Democratic People's Republic of Algeria.
Ministry of Higher Education and Scientific Research
University of May 8, 1945 - Guelma -
Faculty of Mathematics, Computer Science, and Material Sciences
Department of Computer Science

MASTER'S THESIS

Field : Computer Science

Option: Information and Communication Science and Technology

Title :

MELANOMA DETECTION SYSTEM USING DEEP LEARNING TECHNIQUES

Presented By

Salmi Randa

Members of the Jury:

- **President:** Mme Madi Leila
- **Supervisor:** Dr. Bordjiba Yamina
- **Examiner:** Pr. Bencheriet Chemesse Ennehar

June 2024

Acknowledgments

Alhamdulillah for giving me the strength and patience to finish this work.

I would like to express my sincere gratitude to my supervisor, Dr. Borjiba Yamina, whose invaluable support, guidance, and encouragement were instrumental in completing this project, Her expertise and patience were truly appreciated.

I would also like to thank the jury members for their evaluation.

Thanks to my family, especially my mom for her encouragement and belief in me that have been a constant source of strength.

Thanks To my friends, especially my best friends Merabti Chayma(my true friend always), Sara, and Manar, thank you for encouraging me along the way.

And to all my dear friends who supported me with advice , encouragement, resources, or just by being there — thank you. Your support helped me achieve this work.

Dedications

Dedicate this work to all computer science students eager to develop expertise in the healthcare domain.

May it inspire you to use your knowledge and skills in computer science to make meaningful contributions to healthcare and improve, or even save, people's lives.

Abstract

Melanoma is a very aggressive form of skin cancer. Diagnosis by dermatologists is based on a thorough clinical examination, dermatoscopy and, if necessary, a biopsy with histopathological analysis. Early diagnosis is crucial for effective treatment and a favourable prognosis. A computer-assisted diagnostic procedure is more objective and reliable than expert human diagnosis, which is subjective and not necessarily reproducible. Machine learning and image processing techniques have been widely used for the automated diagnosis of cutaneous melanoma. More recently, deep learning techniques are being used more and more and are giving better results, especially with convolutional neural network models.

The rarity of melanoma cases and concerns about data privacy are major obstacles to creating balanced datasets for the automatic diagnosis of melanoma.

In this work, we propose a melanoma detection system based on deep learning techniques. In this system, we address the issue of dataset imbalance in skin data. To this end, we propose using a rebalancing technique by subdividing the database into balanced subsets to reduce bias towards the non-melanoma class. Then, we leverage transfer learning, using the ResNet50 model as the base model.

In the final phase of the proposed system, which involves classification and decision-making, we tested several machine learning techniques, namely SVM, XG-Boost, logistic regression, and voting techniques: the average voting method and the weighted voting method. Each model was trained and evaluated separately, achieving an accuracy of 81%. After combining the decisions, the best results were obtained through weighted voting, with 87.73%, 86.07%, 90%, and 88.01% for precision, accuracy, recall, and F1 score, respectively. Based on these results, the weighted voting technique balanced the predictions and significantly reduced biases. The system proved to be more reliable and accurate, capable of providing effective computer-aided diagnosis in medical practice.

Keywords: CNN, Melanoma, Deep learning , balancing technique, ResNet50.

Résumé

Le mélanome est une forme de cancer de la peau très agressive. Le diagnostic par les dermatologues repose sur un examen clinique approfondi, une dermatoscopie et, si nécessaire, une biopsie avec analyse histopathologique. Un diagnostic précoce est crucial pour un traitement efficace et un pronostic favorable.

Une procédure de diagnostic assistée par ordinateur est plus objective et fiable qu'un diagnostic humain d'expert, qui est subjectif et pas nécessairement reproductible. Les techniques d'apprentissage automatique et de traitement d'images ont été largement utilisées pour le diagnostic automatisé du mélanome cutané. Plus récemment, les techniques d'apprentissage profond sont de plus en plus utilisées et donnent de meilleurs résultats, notamment avec des modèles de réseaux de neurones convolutifs.

La rareté des cas de mélanome et les préoccupations relatives à la confidentialité des données constituent des obstacles majeurs à la création d'ensembles de données équilibrés pour le diagnostic automatique du mélanome.

Dans ce travail, nous proposons un système de détection de mélanome basé sur des techniques d'apprentissage profond. Dans ce système, nous visons la problématique du déséquilibre des ensembles de données sur la peau. À cette fin, nous proposons d'utiliser une technique de déséquilibrage en subdivisant la base de données en sous-ensembles équilibrés afin de réduire le biais vers la classe des non-mélanomes. Ensuite, nous exploitons l'apprentissage par transfert, où le modèle ResNet50 a été utilisé comme modèle de base.

