100	100	100	100	100	95	98	95	95	97	96	96	96
Leaky ReLU	**100**	**98**	**95**	**98**	**98**	**100**	**100**	**100**	**100**	**100**	95	**100**	**100**	97	98	**97**	**97**	**97**
PReLU	85	95	89	87	92	100	100	100	100	100	90	93	86	88	92	91	91	91
Tanh	95	95	90	93	95	100	100	100	100	100	90	98	95	92	95	94	94	94

6.3.1 Experiment with PH2 Dataset

On the PH2 dataset, the proposed model performed better with Clipped ReLU in terms of training accuracy than the other activation functions. Performance is marginally reduced when the Leaky ReLU is applied in the hidden layers. The individual class performance analysis shows that only for Melanoma does Leaky ReLU perform slightly better. The PReLU function performs less accurately when used with CNN's hidden layers, as shown in Table 6.2. For this skin lesion recognition task, the coefficients (ai) of PReLU, a class of learnable parameters, are probably not the best choice. After the Clipped ReLU and Leaky ReLU, the remaining functions fall into the categories of ELU, Tanh, ReLU, and PReLU based on their varying performance levels.

6.3.2 Experiment with Augmented PH2 Dataset

Table 6.3 demonstrates that using the PH2 dataset, Leaky ReLU outperforms Clipped ReLU on the augmented dataset. In the comparison of classes, Leaky ReLU also yields the same result. Nearly all evaluation metrics favor this function (wPR, wRC, and wF1 score are all 98). All metrics for the Melanoma class are at their highest when the Hidden Layers use both the Leaky ReLU and Clipped ReLU functions. The Melanoma class displays results that are very similar to those when the other four activation functions are utilized. It is unknown why the Melanoma class performs more accurately than the other two classes across all six functions.

Visual observation shows that the atypical nevus and common nevus stages of skin disease are less severe than Melanoma. The characteristics of Melanoma lesions are likely more severe than those of the other two classes. This may help to explain why, when using all six functions, the Melanoma class outperformed the other classes. It should be noted that Leaky ReLU produced a relatively higher performance with the common nevus and atypical nevus classes. Both the common and atypical nevus classes perform exceptionally well for the Tanh and ReLU functions.

This mathematical mechanism is better suited than other mechanisms for this image recognition process; it multiplies any input element that is less than zero by a specific scale factor (0.01). Depending on their performance levels from higher to lower, these activation functions can be rearranged as Leaky ReLU, Clipped ReLU, Tanh, ELU, ReLU, and PReLU, respectively.

6.3.3 Experiment with ISIC Archive Dataset

Based on Table 6.4, the Leaky ReLU outperforms the other activation functions on the ISIC archive dataset. In all three classes, Leaky ReLU has outperformed the opposition

(wPR is 97, wRC is 97, and wF1 score is 97) (Benign F1 score is 98, Nevus F1 score is 100, and Melanoma F1 score is 97). The Clipped ReLU and ELU both performed well, tying for second place. The third position belongs to the Tanh function, while the fourth position belongs to the combination of ReLU and PReLU. From higher to lower performance, they are Leaky ReLU, Clipped ReLU, ELU, Tanh, ReLU, and PReLU. In this instance, only the ELU activation function's performance (fourth in the augmented PH2 dataset, third in the PH2 and ISIC archive datasets) is a little bit unstable. The augmented PH2 dataset turns out to be the one where the Tanh function performs best.

These three experiments indicate that the Clipped ReLU function (wF1 score of 74) outperforms other functions when training the proposed CNN model on the PH2 dataset. The Leaky ReLU function also performs remarkably well on the same dataset (wF1 score is 73). On the augmented PH2 and ISIC archive datasets, the Leaky ReLU function outperforms the other functions in training the proposed CNN model (the wF1 score for augmented PH2 is 98, and the wF1 score for ISIC archive is 97). Given the original nature of the PH2 dataset, it appears that the Clipped ReLU exhibits robust performance, whereas the Leaky ReLU demonstrates superiority in scenarios involving augmented PH2 and ISIC archive datasets.

6.4   Comparing the Proposed CNN Model with Other Deep Learning Models

For a thorough evaluation of the proposed CNN model's performance and to confirm its effectiveness for a variety of Melanoma recognition tasks, it is crucial to compare it to other deep learning models. Informed decisions about model selection and optimization strategies can be made by using this information to identify strengths, weaknesses, and potential improvements. The sixth step of Figure 6.2 is represented in this section. Based on their performance on three datasets (PH2, augmented PH2, and ISIC archive), the

proposed CNN model is contrasted with the eight widely used deep learning models. These models' performances are evaluated by different evaluation metrics equations (7) to (16) (see Chapter 3: Research Methodology). Several criteria, such as architecture, performance, training efficiency, generalization, model size, interpretability, availability and community support, and state-of-the-art performance, have been established for comparison. These aspects are considered when selecting, implementing, and training the most popular deep learning models on the augmented PH2 and ISIC archive datasets in comparison to the proposed CNN model. The proposed model's highest performance was noted on the PH2 dataset with the Clipped ReLU function, so the performance of the other models is contrasted with the performance of the Clipped ReLU in this comparison section. The proposed model performs best on the augmented PH2 and ISIC archive datasets when Leaky ReLU is applied in hidden layers. The effectiveness of the other models is compared to Leaky ReLU's performance on the augmented PH2 and ISIC archive datasets.

To compare the proposed model with other deep learning models, the eight most popular deep learning models are chosen such as AlexNet with 25 layers (Krizhevsky et al., 2017), GoogleNet with 144 layers (Szegedy et al., 2015), ResNet18 with 71 layers (He et al., 2016), ResNet50 with 177 layers (He et al., 2016), ResNet 101 with 347 layers (He et al., 2016), SqueezeNet with 68 layers (Bernardo et al., 2022), VGG16 with 41 layers (Russakovsky et al., 2015; Mascarenhas & Agarwal, 2021), and VGG 19 with 47 layers (Russakovsky et al., 2015; Mascarenhas & Agarwal, 2021). Table 6.6 displays the validation accuracy (VA) and training time (TT) for these eight models as well as the proposed CNN model. AlexNet trained in the least amount of time (20 seconds) across all three datasets. SqueezeNet, ResNet18, and GoogleNet, in that order, have the second, third, and fourth shortest training times, respectively. However, the TT for the proposed CNN model was the fifth quickest. However, the proposed model has a top performance

based on the VA for all three datasets (75.5% for PH2, 97.5% for augmented PH2, and 98.33% for ISIC archive).

Table 6.6: Evaluating the accuracy (%) and training time of the proposed CNN model against those of other deep learning models on different datasets.

Models	PH2 Dataset		Augmented PH2 Dataset		ISIC archive Dataset	
	VA (%)	TT (seconds)	VA (%)	TT (seconds)	VA (%)	TT (seconds)
AlexNet	65.50	**20**	92.50	**36**	93.33	**30**
GoogleNet	70.00	42	85.00	80	92.00	60
ResNet18	73.00	32	90.00	58	84.00	46
ResNet50	60.00	98	94.00	194	90.00	140
ResNet101	68.00	198	93.00	360	92.00	300
SqueezeNet	68.00	22	80.00	42	91.67	38
VGG16	65.00	124	71.00	212	76.00	188
VGG19	70.00	162	72.00	302	89.00	209
Proposed model (Clipped ReLU)	**75.50**	56	95.00	110	96.67	58
Proposed model (Leaky ReLU)	72.50	56	**97.50**	113	**98.33**	61

Following that, evaluation metrics for the different datasets are used to evaluate the models' performance. Table 6.7, Table 6.8, and Table 6.9 display the percentage results for the performance evaluation metrics for these eight models on the PH2, augmented PH2 and ISIC archive datasets. To maintain consistency and minimize any skewed comparisons, all eight models were created in the same setting as the proposed CNN model. The convolutional layers of these networks extract the features that the final learnable layer and classification layer use to classify the input images. These two layers had to be swapped out for new ones to train these eight pre-trained models on the three distinct datasets. In these eight networks, the final layer with learnable weights is typically fully connected. In place of the fully connected layer, a new fully connected layer is substituted with the same number of outputs as the classes in the new dataset (3 in this case, including Atypical Nevus, Common Nevus, and Melanoma). However,

SqueezeNet's final learnable layer is a 1-by-1 convolutional layer. It is necessary to swap out the convolutional layer for a new one that has the same number of filters as classes. The eight models are then individually trained using the PH2, augmented PH2, and ISIC archive datasets. On the PH2 dataset, the proposed model (Clipped ReLU) is compared to AlexNet and GoogleNet.

Table 6.7: Performance (%) comparison of different deep learning models on PH2.

Models	Atypical Nevus					Common Nevus					Melanoma					wPR	wRC	wF1
	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC			
AlexNet	50	83	67	57	70	69	75	65	67	73	88	88	64	74	88	66	65	64
GoogleNet	75	71	63	69	73	56	88	75	64	75	88	94	78	82	93	71	70	70
ResNet18	69	75	65	67	73	81	83	76	**79**	83	63	97	83	71	90	73	73	73
ResNet50	**88**	46	52	65	63	31	96	83	45	70	75	97	86	80	93	71	63	60
ResNet101	31	**96**	83	45	70	**94**	54	58	71	70	88	97	88	88	95	74	68	64
SqueezeNet	38	**96**	86	52	73	88	71	67	76	78	88	84	58	70	85	73	68	65
VGG16	75	63	57	65	68	44	92	78	56	73	88	91	70	78	90	68	65	64
VGG19	**88**	63	61	**72**	73	50	**96**	**89**	64	78	75	94	75	75	90	75	70	69
Proposed model (Clipped ReLU)	56	**96**	**90**	69	**80**	81	86	72	76	**85**	**100**	88	67	80	90	**78**	**75**	**74**
Proposed model (Leaky ReLU)	75	71	63	69	73	63	83	71	67	75	88	**100**	**100**	**93**	**98**	74	73	73

The proposed model outperforms these two models, as shown in Table 6.7 (wPR is 78, wRC is 75, and wF1 score is 74). ResNet18, 50, and 101 are once more used to compare the proposed model to. The proposed model performs better than ResNet 101 in the Melanoma class than it does in the other two classes. A comparison of the proposed model with ResNet 101 shows that the proposed model with the Leaky ReLU function outperforms all other models in the Melanoma class. Next, the proposed model (Clipped ReLU) is compared to SqueezeNet, VGG16, and VGG19. Although SqueezeNet, VGG16, and VGG19 show marginally better performances in the classes of Atypical Nevus and Common Nevus, the proposed model outperforms these three models, as expected. The Clipped ReLU function is suggested for the remaining two classes when the training dataset is PH2, and the Leaky ReLU function for the Melanoma class.

Table 6.8: Performance (%) comparison of different deep learning models on augmented PH2.

Models	Atypical Nevus					Common Nevus					Melanoma					wPR	wRC	wF1
	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC			
AlexNet	88	94	90	88	91	91	92	88	89	91	**100**	**100**	**100**	**100**	**100**	91	92	91
GoogleNet	91	81	76	83	85	78	94	89	83	88	88	**100**	**100**	93	98	86	85	85
ResNet18	88	92	88	88	90	97	92	89	93	94	81	**100**	**100**	90	96	91	90	90
ResNet50	84	**100**	**100**	92	94	**100**	90	86	93	94	**100**	**100**	**100**	**100**	**100**	94	94	94
ResNet101	94	92	88	91	93	88	96	93	90	93	**100**	**100**	**100**	**100**	**100**	92	93	92
SqueezeNet	94	71	68	79	80	63	96	91	74	83	88	**100**	**100**	93	98	84	80	80
VGG16	28	**100**	**100**	44	71	**100**	52	58	74	71	**100**	**100**	**100**	**100**	**100**	83	71	67
VGG19	47	90	75	58	73	97	65	65	78	78	75	**100**	**100**	86	95	76	73	72
Proposed model (Leaky ReLU)	**100**	96	94	**97**	**98**	94	**100**	**100**	**97**	**98**	**100**	**100**	**100**	**100**	**100**	**98**	**98**	**98**

Table 6.8 compares the performance of the proposed model (Leaky ReLU) with AlexNet on the augmented PH2 dataset (wPR, wRC, and wF1 score are all 98). AlexNet performs well for the Melanoma class when used as the proposed CNN model in this situation, but not for other classes. Next, the results of GoogleNet and RestNet18 are evaluated. The proposed model will outperform these two going forward. The output is better overall, but the proposed model also performs better than ResNet50 and ResNet101, especially in the Melanoma class. The SqueezeNet's performance cannot outperform the proposed model's performance. The final comparison shows that the performance of the proposed model outperforms that of VGG16 and VGG19. VGG16 performs equally well as ResNet50 and ResNet101 at Melanoma recognition. It is suggested that the Leaky ReLU function be used in the hidden layers of the proposed model to train augmented PH2 dataset.

Table 6.9: Performance (%) comparison of different deep learning models on ISIC archive.

Models	Benign					Nevus					Melanoma					wPR	wRC	wF1
	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC	SE	SP	PR	F1	AC			
AlexNet	95	88	79	86	90	95	95	90	93	95	93	87	70	80	88	79	93	85
GoogleNet	85	100	100	92	95	95	95	90	93	95	100	95	91	95	97	93	92	92
ResNet18	85	88	77	81	87	95	95	90	93	95	75	95	88	81	88	84	84	84
ResNet50	85	98	94	89	93	100	98	95	98	98	90	93	86	88	92	91	91	91
ResNet101	90	95	90	90	93	95	100	100	97	98	95	95	90	93	95	92	92	92
SqueezeNet	85	100	100	92	95	90	97	95	93	95	100	95	91	95	97	94	91	92
VGG16	35	100	100	52	78	95	95	90	93	95	100	70	63	77	80	83	76	73
VGG19	85	95	89	87	92	95	100	100	97	98	90	90	82	86	90	89	89	89
Proposed model (Leaky ReLU)	**100**	**98**	**95**	**98**	**98**	**100**	**100**	**100**	**100**	**100**	**95**	**100**	**100**	**97**	**98**	**97**	**97**	**97**

Table 6.9 contrasts AlexNet's performance with that of the proposed CNN model (Leaky ReLU) on the initial ISIC archive dataset (wPR, wRC, and wF1 scores are all 97). AlexNet did not perform better than the proposed CNN model, according to the calculation. GoogleNet and the proposed model do well in predicting individual classes, but not in all datasets. ResNet18, ResNet50, and ResNet101 did not show any promising results on this dataset, even though ResNet101's performance was the best of the ResNet series. The SqueezeNet's performance is almost identical to the GoogleNet's. Both models produce the same classification for benign and Melanoma. The Nevus classification produces a slightly different result in the performance of these two models. The performance of the proposed model outperforms that of VGG16 and VGG19 when compared to the formers. Finally, the performance of the proposed model on the ISIC archive dataset using the Leaky ReLU activation function is another illustration of its superior prediction capabilities.

Two of these eight deep learning models, ResNet18 and VGG19, are chosen as the top two models based on the data in Table 6.7 and will be compared with the proposed CNN model (Clipped ReLU). Figure 6.11 shows the Receiver Operating Characteristic (ROC) curve plotted for this comparison. As defined by Sahiner et al. (2008), the true positive

rate (sensitivity) versus false-positive rate (1-specificity) ratio is the area under the curve (AUC) of the ROC curve. The AUCs for the proposed ResNet18 and VGG19 models are 0.72, 0.64, and 0.59, respectively, as shown in Figure 6.11. Due to its wider coverage, the proposed CNN model performs better.
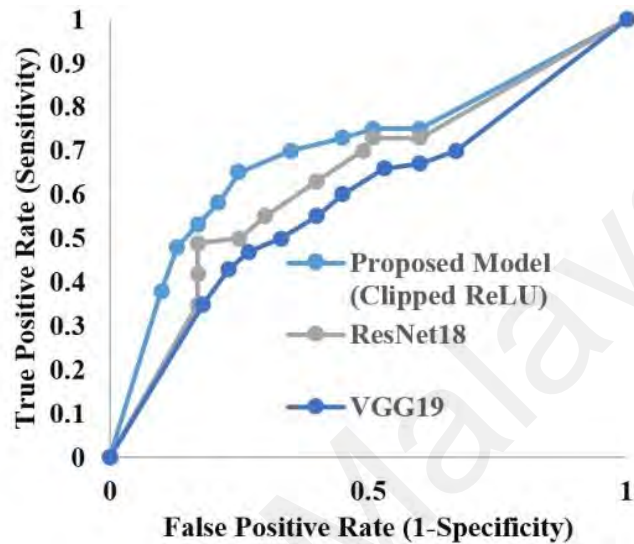
Figure 6.11: Various models' ROC curves on the PH2 dataset.

In Figure 6.12, the same methods are used. From these eight models, the top two models, ResNet50 and ResNet101, are chosen based on the data in Table 6.8. The second ROC curve is used to contrast the proposed model (Leaky ReLU) with the top two models. The proposed ResNet50 and ResNet101 models have AUCs of 0.94, 0.88, and 0.84, respectively, as shown in Figure 6.12. In this situation as well, the proposed model performs better due to its wider coverage.
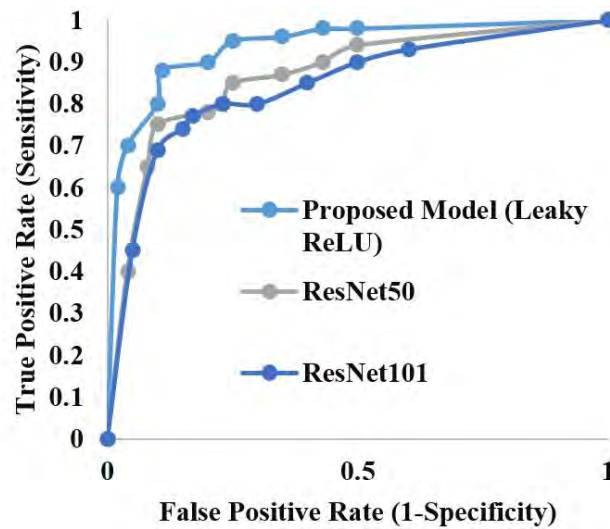
Figure 6.12: Various models' ROC curves on the augmented PH2 dataset.

The top two models once more two of these eight models, GoogleNet and SqueezeNet, are chosen based on the data in Table 6.9. Figure 6.13 depicts the third ROC curve, which contrasts the proposed model (Leaky ReLU) with the top two models. The AUC values for the proposed CNN model, GoogleNet and SqueezeNet are 0.95, 0.85, and 0.84, respectively, as shown in Figure 6.13. Due to its wider coverage, the proposed CNN model performs better in this instance as well.
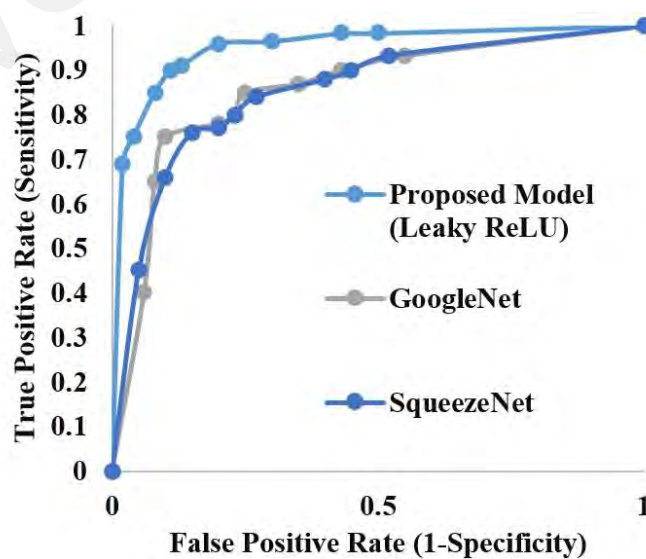
Figure 6.13: Various models' ROC curves on the ISIC archive dataset.

6.5 Comparing the Proposed Classification Model with Transfer Learning

This section discusses Figure 6.2's seventh step in detail. It is crucial to contrast the proposed CNN model for Melanoma detection with well-known transfer learning strategies to assess its potential clinical usefulness and detection accuracy. Task-specific optimization is essential in this specialized field of dermoscopic image analysis. A pre-trained deep learning model often learned on a sizable dataset for a separate job, is used as a starting point for a new, related task in a process known as transfer learning.

Here the experimental results are compared with several recent transfer learning-based deep CNN models; where the pertained Swin transformenr (Ayas et al., 2022) CNN (Anand et al., 2023), AlexNet (Hosny et al., 2018), ResNet50 (Kondaveeti & Edupuganti, 2020), and MobileNet (Younis et al., 2019) are fine-tuned and trained over the PH2, augmented PH2, and ISIC archive datasets in the same training condition as the proposed model by using different evaluation metrics equation (7) to (16) (see Chapter 3: Research Methodology). The validation accuracy (VA) and training time (TT) for the proposed CNN model and transfer learning models are shown in Table 6.10.

Table 6.10: Comparing the accuracy (%) and training duration of the proposed CNN model to transfer learning algorithms on different datasets.

Methods	PH2 Dataset		Augmented PH2 Dataset		ISIC archive Dataset	
	VA (%)	TT (seconds)	VA (%)	TT (seconds)	VA (%)	TT (seconds)
Ayas et al. (2022)	67	103	87	199	88	144
Anand et al. (2023)	69	101	90	197	89	140
KM Hosny et al. (2018)	71	60	91	120	90	70
Kondaveeti & Edupuganti (2020)	71	102	92	198	92	143
Younis et al. (2019)	70	62	91	123	92	72
Proposed model (Clipped ReLU)	**75.50**	**56**	95.00	**110**	96.67	**58**
Proposed model (Leaky ReLU)	72.50	**56**	**97.50**	113	**98.33**	61

Among those transfer learning models, the proposed model's training time was the quickest (56 seconds for PH2, 110 seconds for augmented PH2, and 58 seconds for ISIC

archive), and its validation accuracy was the greatest (75.5% for PH2, 97.5% for augmented PH2, and 98.33% for ISIC archive). Table 6.11 compares the performance of the proposed model to three transfer learning-based models in terms of percentages.