Dans la dernière phase du système proposé, qui est la classification et la décision, nous avons testé plusieurs techniques d'apprentissage automatique, à savoir SVM, XG-Boost, la régression logistique et les techniques de vote : la méthode de vote moyen et la méthode de vote pondéré. Chaque modèle a été formé et évalué séparément, leur précision atteignant 81 %, et après fusion des décisions, les meilleurs résultats ont été obtenus par le vote pondéré avec 87,73 %, 86,07 %, 90 %, 88,01 % pour la précision, l'exactitude, le rappel et le score F1, respectivement. Au vu des résultats, la technique de dépondération a permis d'équilibrer les prédictions et de réduire considérablement les biais. Le système s'est avéré plus fiable et plus précis, capable de fournir un diagnostic assisté par ordinateur efficace dans la pratique médicale.

Mots-clés : CNN, Mélanome, Apprentissage profond, Technique d'équilibrage, ResNet50.

الملخص

الميلانوما هو نوع شديد العدوانية من سرطان الجلد. يعتمد تشخيص أطباء الجلد على فحص سريري شامل، والتنظير الجلدي، و إذا لزم الأمر أخذ خزعة مع التحليل النسيجي. التشخيص المبكر للميلانوما أمر بالغ الأهمية للعلاج الفعال.

تعتبر الإجراءات التشخيصية المدعومة من بالحاسوب أكثر موضوعية وموثوقية من التشخيص البشري الذي يمكن أن يكون ذاتيًّا أو يفتقر للخبرة اللازمة. تم استخدام تقنيات التعلم الآلي ومعالجة الصور على نطاق واسع في للتشخيص الآلي للميلانوما الجلدية في الآونة الأخيرة، حيث يتم استخدام تقنيات التعلم العميق بشكل متزايد و تقدم نتائج أفضل خاصة مع نماذج الشبكات العصبية الالتفافية.

تعد ندرة حالات الميلانوما و المخاوف المتعلقة بخصوصية البيانات عقبات رئيسية أمام إنشاء مجموعات بيانات متوازنة للتشخيص الآلي للميلانوما.

في هذا العمل، نقترح نظامًا للكشف عن الميلانوما يعتمد على تقنيات التعلم العميق. في هذا النظام، نتناول مسألة عدم توازن البيانات في قواعد بيانات سرطان الجلد. لتحقيق هذه الغاية، نقترح استخدام تقنية إعادة التوازن عن طريق تقسيم قاعدة البيانات إلى مجموعات فردية متوازنة لتقليل التحيز نحو فئة "الغير ميلانوما". بعد ذلك، نستفيد من التعلم الانتقالي، باستخدام نموذج ResNet50 كنموذج أساسي.

في المرحلة النهائية من النظام المقترح، والتي تتضمن التصنيف واتخاذ القرار، اختبرنا عدة تقنيات تعلم آلي و هي: SVM و XG-Boost والانحدار اللوجستي وتقنيات التصويت؛ طريقة التصويت المتوسط وطريقة التصويت الموزون. تم تدريب وتقييم كل نموذج بشكل منفصل، محققا دقة بنسبة 81%. بعد دمج القرارات، تم الحصول على أفضل النتائج من خلال التصويت الموزون، حيث بلغت 87.73% ، 86.07% ، 90% و 88.01%. للدقة، والدقة الشاملة، والاسترجاع، ومؤشر F1 على التوالي. بناءً على هذه النتائج، أثبتت تقنية التصويت الموزون فعاليتها في موازنة التنبؤات وتقليل التحيزات بشكل كبير. أثبت النظام أنه أكثر موثوقية ودقة، و قادر على توفير تشخيص مدعوم بالحاسوب بشكل فعال في الممارسة الطبية.

الكلمات المفتاحية: الشبكات العصبية الالتفافية، الميلانوما، التعلم العميق، تقنيات موازنة قواعد البيانات، ResNet50.

Contents

List of Tables

List of Figures

Introduction

Melanoma is a type of skin cancer that originates in melanocytes and develops to invade neighbouring cells and reach various other organs in the body via lymphatic channels. Melanoma accounts for 1% of all cases of skin cancer. Although melanoma is considered a rare tumour, it receives a great deal of attention because it is the most aggressive skin tumour. Diagnosed at its final stage, the survival rate falls to 32%, which is considered fatal for patients. Although doctors have abandoned naked-eye examination in favour of imaging techniques, the accuracy of melanoma diagnosis generally does not exceed 77%, and the rate varies according to the inexperience of the doctor.