Table 6.11: Comparing the proposed CNN model's performance (%) against transfer learning algorithms on various datasets.

Models	PH2 Dataset				Augmented PH2 Dataset				ISIC Archive Dataset			
	AC	wRC	wPR	wF1	AC	wRC	wPR	wF1	AC	wRC	wPR	wF1
Ayas et al., 2022	67	67	68	67	87	86	87	86	88	87	87	87
Anand et al., 2023	69	69	70	69	90	90	90	90	89	89	89	90
KM Hosny et al., 2018	72	72	73	72	91	91	91	91	91	90	91	91
Kondaveeti & Edupuganti, 2020	71	70	71	71	92	91	92	92	92	92	92	92
Younis et al., 2019	70	70	71	70	92	91	91	91	92	93	92	92
Proposed Model (Clipped ReLU)	**76**	**75**	**78**	**74**	95	95	95	95	97	96	96	96
Proposed Model (Leaky ReLU)	73	73	74	73	**98**	**98**	**98**	**98**	**98**	**97**	**97**	**97**

The information in Table 6.10 and Table 6.11 shows that the proposed CNN model performs best on all three datasets, with a wF1 score of 74 for PH2, 98 for augmented PH2, and 97 for the ISIC archive. The five transfer learning models exhibit remarkably similar overall performance results. Among the five transfer learning models, the best three are choosen to show the ROC curves for the PH2, augmented PH2, and ISIC archive datasets in Figure 6.14, Figure 6.15, and Figure 6.16, respectively.
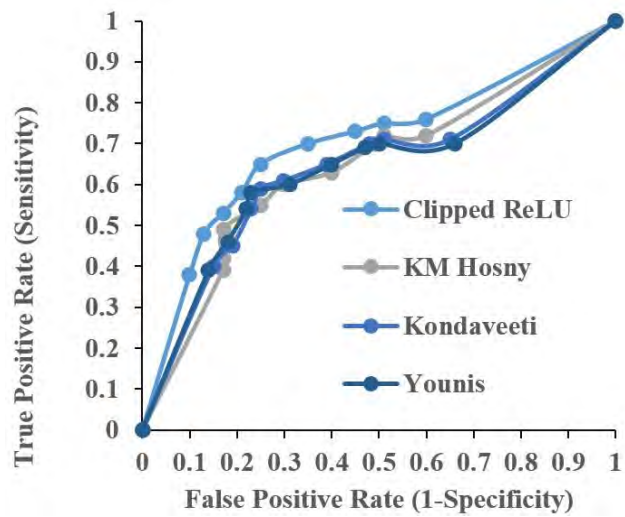
Figure 6.14: The PH2 dataset's ROC Curve for various transfer learning models.
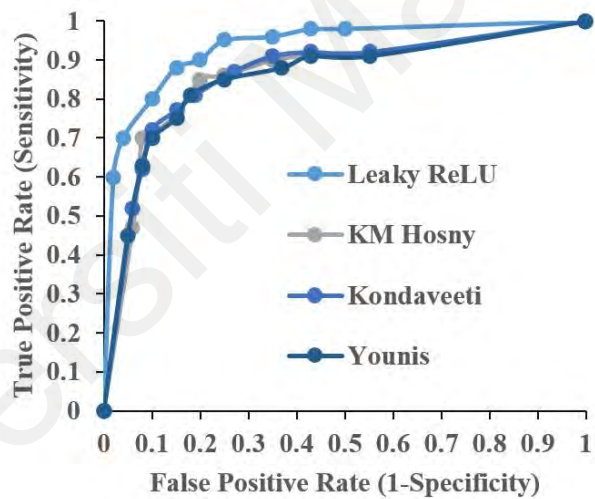
Figure 6.15: The augmented PH2 dataset's ROC Curve for various transfer learning models.
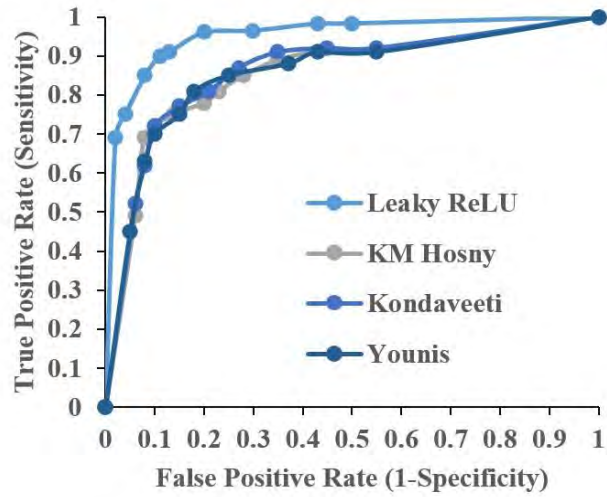
Figure 6.16: The ISIC archive dataset's ROC Curve for various transfer learning models.

These figures provide a visual comparison of the three best transfer learning models' analytical findings with the proposed model by calculating their AUC (the higher AUC value represents the better performance of the model). The AUC values are presented in Table 6.12. The proposed model has the maximum area under the curve in all three datasets (0.72 for PH2, 0.94 for augmented PH2, and 0.95 for the ISIC archive).

Table 6.12: The AUC value of different models.

Models	PH2 Dataset	Augmented PH2 Dataset	ISIC Archive Dataset
KM Hosny	0.68	0.87	0.87
Kondaveeti	0.67	0.86	0.87
Younis	0.67	0.85	0.86
Proposed model	**0.72**	**0.94**	**0.95**

6.6 Applying Proposed Model on the Unfamiliar Test Set of ISIC2018

Finally, the proposed CNN model with Leaky ReLU undergoes final training using all datasets, including PH2, augmented PH2, and ISIC archive. Once training is completed across these datasets, the model is then tested using a completely unfamiliar test set, derived from the ISIC2018 dataset. This test set comprises randomly selected skin lesion images, totaling 100, consisting of 20 Melanoma, 8 Atypical Nevus, and 72 Typical

Nevus. Performance of the proposed CNN on this test set is reported in Tables 6.13 and 6.14, where Phase 2 is compared against Phase 1 and clinical ground truth. Then proposed CNN with Leaky ReLU and Clipped ReLU is evaluated in Table 6.15. The output of this chapter serves as an instant diagnosis method (screening test) for Melanoma detection. Contrastingly, Phase 1 employs the 3-point checklist method to assess the skin condition of the corresponding lesion image, supplemented by the analysis of two additional dermoscopic features to provide further insight. The diagnosis ground truth for these 100 dermoscopic images is provided by the dataset creators, facilitating a direct comparison between the proposed CNN of Phase 2 and the feature engineering of Phase 1 (output of Chapter 5), as presented in Tables 6.13 and 6.14. Both Phase 1 and Phase 2 achieved 100% recall for Melanoma class on this test set.

Table 6.13: Applying the proposed CNN on the test set (1 to 50 of 100 dermoscopic images) to compare Phase 2 against the clinical Ground Truth and Phase 1.

ISIC Archive	Diagnose			ISIC Archive	Diagnose		
	Phase: 1	Phase: 2	Clinical GT		Phase: 1	Phase: 2	Clinical GT
ISIC_0012255	Benign	Ty	Ty	**ISIC_0017474**	**Mel***	**Mel**	**Mel**
ISIC_0012346	Benign	Ty	Ty	ISIC_0017702	Benign	Ty	Ty
ISIC_0012576	Benign	Ty	Ty	ISIC_0017755	Benign	Ty	Ty
ISIC_0012585	Benign	*Mel*	Ty	ISIC_0018111	Benign	Ty	Ty
ISIC_0012623	Benign	Ty	Ty	ISIC_0018179	Benign	Ty	Ty
ISIC_0012627	Benign	Ty	Ty	ISIC_0018248	Benign	Ty	Ty
ISIC_0012633	**Mel***	**Mel**	**Mel**	ISIC_0018375	Benign	Ty	Ty
ISIC_0012643	Benign	Ty	Ty	ISIC_0018472	Benign	Ty	Ty
ISIC_0015294	Benign	Ty	Ty	ISIC_0018521	Benign	Ty	Ty
ISIC_0015351	Benign	Ty	Ty	ISIC_0018556	Benign	Ty	Ty
ISIC_0015370	Benign	Ty	Ty	ISIC_0018611	Benign	Ty	Ty
ISIC_0015462	**Mel****	**Mel**	**Mel**	ISIC_0018680	Benign	Ty	Ty
ISIC_0015480	Benign	Ty	Ty	ISIC_0019049	Benign	Ty	Ty
ISIC_0015492	Benign	Ty	Ty	ISIC_0019309	Benign	Ty	Ty
ISIC_0015518	**Mel****	**Mel**	**Mel**	**ISIC_0019334**	**Mel***	**Mel**	**Mel**
ISIC_0015552	Benign	At	At	ISIC_0019723	Benign	Ty	Ty
ISIC_0015590	Benign	Ty	Ty	ISIC_0019794	Benign	Ty	Ty
ISIC_0015634	Benign	Ty	Ty	ISIC_0019883	Benign	Ty	Ty
ISIC_0016351	Benign	Ty	Ty	ISIC_0020233	Benign	Ty	Ty
ISIC_0016714	Benign	Ty	Ty	ISIC_0020418	Benign	Ty	Ty
ISIC_0016804	Benign	Ty	Ty	ISIC_0020861	Benign	Ty	Ty
ISIC_0017341	Benign	Ty	Ty	ISIC_0020893	Benign	Ty	Ty
ISIC_0017398	**Mel***	**Mel**	**Mel**	ISIC_0020953	Benign	Ty	Ty
ISIC_0017399	**Mel***	**Mel**	**Mel**	ISIC_0020999	Benign	*Mel*	At
ISIC_0017460	**Mel****	**Mel**	**Mel**	ISIC_0021037	Benign	At	At

Note: Ty = Typical, At = Atypical, and Mel = Melanoma.

Table 6.14: Applying the proposed CNN on the test set (51 to 100 of 100 dermoscopic images) to compare Phase 2 against the clinical Ground Truth and Phase 1.