Artificial intelligence is opening up new opportunities in the field of medical diagnosis, particularly for melanoma. Deep learning techniques, such as convolutional neural networks, are powerful tools for developing accurate automatic diagnostic systems. However, their performance is highly dependent on the quality and quantity of the data used to train them. In particular, data imbalance, i.e. the over-representation of non-melanoma tumour images compared with melanoma images, poses a major challenge to accurate model training. In fact, skin cancer databases generally contain a limited number of melanoma images, often less than 4,000, due to the rarity of this disease and restrictions linked to patient confidentiality. This imbalance can lead to biased models that favour the majority class (non-melanoma tumours) and neglect the minority class (melanoma), thus reducing their performance in detecting melanoma cases.

The aim of this study is to develop a melanoma detection system that is robust against unbalanced data. To do this, we propose an approach based on subdividing the database into balanced subsets, each of which is trained in a separate deep learning model. This strategy allows the models to focus on the specific features of the melanoma images and improves their ability to discriminate suspicious lesions. Finally, we use decision fusion techniques to combine the predictions of different models and arrive at a more accurate and reliable final decision.

This master's thesis is structured in four chapters, with a general introduction and a general conclusion. The chapters are organized consistently to explore deep melanoma detection using deep learning.

The first chapter provides an overview of skin cancer and melanoma, its types, location in the skin and risks. The different types of melanoma, their location in the skin and the risk factors associated with their development are explored. In addition, the traditional diagnostic methods and imaging techniques used to identify melanoma are reviewed. Finally, the available databases for evaluating melanoma detection algorithms are reviewed.

The second chapter is devoted to deep learning, by investigating the fundamental concepts of convolutional neural networks, their architecture, layers and hyperparameters. This is followed by a review of existing deep learning models and an overview of the literature on melanoma detection using these techniques.

The third chapter presents our melanoma detection system, describing in detail the proposed methodology for solving the database imbalance problem and providing a detailed explanation of each phase of the system.

The fourth chapter is focused on the implementation and evaluation of our melanoma detection system. It describes the database used, the database division process and the training of the deep learning models. The system's performance is then evaluated using relevant evaluation metrics, and the results obtained are discussed. Finally, the limitations of our approach are discussed and avenues for future research are proposed.

Chapter 1

Melanoma Skin Cancer

1.1 Introduction

Skin cancer is divided primarily into three types: Squamous Cell Carcinoma (SCC), Basal Cell Carcinoma (BCC), and melanoma. Melanoma is the most fatal, causing around 55,000 deaths a year. Early identification of this disease not only spares patients the unpleasant procedure of biopsy, but also increases the patient's chances of survival.

This chapter provides a broad overview of melanoma, with the first section covering skin cancer and its various types, followed by a focus on melanoma. In the second part of the chapter, we discuss the imaging techniques used to diagnosis tumors and explain the diagnostic algorithms generally used.

The chapter concludes with a study of skin cancer databases that are freely available for testing Learning Machine Models for Melanoma Classification.

1.2 Skin

Human skin as the largest organ in the human body is considered the primary defender and recipient of bruises for the body's organs against external threats. It acts as an insulating dinner from heat, sun, microbes, etc (Dwivedi et al., 2022).

The skin consists of layers on top of each other as illustrated in the Figure 1.1, including the two main layers: the epidermis and the dermis, these layers are defined as follow (Sheard, 2021) :

1.2.1 Epidermis

The outer layer of the skin, it is characterized by being a thin layer and is divided into three basic types of cells :

- **a) Squamous cells**

  The top layer of the epidermis is composed of tightly packed flat cells called squamous cells, making it the thickest layer, this property imparts impermeability to the skin.

- **b) Basal cells**

  They are block-like cells located beneath the squamous cells and form a basement membrane. They are responsible for producing new skin cells, then pushing the old ones towards the surface of the skin, where these old cells die and are sloughed off.

- **c) Melanocytes cells**

  They are called melanocytes or pigment cells because they are the cells responsible for skin coloration. They are located between the basal cells and their function is to increase the production of melanin when the skin is exposed to ultraviolet rays in order to protect it from burns.

1.2.2 Dermis

The dermis is a skin layer of fibrous tissue located beneath the epidermis and containing hair roots, blood vessels, nerves, and lymphatic vessels. All of these elements are stabilized by collagen and proteins that ensure skin elasticity (Sheard, 2021).

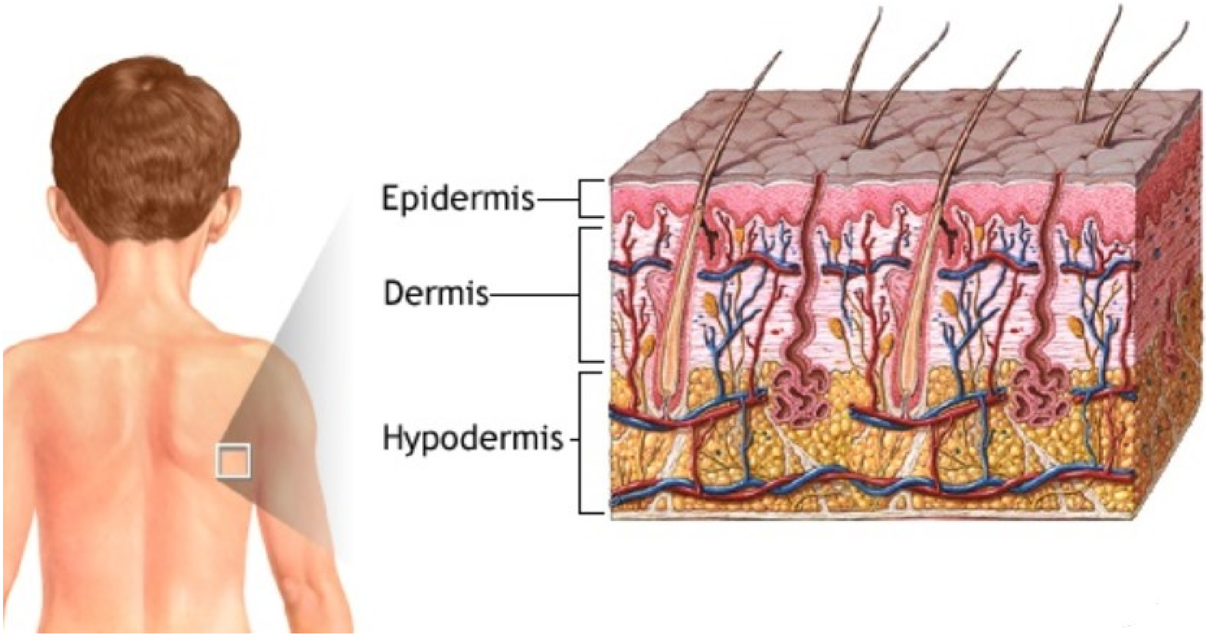


Figure 1.1: *The major layers of the skin (MedlinePlus, 2024).*

1.3 Cancer

The human body is composed of microscopic cells responsible for the functions of the body. During the life cycle of the cell, it grows and divides, then reaches a certain limit and dies according to a controlled mechanism. These cells contain genetic material responsible for growth and division, and when a defect occurs at their level (mutation), The cell loses control and continues to divide, forming an abnormal growth called a tumor (Dwivedi et al., 2022).

A tumor is considered benign, if it does not attack nearby tissues or other parts of the body,Its growth stops at a certain size and is considered fairly controlled and removed with surgery. But if the tumor invades and destroys neighboring cells, is labeled malignant or cancerous, which makes it life-threatening, unlike the first (Frederick O. Stephens, 2015), the difference is illustrated in the Figure 1.2.