ISIC Archive	Diagnose			ISIC Archive	Diagnose		
	Phase: 1	Phase: 2	Clinical GT		Phase: 1	Phase: 2	Clinical GT
ISIC_0021041	Benign	At	At	**ISIC_0023755**	**Mel****	**Mel**	**Mel**
ISIC_0021152	Benign	Ty	Ty	ISIC_0023831	Benign	Ty	Ty
ISIC_0021158	Benign	Ty	Ty	**ISIC_0023900**	**Mel***	**Mel**	**Mel**
ISIC_0021202	Benign	Ty	Ty	ISIC_0023904	Benign	Ty	Ty
ISIC_0021251	Benign	*At*	Ty	**ISIC_0023924**	**Mel****	**Mel**	**Mel**
ISIC_0021448	*Mel**	Ty	Ty	ISIC_0023936	Benign	Ty	Ty
ISIC_0021449	*Mel**	Ty	Ty	ISIC_0024135	Benign	Ty	Ty
ISIC_0021504	Benign	Ty	Ty	**ISIC_0036073**	**Mel****	**Mel**	**Mel**
ISIC_0021714	Benign	At	At	**ISIC_0036085**	**Mel***	**Mel**	**Mel**
ISIC_0021762	Benign	Ty	Ty	**ISIC_0036098**	**Mel***	**Mel**	**Mel**
ISIC_0021816	**Mel****	**Mel**	**Mel**	**ISIC_0036101**	**Mel***	**Mel**	**Mel**
ISIC_0021904	Benign	Ty	Ty	**ISIC_0036121**	**Mel***	**Mel**	**Mel**
ISIC_0021914	Benign	At	At	ISIC_0036147	Benign	Ty	Ty
ISIC_0021990	Benign	Ty	Ty	ISIC_0036174	Benign	Ty	Ty
ISIC_0022029	Benign	Ty	Ty	ISIC_0036206	Benign	Ty	Ty
ISIC_0022039	Benign	Ty	Ty	ISIC_0036236	Benign	Ty	Ty
ISIC_0022147	Benign	Ty	Ty	ISIC_0036237	Benign	Ty	Ty
ISIC_0022192	Benign	At	At	ISIC_0036240	Benign	Ty	Ty
ISIC_0022221	Benign	Ty	Ty	ISIC_0036247	Benign	At	At
ISIC_0022657	Benign	Ty	Ty	**ISIC_0036281**	**Mel***	**Mel**	**Mel**
ISIC_0022738	Benign	Ty	Ty	ISIC_0036291	Benign	Ty	Ty
ISIC_0023371	**Mel***	**Mel**	**Mel**	ISIC_0036306	*Mel**	Ty	Ty
ISIC_0023508	**Mel***	**Mel**	**Mel**	ISIC_0036321	*Mel**	Ty	Ty
ISIC_0023628	Benign	Ty	Ty	ISIC_0036328	Benign	Ty	Ty
ISIC_0023678	Benign	Ty	Ty	ISIC_0036333	Benign	Ty	Ty

Note: Ty = Typical, At = Atypical, and Mel = Melanoma.

Table 6.15: Evaluation of the proposed CNN (Phase 2) on the test set of ISIC2018.

Method	Melanoma			Atypical Nevus			Typical Nevus			Overall		
	PR	RC	F1	PR	RC	F1	PR	RC	F1	AC	mF1	wF1
Proposed CNN (Leaky ReLU)	91	100	95	88	88	88	100	97	99	97	94	97
Proposed CNN (Clipped ReLU)	81	85	83	60	75	67	99	94	96	91	82	91

6.7 Chapter Summary

This chapter presents Phase 2, the second and third research objectives of this thesis (see Figure 6.1). The purpose of this chapter is to suggest a technique for automatically identifying Melanoma on dermoscopic images without using any supervised techniques. The proposed CNN is trained on multiracial (PH2, augmented PH2, and ISIC archive) datasets to save time and provide a remote Melanoma recognition process (without using clinical rules-based methods). The nonlinear activation functions used to parameterize the proposed CNN are various. The experimental results are contrasted with other common deep learning models and transfer learning models to demonstrate the viability of the proposed CNN-based classification method. This study employs the proposed model to predict skin lesions, particularly Melanoma, within a few minutes of the CNN training period, enabling the possibility of a highly accurate diagnosis. This model has several benefits, including the ability to learn from small datasets, high accuracy, effective diagnosis, non-invasive approach, scalability, and accessibility, as well as the potential to support decision-making. The proposed model enables investigation into CNN's internal operations to increase its dependability and effectiveness. In terms of performance (sensitivity-based), this model outperforms the other models that are currently in the literature. The average sensitivity and F1 score for Melanoma in the three datasets are 98.33% and 96.67%, respectively. It is possible to develop and test this proposed model to more accurately and sensitively identify other related skin conditions from

dermoscopic images, such as Seborrheic keratosis, Basal cell carcinoma, Squamous cell carcinoma, melanocytic nevus, and others. This is important to note that while the proposed CNN offers various advantages, it is not intended to replace medical professionals. The role of the proposed method (Classification phase) is to augment the expertise of dermatologists and enhance the accuracy and efficiency of Melanoma diagnosis.

In the following chapter, the summary, contribution, limitation, future work recommendation, and conclusion of this research work will be presented based on the two proposed methods in Phase 1 and Phase 2.

CHAPTER 7: CONCLUSION

7.1   Background

The incidence of Melanoma is increasing globally, affecting countries such as the USA, Europe, Australia, and New Zealand (Blundo et al., 2021). Each year, a significant number of individuals succumb to Melanoma due to delayed detection, emphasizing the critical importance of early diagnosis to prevent these untimely deaths. Standard clinical methods (Davis et al., 2019) for Melanoma assessment utilize various techniques such as the ABCD rules, 3-point checklist, 7-point checklist, and CASH, which are employed by diagnostic centers, clinics, and hospitals. These methods combine different key features of skin conditions, including lesion asymmetry, regularity of pigment network, blue white veil, dots-globules, and lesion colors, to accurately analyze dermoscopic images. However, the existing clinical method for Melanoma assessment, which involves manual assessment, has several drawbacks. Firstly, it heavily relies on the expertise and subjective interpretation of the clinician, introducing variability and inconsistency in diagnoses. This subjectivity increases the risk of misdiagnosis and delayed treatment, potentially compromising patient outcomes. Furthermore, manual assessment is time-consuming, requiring significant effort from clinicians to carefully analyze dermoscopic images and assess various key features of skin conditions. This time-intensive process can contribute to delays in diagnosis and treatment initiation. Moreover, relying solely on manual assessment can impede the scalability of Melanoma screening programs, as it necessitates skilled dermatologists or clinicians to perform the evaluations. This limitation may hinder access to timely and accurate diagnoses, especially in regions with limited healthcare resources or a scarcity of specialists. Consequently, the disadvantages of manual assessment underscore the need for more objective and automated methods to enhance Melanoma detection and diagnosis.

In this research, the most common features associated with Melanoma are analyzed and classified to fit into the clinically practiced methods (ABCD rules, 3-point checklist, 7-point checklist, and CASH). The results of analyzing these selected features are then combined to explain the screening test results using deep learning. Finally, a deep learning-based screening test is proposed to provide instant Melanoma diagnosis results. This chapter revisits the proposed research works based on the two phases. It highlights the contributions, limitations, and future recommendations of this research in separate sections.

7.2   Summary of the Proposed Work

The aim of this research is to analyze dermoscopic images based on standard clinical methods (see Figure 3.5 of Chapter 3: Research Methodology). Different algorithms such as image processing, pattern recognition, computer vision, and deep learning were engaged to extract the feature information from the dermoscopic images. The whole proposed approach was divided into two phases based on three research objectives. Phase 1 (Feature Engineering) has two steps: Step 1 (Skin Lesion Segmentation) and Step 2 (Dermatological Feature Analysis), together representing Objectives 01 and 03. On the other hand, Phase 2 (Classification) represents Objectives 02 and 03.

Dermoscopic images are segmented by isolating the lesion from the healthy skin area. Deep learning-based semantic segmentation methods are employed for this completely automatic segmentation process. This process facilitates the determination of the disease-affected area, enabling the precise execution of the next step, which involves feature extraction in this research.

Dermatological asymmetry property is then determined from the segmented images, encompassing both symmetric and asymmetric patterns on both sides or on one side. The

pigment network is detected and classified as either typical (regular) or atypical (irregular). Additionally, the presence or absence of a blue white veil is identified on the skin lesion in the dermoscopic images. The next step involves detecting dots and globules as part of the feature extraction process from the dermoscopic images. Based on the color and shape of the dots and globules, they are classified into atypical, present, and absent categories. Finally, the dermoscopic images undergo lesion color analysis. This process involves counting the total number of different colors present on the skin lesion, enabling the classification of skin lesions into different classes based on the total number of colors contained within the lesion. By extracting these five features, the feature extraction process is completed. These five features provide an explanation for diagnosing Melanoma that is comprehensible to both experts and non-experts. At the end, three features (asymmetry, pigment network, and blue white veil) are engaged to apply on clinically 3-point checklist method to diagnose Melanoma and then other two features (dots-globules, and lesion colors) are checked also to make the final diagnosis. Fusion of these five dermoscopic features describe the skin condition too.

A classificaition method is proposed which is based on deep learning for instant diagnosis results without any explanation (Phase 2). Later fusion of five features and classification methods are compared by using a completely unknown test set. The method of Phase 2 can be used as a screening test for diagnosing Melanoma.

7.3   Contribution of the Proposed Work

The contributions of this research are reported based on the research phases. The primary aim of this research was to analyze dermoscopic images with the goal of early Melanoma detection. To accomplish this overarching objective, several supporting sub-goals were pursued. Throughout the course of achieving these goals, several notable

contributions have emerged, which can be categorized based on the two research phases. Phase 1 with two steps is the combination of Objective 01 and Objective 03. On the other hand, Phase 2 is the combination of Objective 02 and Objective 03.

Objective 01: To perform feature engineering for five dermoscopic features from dermoscopic images based on lesion segmentation and feature-specific deep learning models.

Objective 03: To evaluate feature-engineered and automated feature-identified CNN methods against state-of-the-art techniques in determining the efficacy of Melanoma detection.

Phase 1 (Step 1) Contribution: At first, the Objective 01 focuses on the segmentation of dermoscopic images. Five algorithms (i.e., simple semantic segmentation, semantic segmentation using dilated convolutions, semantic segmentation with Tversky loss function, semantic segmentation with Dice loss function, and modified U-Net) were proposed and applied to different datasets to achieve automated segmentation of skin lesions (see contribution of first and third objectives in Table 7.1). Additionally, an ablation study was conducted to identify a suitable loss function (Tversky) for semantic segmentation. This finding is valuable for the specific task of lesion segmentation and for similar segmentation tasks. These advancements contribute to effectively determining the disease-infected area within dermoscopic images.

Phase 1 (Step 2) Contribution: Then, Objective 01 focuses on feature extraction. Various algorithms and classifiers were utilized to analyze and classify the five common dermoscopic features. Some of these algorithms and classifiers were proposed and demonstrated as the top performers in their respective analyses. Each process of analyzing dermoscopic features, including asymmetry, pigment network, blue white veil, dots-

globules, and lesion colors, has made its own unique contribution (i.e., an algorithm for dermatological asymmetry analysis, directional imaging algorithm to isolate pigment network, a CNN model to classify pigment network, an algorithm for detecting blue white veil, an algorithm for analyzing dots-globules, and an algorithm for detecting different lesion colors). Building upon these contributions (see the contribution of the first and third objectives in Table 7.1), a comprehensive explanation-based dermoscopic feature extraction and analysis method was developed on multiracial datasets. This feature extraction and analysis method serves as an interpreter for the 2nd Research Objective, catering to both experts (skin experts, dermatologists, and medical professionals) and non-experts, providing a comprehensive explanation of the diagnostic outcomes. Additionally, during the features analysis process, multiple non-annotated datasets became annotated datasets such as- MED-NODE (annotated with symmetrical, blue white veil, and lesion colors information) and ISIC2016 (annotated with symmetrical information). Objective 03 evaluates the proposed models of both steps against state-of-the-art models for the dermoscopic image segmentation and features analysis.