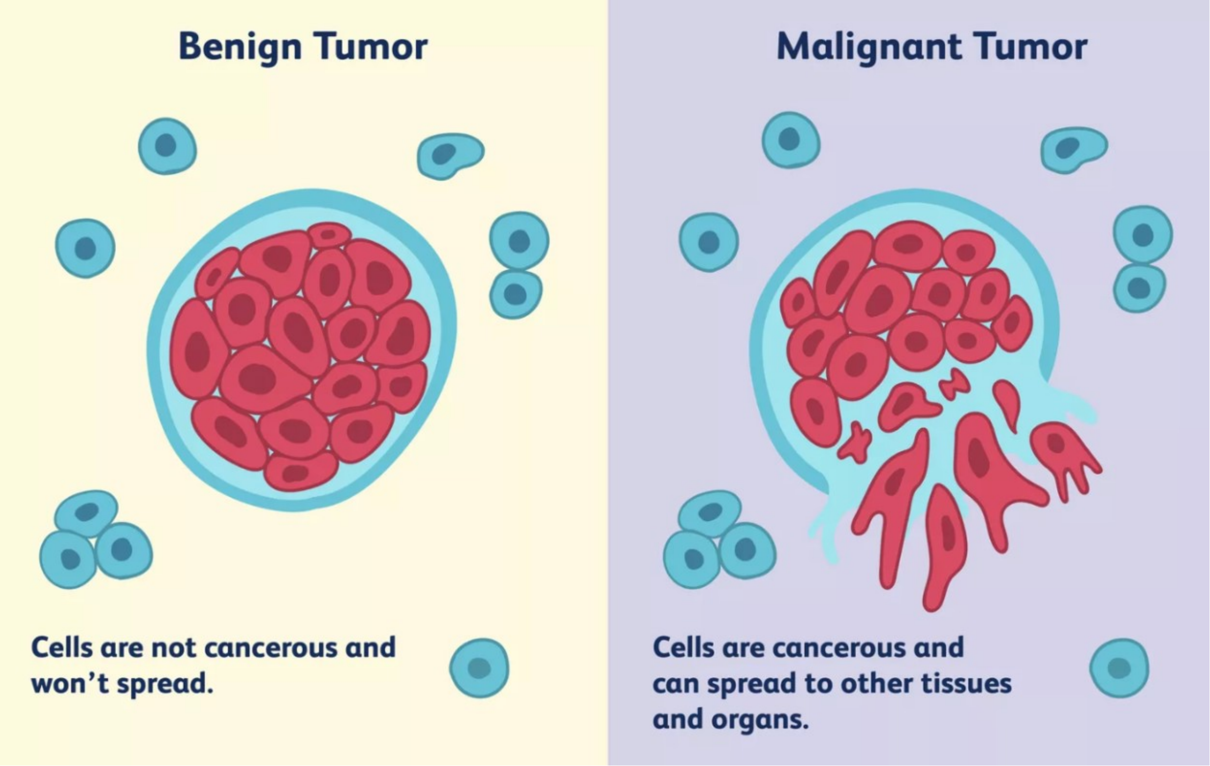


Figure 1.2: *Malignant tumor Vs Benign tumor (Verywell Health, 2023).*

1.4 Skin cancer

1.4.1 Definition

Cancer cells are characterized by their spread to adjacent tissues and organs and continue to spread uncontrollably (Frederick O. Stephens, 2015). Cancer is named according to its origin, and more precisely the place where it begins to appear, even if it spreads to the rest of the body's organs .

In this context, we can define skin cancer as follows: Skin cancer is a lesion or malignant tumor that begins in skin cells through uncontrolled growth (Sheard, 2021).

1.4.2 Types

Because the skin consists of several layers, skin cancer can form in the cells of one of these layers, and it is divided into two basic categories, illustrated in the Figures 1.3 and 1.4, and classified in the Table 1.1, The cells categories are defined as follows:

Tumor Types	**Tumor Classification**	**Sub-Types**
Non-Melanoma	Benign Tumors	Melanocytic nevus
		Solar lentigo (SL)
		Seborrheic keratosis (SK)
		Dermatofibroma (DF)
		Vascular lesion (VASC)
	Pre-cancerous	Actinic keratosis (AK)
	Malignant	Basal Cell Carcinoma (BCC)
		Squamous Cell Carcinoma (SCC)
Melanoma	Malignant	Superficial Spreading Melanoma
		Nodular Melanoma
		Lentigo Maligna Melanoma
		Acral Lentiginous Melanoma

Table 1.1: *Classification of Tumor Types.*

- **a) Basal Cell Carcinoma (BCC)** : The first common type of skin cancer, accounting for sixty-six percent of them. This type begins to appear in the basal cells of skin exposed to the sun for long periods of time, and the probability of contracting it increases with age, so its spread does not usually occur in younger age groups. From the age of forty.
  - **Appearance:** Often manifests on the skin as small gray sometimes pink spots or skin ulcers.
  - **Location:** Appears on sun-exposed skin, particularly on the face, followed by the neck, forearms, and back of the hands.
  - **Growth and Spread:** Characterized by slow growth, limited spread, and a rare possibility of spreading to the lymph nodes and nearby tissues.
  - **Risk Factors:** Due to its painless and slow-spreading nature, neglecting treatment poses a life-threatening risk and erosion of tissues adjacent to the tumor site.
- **b) Squamous Cell Carcinoma (SCC)**: This type constitutes approximately 33 percent of skin cancers, making it the second most prevalent type after basal cell carcinoma (BCC), It originates from squamous cells and commonly associated with skin that has been chronically exposed to prolonged sunlight. In its early stages, it is referred to as Bowen's disease or intra-epidermal cancer when located in the uppermost layer of the skin, SCC is also called invasive SCC if Invaded the membrane of basal cells.
  - **Appearance:** frequently manifests on the skin surface as a small, crusty lump or may present as a skin ulcer. The (SCC) typically exhibits a red coloration and is painful to touch.
  - **Location:** It is located in areas exposed to the sun, specifically the face, nose, chin, and lower lips, in addition to the back, neck, hands, and lower legs.
  - **Growth and Spread:** It is characterized by rapid growth and the ability to

spread to neighboring tissues and lymph nodes.

  - **Risk Factors:** Prolonged exposure to the sun and neglecting early examination of the tumor can result in its progression to bleeding, painful, and potentially life-threatening ulcers.

- **c) Melanoma** : These are cancerous tumors that begin in melanoma cells It constitutes the lowest percentage of skin cancers, with a percentage ranging between 1-2 percent.

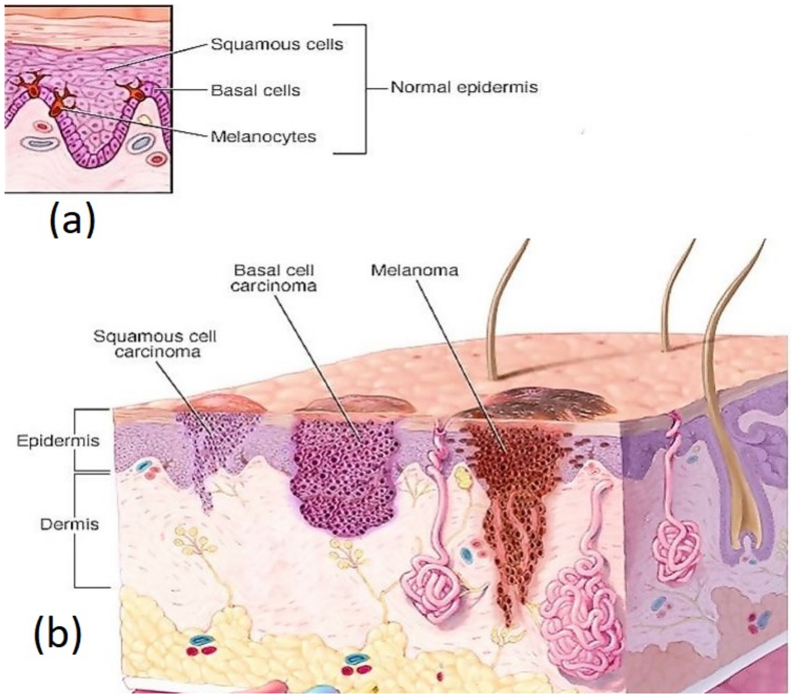


Figure 1.3: *The impact of skin cancer presence: (a) illustrates the normal epidermis; (b) depicts the three types of skin cancer (Moscow Clinical Scientific Center, 2024).*

Figure 1.4: *Images of actual skin cancer lesions illustrating the three primary types:Squamous Cell Carcinoma,Basal Cell Carcinoma, Melanoma (Science Photo Library, n.d.).*

1.4.3 Benign skin spots

- **a) Melanocytic nevus**

The melanocytic nevus, commonly referred to as a mole or naevus, is prevalent among individuals with diverse skin tones. It is defined as skin growths resulting from the growth of melanocytes, these growths may arise due to genetic factors or excessive exposure to sunlight during childhood, potentially leading to an increase in their presence (Sheard, 2021) :

- **Classification :** benign and harmless skin spot.
- **Appearance:** Its shape commonly appears oval or round, and it can be black, brown, pink, or match the skin color.
- **Location :**it can be found anywhere on the body.

- **b) Actinic keratosis**

  Commonly referred to as sunspots, actinic keratoses, or solar keratoses, these are considered precancerous lesions that appear in individuals of all ages but are more prevalent in those older than 40 due to prolonged sun exposure (Sheard, 2021):

  - **Classification :** Pre-cancerous.
  - **Appearance:** Appear in the form of spots of the same skin color or redder, with a rough texture, flat, and scaly.
  - **Location :** It commonly appears in areas of the body that are frequently exposed to the sun, such as the head, neck, hands, forearms, and legs.

- **c) Benign keratosis**

  1. **Solar lentigo**

     Solar lentigo (SL),This benign tumor is commonly known in books and articles under names such as age spots or senile lentigo, senile lentigines , or as mottled pigmentation. This type begins to appear with age because it is closely linked to skin aging ,due to genetic factors such as heritability and chronic exposure to ultraviolet radiation (UVR) (Hasegawa et al., 2015)(J. C. Kim et al., 2022):

     - **Classification :** Benign tumor (Hodgson, 1963).
     - **Appearance:** It appears in the form of spots darker than the skin color, which can be brown or dark brown (Hodgson, 1963).
     - **Location :** Any area of the human body that is exposed to a lot of sun is vulnerable to this type, especially the face, shoulders, and upper back (Hasegawa et al., 2015).

  2. **Seborrheic keratosis**

     Seborrheic keratosis (SK), also commonly known in books and resources under names such as verruca seborrhoeica, seborrheic or senile wart, and verruca senilis. It is more common in age groups over the age of 30, due to genetic factors.

     - **Classification :** Noncancerous or, in other words, benign.
     - **Appearance:** They take the form of multiple well-circumscribed lesions and have a relatively greasy, warty, or even velvety texture, with their color ranging from yellow to dark brown.