Objective 02: To experiment with diverse CNN models for automatic computer-based image classification for Melanoma detection from dermoscopic images.

Objective 03: To evaluate feature-engineered and automated feature-identified CNN methods against state-of-the-art techniques in determining the efficacy of Melanoma detection.

Phase 2 Contribution: The Objective 02 focuses on dermoscopic image classification using deep learning as a rapid Melanoma detection method without any explanation. This proposed deep learning-based detection method (does not provide explanations) contributes to the detection of malignant Melanoma as a screening test approach (see the contribution of the second and third objectives in Table 7.1). Objective 03 evaluates the

proposed model of Phase 2 against state-of-the-art models for the dermoscopic image classification. Table 7.1 presents a comprehensive overview of all the contributions made in this research, categorized based on the two research phases and three research objectives.

Table 7.1: The research contributions.

Phase	Objectives		Contributions
1	1st+3rd: **Segmentation** (Chapter 4)		1. Simple semantic segmentation network. 2. Semantic network with dilated convolution. 3. Semantic segmentation with Tversky loss function. 4. Semantic segmentation with Dice loss function 5. Modified U-Net.
	1st+3rd: **Feature analysis** (Chapter 5)	Asymmetry	6. An algorithm for Dermatological asymmetry analysis. 7. MED-NODE becomes annotated dataset with symmetrical information. 8. ISIC2016 becomes annotated dataset with symmetrical information.
		Pigment Network	9. Directional imaging algorithm. 10. Proposed CNN model.
		Blue white veil	11. An algorithm for detecting blue white veils. 12. MED-NODE becomes annotated dataset with blue white veil information.
		Dots-globules	13. An algorithm for analyzing Dots-globules.
		Lesion Colors	14. An algorithm for detecting different lesion colors. 15. MED-NODE becomes annotated dataset with lesion color information.
2	2nd+3rd: **Classification** (Chapter 6)		16. A deep CNN model parameterized with activation function.

7.4   Limitation of the Proposed Work

To save lives, early diagnosis of Melanoma is crucial. The process of early Melanoma diagnosis relies on clinically established conventional methods, including the ABCD rules, 3-point and 7-point checklists, and CASH. This research primarily focuses on exploring both conventional and non-conventional approaches for Melanoma recognition. However, it is critical to recognize the shortcomings of this research, which are listed below.

Lack of high-quality labeled datasets: In this research, different deep learning models heavily rely on large and well-curated datasets for training. The scarcity of comprehensive, accurately labeled dermoscopic image datasets can limit the performance and generalizability of those proposed deep learning models.

Variation of dermoscopic images: Dermoscopic images can exhibit variations in image quality, such as differences in lighting, resolution, and image artifacts. Here, these variations can impact the performance of feature extraction proposed algorithms (such as pigment network, blue white veil, and total number of colors) and proposed deep learning models, leading to less reliable results.

Influence of image artifacts and noise: The proposed image processing and computer vision algorithms used for feature extraction can be sensitive to artifacts and noise present in dermoscopic images. These artifacts, such as hairs, air bubbles, gels, or dust, can interfere with the accurate extraction of relevant features (pigment network and lesion color distribution), potentially leading to false positive or false negative results.

Classification of extracted features: The proposed algorithms and techniques of extracting features from dermoscopic images involve subjective decisions, such as

determining the presence or absence of certain features. Inter-observer variability among experts can introduce inconsistencies and affect the accuracy of feature-based methods.

Data imbalance and class distribution: A data set with imbalanced class proportions is considered imbalanced. Majority classes refer to those that make up a large proportion of the data set. Melanoma is a relatively rare condition compared to other skin lesions, leading to imbalanced class distributions in all datasets (including PH2, MED-NODE, and HAM10000). This imbalance can affect the performance of proposed deep learning models, making them more prone to biases and favoring the majority class.

Limited generalization to diverse populations: Most dermoscopic image datasets used in research are often biased towards specific populations, which may not represent the diversity seen in real-world clinical settings. The lack of diversity in dermoscopic image datasets poses challenges for accurate medical imaging algorithms and deep learning models. This limitation can impact the generalization of the proposed two methods to different ethnicities and skin types.

Analyzing the limited number of features: In this research, the analysis focuses on the five most prevalent features (asymmetry, pigment network, blue white veil, dots and globules, and total number of colors) for diagnosing Melanoma on dermoscopic images. However, it is essential to confess that based on standard methods like the ABCD rules, 3-point and 7-point checklists, and CASH, there exist additional features on skin lesions that can serve as potential indicators of Melanoma.

7.5   Recommendation for Future Works

This research document reports a journey that can be divided into two main phases: a road toward developing a clinically inspired dermoscopic image analysis method; and a

thorough study of the deep learning-based Melanoma detection method. Both phases resulted in relevant outputs for the research community. Based on the outputs and the knowledge acquired during the analysis of dermoscopic images in two phases, there are some possible guidelines for future work in this field. Table 7.2 recommends some future works based on the research objectives.

Table 7.2: The recommended future works.

Objective		Future Works
1st+3rd: Segmentation (Chapter 4)		1. Aim for higher accuracy. 2. Increasing the class numbers 2 to 3; such as background, region of interest, and lesion. 3. Comparing with other semantic segmentation networks such as Xnet, FCN, and Znet.
1st+3rd: Feature analysis (Chapter 5)	Asymmetry	4. Splitting into 8 portions for more precision.
	Pigment Network	5. Classifying atypical networks into prominent and delicate networks.
	Blue white veil	6. Aim for higher accuracy and increasing training data. Measuring the area in the percentage of lesion area.
	Dots-Globules	7. Color and shape analysis individually and counting the percentage of area. 8. Aim for more precise detection of the lesion.
	Lesion Colors	9. Diagnosing Melanoma by using color feature only.
	Fusion	10. Including more features (irregular borders, irregular streaks & blotches, and atypical vascular patterns) in the Melanoma diagnosis system.
2nd+3rd: Classification (Chapter 6)		11. Increasing the number of disease classes and training data. 12. Using custom activation function. 13. Design vision transformer model for complex analysis.

Based on the experiments from Chapters 4 to Chapter 6, here are more recommendations for future work in the field of automatic Melanoma detection on dermoscopic images using image processing, pattern recognition, and deep learning methods:

Diversity of the dataset: Collect and curate a more diverse and comprehensive dataset that includes a wider range of skin types, ethnicities, and clinical variations with the ground truth of different dermatological features. This will help to improve the generalization and robustness of those proposed methods.

Understanding of deep learning models: To enhance clinicians' trust in deep learning models for Melanoma detection from dermoscopic images, interpretability techniques like attention mechanisms, saliency maps, and model visualization can be employed. Attention mechanisms highlight regions of interest in an image, aiding clinicians in understanding the model's decision-making process. Saliency maps provide visual explanations for model predictions, offering insights into lesion features influencing diagnoses. Model visualization methods reveal how different layers capture distinct image aspects. These techniques bridge the gap between model predictions and clinical understanding, increasing transparency and acceptance of AI-driven diagnoses in Melanoma detection.

Approaches to lifelong learning and incremental learning: Create techniques that allow deep learning models to continuously pick up new information over time, allowing them to adjust to new data and continually enhance their performance. This can be especially helpful in actual clinical situations where the distribution of Melanoma patients is subject to vary over time.

Telemedicine and mobile applications integration: Investigate the integration of proposed Melanoma detection methods with telemedicine platforms and mobile applications. This can enable remote screening, monitoring, and early detection of Melanoma, especially in underserved or remote areas where access to dermatologists is limited.

Clinical trials and verification: Conduct rigorous validation studies and robust clinical trials to thoroughly assess the efficacy and clinical viability of the proposed Melanoma detection procedure. To ensure seamless integration of this system into current clinical practices, work closely with dermatologists and other medical professionals. Collaborating with healthcare professionals can confirm the proposed method's influence on patient outcomes, confirm its dependability, and make any necessary adjustments to the procedure, fostering a more reliable and effective detection approach.

By pursuing these future research directions, researchers can advance the field of automatic Melanoma diagnosis and contribute to more accurate, efficient, and accessible Melanoma screening and early detection, ultimately improving patient care and outcomes.

7.6 Conclusion

In conclusion, the importance of timely and automated Melanoma detection with proper interpretation cannot be overstated. This research output brings numerous benefits to various groups, including researchers, clinical practitioners, and the community at large. For researchers, automated Melanoma detection with explanation offers opportunities for further advancements in imaging algorithm development and the exploration of novel computer vision techniques. It allows for the development of more

accurate and efficient methods, ultimately contributing to the overall progress in Melanoma diagnosis.

For clinical practitioners, automated Melanoma detection provides a valuable tool that can aid in early and accurate diagnosis. By leveraging advanced algorithms and technologies, clinicians can enhance their diagnostic capabilities, leading to improved patient outcomes. Rapid and reliable detection enables prompt intervention, which is crucial for effective treatment and potentially lifesaving for patients.

The community also reaps significant benefits from timely automated Melanoma detection. Early detection facilitates the identification of Melanoma at its earliest stages when it is most treatable. This translates to higher survival rates, reduced healthcare costs, and improved quality of life for individuals affected by Melanoma. Additionally, widespread implementation of automated detection systems can increase accessibility, allowing for earlier detection in underserved areas or regions with limited healthcare resources.

Overall, the importance of timely and automated Melanoma detection cannot be emphasized enough. It has the potential to revolutionize Melanoma diagnosis, empowering researchers, enabling accurate clinical decision-making, and ultimately saving lives. By embracing and advancing these imaging and computer vision technologies, we can make substantial progress in the fight against Melanoma and contribute to the well-being of individuals and communities worldwide.

REFERENCES

Abbas, Q., Celebi, M., Serrano, C., Fondón García, I., & Ma, G. (2013, January). Pattern classification of dermoscopy images: A perceptually uniform model. *Pattern Recognition*, *46*(1), 86–97. https://doi.org/10.1016/j.patcog.2012.07.027

Abood, L. K. (2018, September 11). Contrast enhancement of infrared images using Adaptive Histogram Equalization (AHE) with Contrast Limited Adaptive Histogram Equalization (CLAHE). *Iraqi Journal of Physics*, *16*(37), 127–135. https://doi.org/10.30723/ijp.v16i37.84

Adegun, A. A., & Viriri, S. (2020). Deep Learning-Based System for Automatic Melanoma Detection. *IEEE Access*, *8*, 7160–7172. https://doi.org/10.1109/access.2019.2962812.

Adler, N. R., Kelly, J. W., Guitera, P., Menzies, S. W., Chamberlain, A. J., Fishburn, P., Button-Sloan, A. E., Heal, C., Soyer, H. P., & Thompson, J. F. (2018, December 2). Methods of melanoma detection and of skin monitoring for individuals at high risk of melanoma: new Australian clinical practice. *Medical Journal of Australia*, *210*(1), 41–47. https://doi.org/10.5694/mja2.12033

Al-Ahmad, M., Arifohdzic, N., & Fakim, N. (2013). Allergic diseases of the skin and drug allergies – 2005. Common causes of allergic contact dermatitis in Kuwait. *World Allergy Organization Journal*, *6*, P95. https://doi.org/10.1186/1939-4551-6-s1-p95

Al-Azzawi, N. A. (2018, July). Color Medical Imaging Fusion Based on Principle Component Analysis and F-Transform. *Pattern Recognition and Image Analysis*, *28*(3), 393–399. https://doi.org/10.1134/s105466181803001x

Alfed, N., Khelifi, F., Bouridane, A., & Seker, H. (2015, August). Pigment network-based skin cancer detection. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Presented at the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan. doi:10.1109/embc.2015.7320056.

Ali, A. R., Li, J., & O'Shea, S. J. (2020, June 16). Towards the automatic detection of skin lesion shape asymmetry, color variegation and diameter in dermoscopic images. *PLOS ONE*, *15*(6), e0234352. https://doi.org/10.1371/journal.pone.0234352

Almubarak, H., Stanley, R., Stoecker, W., & Moss, R. (2017, July 25). Fuzzy Color Clustering for Melanoma Diagnosis in Dermoscopy Images. *Information*, *8*(3), 89. https://doi.org/10.3390/info8030089.

Alwakid, G., Gouda, W., Humayun, M., & Jhanjhi, N. Z. (2023, May 22). Diagnosing Melanomas in Dermoscopy Images Using Deep Learning. *Diagnostics*, *13*(10), 1815. https://doi.org/10.3390/diagnostics13101815

Amelard, R., Glaister, J., Wong, A., & Clausi, D. A. (2015, March). High-Level Intuitive Features (HLIFs) for Intuitive Skin Lesion Description. *IEEE Transactions on Biomedical Engineering*, *62*(3), 820–831. https://doi.org/10.1109/tbme.2014.2365518

Anand, V., Gupta, S., Koundal, D., & Singh, K. (2023, March). Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images. *Expert Systems with Applications*, *213*, 119230. https://doi.org/10.1016/j.eswa.2022.119230

Argenziano, G., Catricalà, C., Ardigo, M., Buccini, P., De Simone, P., Eibenschutz, L., Ferrari, A., Mariani, G., Silipo, V., Sperduti, I., & Zalaudek, I. (2011, March 10). Seven-point checklist of dermoscopy revisited. *British Journal of Dermatology*, *164*(4), 785–790. https://doi.org/10.1111/j.1365-2133.2010.10194.x

Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., & Delfino, M. (1998, December 1). Epiluminescence Microscopy for the Diagnosis of Doubtful Melanocytic Skin Lesions. *Archives of Dermatology*, *134*(12). https://doi.org/10.1001/archderm.134.12.1563

Argenziano, G., Zalaudek, I., Corona, R., Sera, F., Cicale, L., Petrillo, G., Ruocco, E., Hofmann-Wellenhof, R., & Soyer, H. P. (2004, December 1). Vascular Structures in Skin Tumors. *Archives of Dermatology*, *140*(12). https://doi.org/10.1001/archderm.140.12.1485

Ayas, S. (2022, November 28). Multiclass skin lesion classification in dermoscopic images using swin transformer model. *Neural Computing and Applications*, *35*(9), 6713–6722. https://doi.org/10.1007/s00521-022-08053-z

Bansal, P., Garg, R., & Soni, P. (2022, June). Detection of melanoma in dermoscopic images by integrating features extracted using handcrafted and deep learning models. *Computers & Industrial Engineering*, *168*, 108060. https://doi.org/10.1016/j.cie.2022.108060

Barata, C., Marques, J. S., & Rozeira, J. (2012). A system for the detection of pigment network in dermoscopy images using directional filters. *IEEE Transactions on Bio-Medical Engineering*, *59*(10), 2744–2754. doi:10.1109/TBME.2012.2209423.

Bareiro Paniagua, L. R., Leguizamón Correa, D. N., Pinto-Roa, D., Vázquez Noguera, J. L., & Salgueiro Toledo, L. A. (2016, August 1). Computerized Medical Diagnosis of Melanocytic Lesions based on the ABCD approach. *CLEI Electronic Journal*. https://doi.org/10.19153/cleiej.19.2.5

Barhoumi, W., & Baâzaoui, A. (2014, June). Pigment network detection in dermatoscopic images for melanoma diagnosis. *IRBM*, *35*(3), 128–138. https://doi.org/10.1016/j.irbm.2013.12.010

Barın, S., & Güraksın, G. E. (2022, October). An automatic skin lesion segmentation system with hybrid FCN-ResAlexNet. *Engineering Science and Technology, an International Journal*, *34*, 101174. https://doi.org/10.1016/j.jestch.2022.101174

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. *Computer Vision – ECCV 2006*, 404–417. https://doi.org/10.1007/11744023_32

Bernardo, L. S., Damaševičius, R., Ling, S. H., de Albuquerque, V. H. C., & Tavares, J. M. R. S. (2022, October 28). Modified SqueezeNet Architecture for Parkinson's Disease Detection Based on Keypress Data. *Biomedicines*, *10*(11), 2746. https://doi.org/10.3390/biomedicines10112746

Blundo, A., Cignoni, A., Banfi, T., & Ciuti, G. (2021, April 21). Comparative Analysis of Diagnostic Techniques for Melanoma Detection: A Systematic Review of Diagnostic Test Accuracy Studies and Meta-Analysis. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.637069

Boustany, J. (2013, March 30). Accès et réutilisation des données publiques. état des lieux en France. *Les Cahiers Du Numérique*, *9*(1), 9–19. https://doi.org/10.3166/lcn.9.1.9-19

Cacciapuoti, S., Di Leo, G., Ferro, M., Liguori, C., Masarà, A., Scalvenzi, M., Sommella, P., & Fabbrocini, G. (2020, June 24). A Measurement Software for Professional Training in Early Detection of Melanoma. *Applied Sciences*, *10*(12), 4351. https://doi.org/10.3390/app10124351.

Cavalcanti, P. G., & Scharcanski, J. (2011, September). Automated prescreening of pigmented skin lesions using standard cameras. *Computerized Medical Imaging and Graphics*, *35*(6), 481–491. https://doi.org/10.1016/j.compmedimag.2011.02.007

Celebi, M. E., Iyatomi, H., Stoecker, W. V., Moss, R. H., Rabinovitz, H. S., Argenziano, G., & Soyer, H. P. (2008, December). Automatic detection of blue-white veil and related structures in dermoscopy images. *Computerized Medical Imaging and Graphics*, *32*(8), 670–677. https://doi.org/10.1016/j.compmedimag.2008.08.003.

Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007, September). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, *31*(6), 362–373. https://doi.org/10.1016/j.compmedimag.2007.01.003

Chakravorty, R., Liang, S., Abedini, M., & Garnavi, R. (2016, August). Dermatologist-like feature extraction from skin lesion for improved asymmetry classification in PH2 database. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Presented at the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA. doi:10.1109/embc.2016.7591569

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018, April 1). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/tpami.2017.2699184

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Lecture*

Notes in Computer Science (pp. 833–851). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-01234-2_49

Chen, S., Zhao, S., & Lan, Q. (2022, March 9). Residual Block Based Nested U-Type Architecture for Multi-Modal Brain Tumor Image Segmentation. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.832824

Chen, X., & Bagci, U. (2011, July 25). 3D automatic anatomy segmentation based on iterative graph-cut-ASM. *Medical Physics*, *38*(8), 4610–4622. https://doi.org/10.1118/1.3602070

Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., … Halpern, A. (2017). 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). Retrieved from http://arxiv.org/abs/1710.05006

Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. (2018, April). 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). *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. https://doi.org/10.1109/isbi.2018.8363547

Combalia, M., Codella, N. C. F., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., … Malvehy, J. (2019). BCN20000: Dermoscopic Lesions in the Wild. Retrieved from http://arxiv.org/abs/1908.02288

Cortes, C., & Vapnik, V. (1995, September). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

Crnjanski, J., Krstić, M., Totović, A., Pleros, N., & Gvozdić, D. (2021, April 16). Adaptive sigmoid-like and PReLU activation functions for all-optical perceptron. *Optics Letters*, *46*(9), 2003. https://doi.org/10.1364/ol.422930

Crum, W., Camara, O., & Hill, D. (2006, November). Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. *IEEE Transactions on Medical Imaging, 25*(11), 1451–1461. https://doi.org/10.1109/tmi.2006.880587

Damian, F. A., Moldovanu, S., & Moraru, L. (2021, October 20). Skin lesions asymmetry estimation using artificial neural networks. *2021 25th International Conference on System Theory, Control and Computing (ICSTCC)*. Presented at the 2021 25th International Conference on System Theory, Control and Computing (ICSTCC), Iasi, Romania. doi:10.1109/icstcc52150.2021.9607133

Davis, L. E., Shalin, S. C., & Tackett, A. J. (2019, August 1). Current state of melanoma diagnosis and treatment. *Cancer Biology & Therapy*, *20*(11), 1366–1379. https://doi.org/10.1080/15384047.2019.1640032

Di Leo, G., Paolillo, A., Sommella, P., & Fabbrocini, G. (2010). Automatic diagnosis of melanoma: A software system based on the 7-point check-list. *2010 43rd Hawaii International Conference on System Sciences*. Presented at the 2010 43rd Hawaii

International Conference on System Sciences, Honolulu, Hawaii, USA. doi:10.1109/hicss.2010.76

Doi, K. (2005, January). Current status and future potential of computer-aided diagnosis in medical imaging. *The British Journal of Radiology*, *78*(suppl_1), s3–s19. https://doi.org/10.1259/bjr/82933343

Doi, K. (2007, June). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, *31*(4–5), 198–211. https://doi.org/10.1016/j.compmedimag.2007.02.002

Eadie, L. H., Taylor, P., & Gibson, A. P. (2012, January). A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *European Journal of Radiology*, *81*(1), e70–e76. https://doi.org/10.1016/j.ejrad.2011.01.098

Eltayef, K., Li, Y., & Liu, X. (2017, February). Detection of pigment networks in dermoscopy images. In *Journal of Physics: Conference Series* (Vol. 787, No. 1, p. 012033). IOP Publishing.