    - **Location :** The possibility of its appearance includes all areas of the body, especially the most common areas such as the face and upper torso, with the exception of the palms and soles.

- **d) Dermatofibroma**

  Dermatofibroma (DF), It usually falls under the following keywords: cutaneous fibrous histiocytomas (Chen et al., 2000; Luzar & Calonje, 2010),It commonly appears in the age group of twenty to fifty years old (J. Higgins et al., 2015).

  According to (Chen et al., 2000; J. Higgins et al., 2015; Luzar & Calonje, 2010) his location,classification and appearance are defined as follow :

  - **Classification :** benign tumor.
  - **Appearance:** The tumor is known as papules or nodules of a dark brown or reddish color that are darker in the center than at the edges, They are also characterized by being firm and raised.
  - **Location :** The possibility of its appearance includes any cutaneous in the human body, but the most common site is the lower extremities.

- **e) Vascular lesion**

  Vascular lesion (Vasc), are heterogeneous illnesses that affect the skin's blood vessels and lymphatic systems. They are distinguished by dilatation, proliferation, abnormal development, or growth of blood vessels.It commonly appears in the age group of 6 to 95 years old (Choi & Oakley, 2021).

  - **Classification :** Benign tumor (Choi & Oakley, 2021; Patel et al., 2012).
  - **Appearance:** Depends on the type (Patel et al., 2012).
  - **Location :** The location of its appearance varies depending on the type of tumor, as indicated in this study by Patel et al. (Patel et al., 2012), however it is typically found in the neck, head, extremities, or trunk (Choi & Oakley, 2021).

1.5 Melanoma

1.5.1 Definition

Melanoma is a form of cancer that originates in the cells that produce melanin, which is the pigment responsible for skin color. Despite not being the most common form, melanoma is deadly, causing 80% of skin cancer deaths (Penman et al., 2022).

1.5.2 Melanoma's types

Melanoma's types are divided into tumors with metastatic potential and those without. The main types with metastatic potential, which is the type of cancer that has the ability to spread or metastasize to other parts of the body, are presented in the Figure 1.5 and defined as follow:

Nodular Melanoma

Superficial Spreading

Acral Lentiginous Melanoma

Desmoplastic Melanoma

Lentigo Maligna Melanoma

Figure 1.5: *Melanoma's types (Science Photo Library, n.d.).*

- **a) Superficial spreading melanoma**

  This subtype accounts for 60% of melanomas, rendering it the most prevalent type. It typically manifests in individuals under the age of 40 , presenting as brown or black spots that undergo sudden growth on the skin or arise from alterations in the color, shape, or size of freckles or moles (Mothoneos, 2023).

  It is characterized by the following :

  - Appears between the shoulders and hip (Mothoneos, 2023).
  - Radial growth - grows horizontally across epidermal and superficial epidermal planes (Penman et al., 2022).
  - Does not reach lymphatics and blood vessels (Penman et al., 2022).
  - Exhibits a slow rate of growth (Mothoneos, 2023).

- **b) Nodular melanoma**

  This subtype constitutes 10-15 percent of melanomas, with a higher prevalence observed in individuals aged 65 and older (Mothoneos, 2023), Additionally, it is the most frequent subtype in prepubertal cases, occurring in 12 to 34% of cases (Riker, 2018). It manifests in sun-damaged skin as a nodule characterized by a raised, round shape, firm texture, and distinct color that can be black, brown ,or even red (Mothoneos, 2023).

  It is characterized by the following :

  - Appears in sun-damaged skin (Mothoneos, 2023).
  - Enters the vertical fast-growth phase early in lesion development (Elder et al., 1980).
  - Spreads rapidly into the lower layer of the skin (Mothoneos, 2023).
  - Commonly associated with ulceration and increased thickness (Riker, 2018).

- **c) Lentigo maligna melanoma**

  It constitutes 5 to 15 percent of melanomas, ranking as the third most prevalent sub type following superficial spread and nodular melanoma (H. 2. Higgins et al., 2015). It manifests as a an asymmetric, brown-black (H. 2. Higgins et al., 2015) sizable pigmented spot in individuals aged over 40, particularly among older individuals, and is attributed to chronic sun damage (Mothoneos, 2023).

  It is characterized by the following (H. 2. Higgins et al., 2015)(Mothoneos, 2023):

  - Appears in the facial area, specifically on the neck, ears, nose, or cheeks.
  - Manifests as a slow-growing lesion capable of penetrating deep into the skin over the course of years.
  - The substantial size poses challenges for effective treatment .