Fang, J., Xu, H., Xu, P., & Wang, Z. (2017, February). Spectral estimation of fluorescent lamps using RGB digital camera and standard color chart. *Optik*, *130*, 50–60. https://doi.org/10.1016/j.ijleo.2016.11.014

Fei-Fei, L., Deng, J., & Li, K. (2010, March 22). ImageNet: Constructing a large-scale image database. *Journal of Vision*, *9*(8), 1037–1037. https://doi.org/10.1167/9.8.1037

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J., Comber, H., Forman, D., & Bray, F. (2013, April). Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *European Journal of Cancer*, *49*(6), 1374–1403. https://doi.org/10.1016/j.ejca.2012.12.027

Friedman, R. J., Rigel, D. S., & Kopf, A. W. (1985, May 1). Early Detection of Malignant Melanoma: The Role of Physician Examination and Self-Examination of the Skin. *CA: A Cancer Journal for Clinicians*, *35*(3), 130–151. https://doi.org/10.3322/canjclin.35.3.130

Frühauf, J., Leinweber, B., Fink-Puches, R., Ahlgrimm-Siess, V., Richtig, E., Wolf, I., Niederkorn, A., Quehenberger, F., & Hofmann-Wellenhof, R. (2011, April 20). Patient acceptance and diagnostic utility of automated digital image analysis of pigmented skin lesions. *Journal of the European Academy of Dermatology and Venereology*, *26*(3), 368–372. https://doi.org/10.1111/j.1468-3083.2011.04081.x

Gadeliya Goodson, A., & Grossman, D. (2009, May). Strategies for early melanoma detection: Approaches to the patient with nevi. *Journal of the American Academy of Dermatology*, *60*(5), 719–735. https://doi.org/10.1016/j.jaad.2008.10.065

García Arroyo, J. L., & García Zapirain, B. (2014, January). Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis. *Computers in Biology and Medicine*, *44*, 144–157. https://doi.org/10.1016/j.compbiomed.2013.11.002

García Arroyo, J. L., & García Zapirain, B. (2014, January). Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis. *Computers in Biology and Medicine*, *44*, 144–157. https://doi.org/10.1016/j.compbiomed.2013.11.002

Garcia-Arroyo, J. L., & Garcia-Zapirain, B. (2018, January). Recognition of pigment network pattern in dermoscopy images based on fuzzy classification of pixels. *Computer Methods and Programs in Biomedicine*, *153*, 61–69. https://doi.org/10.1016/j.cmpb.2017.10.005

Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Kniep, H., Baltruschat, I., Werner, R., & Schlaefer, A. (2020, February). Skin Lesion Classification Using CNNs With Patch-Based Attention and Diagnosis-Guided Loss Weighting. *IEEE Transactions on Biomedical Engineering*, *67*(2), 495–503. https://doi.org/10.1109/tbme.2019.2915839

Giotis, I., Molders, N., Land, S., Biehl, M., Jonkman, M. F., & Petkov, N. (2015, November). MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Systems With Applications*, *42*(19), 6578–6585. https://doi.org/10.1016/j.eswa.2015.04.034

Goyal, M., Oakley, A., Bansal, P., Dancey, D., & Yap, M. H. (2020). Skin Lesion Segmentation in Dermoscopic Images With Ensemble Deep Learning Methods. *IEEE Access*, *8*, 4171–4181. https://doi.org/10.1109/access.2019.2960504.

Greenspan, H., Müller, H., & Syeda-Mahmood, T. (2013). *Medical content-based retrieval for clinical decision support*. Springer Berlin Heidelberg.

Gutman, D., Codella, N. C. F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. (2016). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). Retrieved from http://arxiv.org/abs/1605.01397

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/iccv.2015.123

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*. https://doi.org/10.1109/cvpr.2016.90

Henning, J. S., Stein, J. A., Yeung, J., Dusza, S. W., Marghoob, A. A., Rabinovitz, H. S., Polsky, D., & Kopf, A. W. (2008, April 1). CASH Algorithm for Dermoscopy Revisited. *Archives of Dermatology*, *144*(4). https://doi.org/10.1001/archderm.144.4.554

Hosny, K. M., Kassem, M. A., & Foaud, M. M. (2018, December). Skin Cancer Classification using Deep Learning and Transfer Learning. *2018 9th Cairo*

International Biomedical Engineering Conference (CIBEC). https://doi.org/10.1109/cibec.2018.8641762

Hosny, K. M., Kassem, M. A., & Fouad, M. M. (2020, June 30). Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet. *Journal of Digital Imaging*, *33*(5), 1325–1334. https://doi.org/10.1007/s10278-020-00371-9

Hurtado, J., & Reales, F. (2021, August 1). A machine learning approach for the recognition of melanoma skin cancer on macroscopic images. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *19*(4), 1357. https://doi.org/10.12928/telkomnika.v19i4.20292

Janney, B. J., Roslin, S. E., & Shelcy, M. J. (2018, July). A Comparative Analysis of Skin Cancer Detection based on SVM, ANN and Naive Bayes Classifier. *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)*. https://doi.org/10.1109/icrieece44171.2018.9008943

Jaworek-Korjakowska, J., & Tadeusiewicz, R. (2013, September). Assessment of dots and globules in dermoscopic color images as one of the 7-point check list criteria. *2013 IEEE International Conference on Image Processing*. Presented at the 2013 20th IEEE International Conference on Image Processing (ICIP), Melbourne, Australia. doi:10.1109/icip.2013.6738299.

Kassem, M. A., Hosny, K. M., & Fouad, M. M. (2020). Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning. *IEEE Access*, *8*, 114822–114832. https://doi.org/10.1109/access.2020.3003890

Kawahara, J., Daneshvar, S., Argenziano, G., & Hamarneh, G. (2019, March). Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 538–546. https://doi.org/10.1109/jbhi.2018.2824327

Khan, M. A., Sharif, M., Akram, T., Damaševičius, R., & Maskeliūnas, R. (2021, April 29). Skin Lesion Segmentation and Multiclass Classification Using Deep Learning Features and Improved Moth Flame Optimization. *Diagnostics*, *11*(5), 811. https://doi.org/10.3390/diagnostics11050811

Kondaveeti, H. K., & Edupuganti, P. (2020, December 30). Skin Cancer Classification using Transfer Learning. *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*. https://doi.org/10.1109/icatmri51801.2020.9398388

Kopf, A. W., Elbaum, M., & Provost, N. (1997, February). The use of dermoscopy and digital imaging in the diagnosis of cutaneous malignant melanoma. *Skin Research and Technology*, *3*(1), 1–7. https://doi.org/10.1111/j.1600-0846.1997.tb00152.x

Korotkov, K., & Garcia, R. (2012, October). Computerized analysis of pigmented skin lesions: A review. *Artificial Intelligence in Medicine*, *56*(2), 69–90. https://doi.org/10.1016/j.artmed.2012.08.002

Korotkov, K., Quintana, J., Puig, S., Malvehy, J., & Garcia, R. (2015, January). A New Total Body Scanning System for Automatic Change Detection in Multiple Pigmented Skin Lesions. *IEEE Transactions on Medical Imaging*, *34*(1), 317–338. https://doi.org/10.1109/tmi.2014.2357715

Krishna, S., & Remya Ajai, A. S. (2019, December). Analysis of Three Point checklist and ABCD methods for the feature extraction of dermoscopic images to detect melanoma. *2019 9th International Symposium on Embedded Computing and System Design (ISED)*. Presented at the 2019 9th International Symposium on Embedded Computing and System Design (ISED), Kollam, India. doi:10.1109/ised48680.2019.9096226

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017, May 24). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kropidlowski, K., Kociolek, M., Strzelecki, M., & Czubinski, D. (2016). Blue whitish veil, atypical vascular pattern and regression structures detection in skin lesions images. In *Lecture Notes in Computer Science*. *Computer Vision and Graphics* (pp. 418–428). doi:10.1007/978-3-319-46418-3_37.

Kropidlowski, K., Kociolek, M., Strzelecki, M., & Czubinski, D. (2015, September). Nevus atypical pigment network distinction and irregular streaks detection in skin lesions images. *2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. https://doi.org/10.1109/spa.2015.7365135

Kuehni, R. G. (1976, May 1). Color-tolerance data and the tentative CIE 1976 L*a*b* formula. *Journal of the Optical Society of America*, *66*(5), 497. https://doi.org/10.1364/josa.66.000497

Lamb, J., Stone, A., Li, Z., Hu, J., & James, A. (2022, August). 670 Average RGB color value to categorize skin color of dermatology images. *Journal of Investigative Dermatology*, *142*(8), S115. https://doi.org/10.1016/j.jid.2022.05.681

Landa, E., & Fairchild, M. (2005). Charting Color from the Eye of the Beholder. *American Scientist*, *93*(5), 436. https://doi.org/10.1511/2005.55.975

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Li, Z., Zhang, H., Li, Z., & Ren, Z. (2022, July 15). Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation. *Applied Sciences*, *12*(14), 7149. https://doi.org/10.3390/app12147149

Lio, P. A., & Nghiem, P. (2004, May). Interactive Atlas of Dermoscopy. *Journal of the American Academy of Dermatology*, *50*(5), 807–808. https://doi.org/10.1016/j.jaad.2003.07.029

Lu, L. (2020, June). Dying ReLU and Initialization: Theory and Numerical Examples. *Communications in Computational Physics*, *28*(5), 1671–1706. https://doi.org/10.4208/cicp.oa-2020-0165

Ma, Z., & Yin, S. (2018, October). Deep attention network for melanoma detection improved by color constancy. *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*. Presented at the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou. doi:10.1109/itme.2018.00037.

Mackie, R., & Doherty, V. (1991, March). Seven-point checklist for melanoma. *Clinical and Experimental Dermatology*, *16*(2), 151–152. https://doi.org/10.1111/j.1365-2230.1991.tb00329.x

Madooei, A., & Drew, M. S. (2013, January 1). A Colour Palette for Automatic Detection of Blue-White Veil. *Color and Imaging Conference*, *21*(1), 200–205. https://doi.org/10.2352/cic.2013.21.1.art00036

Madooei, A., Drew, M. S., & Hajimirsadeghi, H. (2019, March). Learning to Detect Blue–White Structures in Dermoscopy Images With Weak Supervision. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 779–786. https://doi.org/10.1109/jbhi.2018.2835405

Madooei, A., Drew, M. S., Sadeghi, M., & Atkins, M. S. (2013). Automatic Detection of Blue-White Veil by Discrete Colour Matching in Dermoscopy Images. *Advanced Information Systems Engineering*, 453–460. https://doi.org/10.1007/978-3-642-40760-4_57

Maglogiannis, I., & Delibasis, K. K. (2015, February). Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy. *Computer Methods and Programs in Biomedicine*, *118*(2), 124–133. https://doi.org/10.1016/j.cmpb.2014.12.001.

Maglogiannis, I., & Doukas, C. (2009, September). Overview of Advanced Computer Vision Systems for Skin Lesions Characterization. *IEEE Transactions on Information Technology in Biomedicine*, *13*(5), 721–733. https://doi.org/10.1109/titb.2009.2017529

Mascarenhas, S., & Agarwal, M. (2021, November 19). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*. https://doi.org/10.1109/centcon52345.2021.9687944

Mayer, J. (1997, August). Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. *Medical Journal of Australia*, *167*(4), 206–210. https://doi.org/10.5694/j.1326-5377.1997.tb138847.x

Mendonca, T., Ferreira, P. M., Marques, J. S., Marcal, A. R. S., & Rozeira, J. (2013, July). PH^2 - A dermoscopic image database for research and benchmarking. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Presented at the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka. doi:10.1109/embc.2013.6610779.

Milczarski, P. (2017, September). Skin lesion symmetry of hue distribution. *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. Presented at the 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest. doi:10.1109/idaacs.2017.8095238.

Nambisan, A. K., Lama, N., Phan, T., Swinfard, S., Lama, B., Smith, C., Rajeh, A., Patel, G., Hagerty, J., Stoecker, W. V., & Stanley, R. J. (2022, November). Deep learning-based dot and globule segmentation with pixel and blob-based metrics for evaluation. *Intelligent Systems With Applications*, *16*, 200126. https://doi.org/10.1016/j.iswa.2022.200126.

Nezhadian, F. K., & Rashidi, S. (2017, October). Melanoma skin cancer detection using color and new texture features. *2017 Artificial Intelligence and Signal Processing Conference (AISP)*. Presented at the 2017 Artificial Intelligence and Signal Processing Conference (AISP), Shiraz. doi:10.1109/aisp.2017.8324108.

Ng, V. T., Fung, B. Y., & Lee, T. K. (2005, February). Determining the asymmetry of skin lesion with fuzzy borders. *Computers in Biology and Medicine*, *35*(2), 103–120. https://doi.org/10.1016/j.compbiomed.2003.11.004

Nowak, L. A., Grzesiak-Kopec[acute], K., & Ogorza[lstrok]ek, M. J. (2015). Melanocytic globules detection in skin lesion images. *12th International Symposium on Operations Research and Its Applications in Engineering, Technology and Management (ISORA 2015)*. Presented at the 12th International Symposium on Operations Research and its Applications in Engineering, Technology and Management (ISORA 2015), Luoyang, China. doi:10.1049/cp.2015.0624.

Nugraha, W., & Sasongko, A. (2022, May 21). Hyperparameter Tuning on Classification Algorithm with Grid Search. *SISTEMASI*, *11*(2), 391. https://doi.org/10.32520/stmsi.v11i2.1750

Oliveira, R. B., Papa, J. P., Pereira, A. S., & Tavares, J. M. R. S. (2016, July 15). Computational methods for pigmented skin lesion classification in images: review and future trends. *Neural Computing and Applications*, *29*(3), 613–636. https://doi.org/10.1007/s00521-016-2482-6

Pathan, S., Prabhu, K. G., & Siddalingaswamy, P. (2018, July). A methodological approach to classify typical and atypical pigment network patterns for melanoma diagnosis. *Biomedical Signal Processing and Control*, *44*, 25–37. https://doi.org/10.1016/j.bspc.2018.03.017.

Raikar, A., Sangani, S. P., & Hanabaratti, K. D. (2013, July). Diagnosis of melanomas by check-list method. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. Presented at the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode. doi:10.1109/icccnt.2013.6726775

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August 13). "Why Should I Trust You?" *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939778

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.

Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., Halpern, A., Helba, B., Kittler, H., Kose, K., Langer, S., Lioprys, K., Malvehy, J., Musthaq, S., Nanda, J., . . . Soyer, H. P. (2021, January 28). A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific Data*, *8*(1). https://doi.org/10.1038/s41597-021-00815-z

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015, April 11). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sabbaghi Mahmouei, S., Aldeen, M., Stoecker, W. V., & Garnavi, R. (2019, March). Biologically Inspired QuadTree Color Detection in Dermoscopy Images of Melanoma. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 570–577. https://doi.org/10.1109/jbhi.2018.2841428.

Sadeghi, M., Razmara, M., Lee, T. K., & Atkins, M. (2011, March). A novel method for detection of pigment network in dermoscopic images using graphs. *Computerized Medical Imaging and Graphics*, *35*(2), 137–143. https://doi.org/10.1016/j.compmedimag.2010.07.002

Sahiner, B., Heang-Ping Chan, & Hadjiiski, L. (2008, February). Performance Analysis of Three-Class Classifiers: Properties of a 3-D ROC Surface and the Normalized Volume Under the Surface for the Ideal Observer. *IEEE Transactions on Medical Imaging*, *27*(2), 215–227. https://doi.org/10.1109/tmi.2007.905822

Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. In *Springer eBooks* (pp. 379–387). https://doi.org/10.1007/978-3-319-67389-9_44

Schadendorf, D., Lebbe, C., Lorigan, P., & Puig, S. (2023). European Journal of Skin Cancer Editorial. *EJC Skin Cancer*, *1*, 100001. https://doi.org/10.1016/j.ejcskn.2023.100001

Sharma, A. K., Tiwari, S., Aggarwal, G., Goenka, N., Kumar, A., Chakrabarti, P., Chakrabarti, T., Gono, R., Leonowicz, Z., & Jasinski, M. (2022). Dermatologist-Level Classification of Skin Cancer Using Cascaded Ensembling of Convolutional Neural Network and Handcrafted Features Based Deep Neural Network. *IEEE Access*, *10*, 17920–17932. https://doi.org/10.1109/access.2022.3149824

She, Z., Liu, Y., & Damatoa, A. (2007, February). Combination of features from skin pattern and ABCD analysis for lesion classification. *Skin Research and Technology*, *13*(1), 25–33. https://doi.org/10.1111/j.1600-0846.2007.00181.x

Singh, R., & Rani, R. (2020). Semantic Segmentation using Deep Convolutional Neural Network: A Review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3565919

Soyer, H. P., Argenziano, G., Zalaudek, I., Corona, R., Sera, F., Talamini, R., Barbato, F., Baroni, A., Cicale, L., Di Stefani, A., Farro, P., Rossiello, L., Ruocco, E., & Chimenti, S. (2004). Three-Point Checklist of Dermoscopy. *Dermatology*, *208*(1), 27–31. https://doi.org/10.1159/000075042

Stanley, R. J., Stoecker, W. V., & Moss, R. H. (2007, February). A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images. *Skin Research and Technology*, *13*(1), 62–72. https://doi.org/10.1111/j.1600-0846.2007.00192.x.

Stolz, W. (1995, October). The ABCD rule of dermatoscopy: high negative predictive value for the recognition of malignant melanomas. *Journal of the European Academy of Dermatology and Venereology*, *5*(1), S59. https://doi.org/10.1016/0926-9959(95)95977-9

Sumithra, R., Suhil, M., & Guru, D. (2015). Segmentation and Classification of Skin Lesions for Disease Diagnosis. *Procedia Computer Science*, *45*, 76–85. https://doi.org/10.1016/j.procs.2015.03.090

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015, June). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2015.7298594

Talavera-Martínez, L., Bibiloni, P., Giacaman, A., Taberner, R., Hernando, L. J. D. P., & González-Hidalgo, M. (2022, June). A novel approach for skin lesion symmetry classification with a deep learning model. *Computers in Biology and Medicine*, *145*, 105450. https://doi.org/10.1016/j.compbiomed.2022.105450.

Tarver, T. (2012, July). Cancer Facts & Figures 2012. American Cancer Society (ACS). *Journal of Consumer Health on the Internet*, *16*(3), 366–367. https://doi.org/10.1080/15398285.2012.701177

Thanh, D. N. H., Prasath, V. B. S., Hieu, L. M., & Hien, N. N. (2019, December 17). Melanoma Skin Cancer Detection Method Based on Adaptive Principal Curvature, Colour Normalisation and Feature Extraction with the ABCD Rule.

Journal of Digital Imaging, *33*(3), 574–585. https://doi.org/10.1007/s10278-019-00316-x

Tschandl, P., Rosendahl, C., & Kittler, H. (2018, August 14). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, *5*(1). https://doi.org/10.1038/sdata.2018.161

Valueva, M., Nagornov, N., Lyakhov, P., Valuev, G., & Chervyakov, N. (2020, November). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathematics and Computers in Simulation*, *177*, 232–243. https://doi.org/10.1016/j.matcom.2020.04.031

Vance, K. W., & Goding, C. R. (2004, August). The Transcription Network Regulating Melanocyte Development and Melanoma. *Pigment Cell Research*, *17*(4), 318–325. https://doi.org/10.1111/j.1600-0749.2004.00164.x

WANG, H. L. (2008, May 20). Approximation of Kullback-Leibler Divergence between Two Gaussian Mixture Distributions. *Acta Automatica Sinica*, *34*(5), 529–534. https://doi.org/10.3724/sp.j.1004.2008.00529

Wang, L., Zhang, L., Shu, X., & Yi, Z. (2023, April). Intra-class consistency and inter-class discrimination feature learning for automatic skin lesion classification. *Medical Image Analysis*, *85*, 102746. https://doi.org/10.1016/j.media.2023.102746

Wen, C., Wang, T., & Hu, J. (2007, January). Relative principle component and relative principle component analysis algorithm. *Journal of Electronics (China)*, *24*(1), 108–111. https://doi.org/10.1007/s11767-006-0097-2

Wolf, I. H., Smolle, J., Soyer, H. P., & Kerl, H. (1998, October). Sensitivity in the clinical diagnosis of malignant melanoma. *Melanoma Research*, *8*(5), 425–429. https://doi.org/10.1097/00008390-199810000-00007

Xiao, F., & Wu, Q. (2020, June 22). Visual saliency based global–local feature representation for skin cancer classification. *IET Image Processing*, *14*(10), 2140–2148. https://doi.org/10.1049/iet-ipr.2019.1018

Younis, H., Bhatti, M. H., & Azeem, M. (2019, December). Classification of Skin Cancer Dermoscopy Images using Transfer Learning. *2019 15th International Conference on Emerging Technologies (ICET)*. https://doi.org/10.1109/icet48972.2019.8994508

Yu, J., Qin, Z., Wan, T., & Zhang, X. (2013, November). Feature integration analysis of bag-of-features model for image retrieval. *Neurocomputing*, *120*, 355–364. https://doi.org/10.1016/j.neucom.2012.08.061

Yuan, Y., Chao, M., & Lo, Y. C. (2017, September). Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance. *IEEE Transactions on Medical Imaging*, *36*(9), 1876–1886. https://doi.org/10.1109/tmi.2017.2695227

Zhang, G., & Guo, S. (2021, April 1). Asymmetry analysis of melanoma based on ABCD rule. *Journal of Physics: Conference Series*, *1883*(1), 012070. https://doi.org/10.1088/1742-6596/1883/1/012070

Zhang, Y., Liu, S., Li, C., & Wang, J. (2021, January 26). Rethinking the Dice Loss for Deep Learning Lesion Segmentation in Medical Images. *Journal of Shanghai Jiaotong University (Science), 26*(1), 93–102. https://doi.org/10.1007/s12204-021-2264-x

Zhao, C., Shuai, R., Ma, L., Liu, W., Hu, D., & Wu, M. (2021). Dermoscopy Image Classification Based on StyleGAN and DenseNet201. *IEEE Access*, *9*, 8659–8679. https://doi.org/10.1109/access.2021.3049600