- **d) Acral lentiginous melanoma**

  This subtype is relatively less prevalent, comprising only 1 to 2 percent of melanomas. It manifests as dull or light-colored spots bearing a striking resemblance to bruises, unusual wart. the acral lentiginous melanoma (ALM) occurrence is most frequently observed in individuals aged over 40 (Mothoneos, 2023).

  It is characterized by the following :

  - Characterized by slow growth and the capacity to extend deep into the skin.
  - ALM development may be contributed to by mechanical stress or trauma, along with specific genetic factors.
  - Predominantly located on acral sites, notably the palms, soles, or under fingernails/toenails.

1.5.3 Diagnostic methods

In diagnosing melanoma, the examination of tumor shape is pivotal, where several diagnostic methods rely on discernible symptoms of the tumor, including the following:

- **a) Pattern analysis**

  All of the criteria for assessing patterns that separate benign tumors from melanomas have been sorted. More information is provided in the publication, and the patterns are organized in a table per tumor type. The criteria considered are stated below (Pehamberger et al., 1987):

  1. The margin of pigmented lesions of the skin (regular, uneven, or with unique pigment patterns).
  2. The study defines the pigmentation pattern (color, intensity, and depigmentation) as well as specific pigment patterns (pigment network, brown globules, and black dots).
  3. The overall appearance of pigmented skin lesions (uniform or diverse).
  4. The outline of the pigmented spots inside the surrounding skin.

- **b) ABCD rule**:

  The ABCD rule is an abbreviation of the four significant criteria: asymmetry (A), borders (B), colors (C), and different structural components (D) (Nachbar et al., 1994) as shown in the Figure 1.6. It follows the pattern analysis and focuses simply on the four mentioned criteria to classify tumors as malignant or benign (Johr, 2002).

  The TDS, or total dermoscopy score, is the sum of the scores assigned to each criterion. A score below 4.75 is classified as benign, 4.80-5.45 as malignant, and scores in between are considered precancerous (Johr, 2002; Nachbar et al., 1994).

  This method sometimes mistakenly classifies certain lesions, such as melanocytic nevi, dermatofibroma, ink-spot lentigo, actinic lentigo, etc (Johr, 2002).

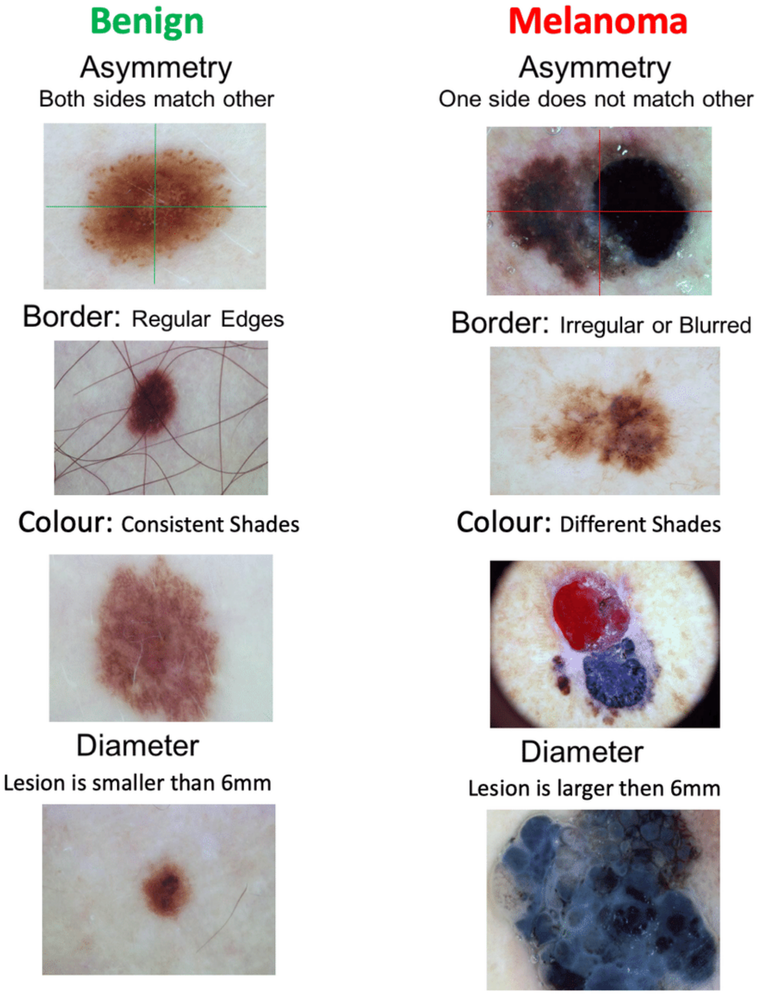


Figure 1.6: *The ABCD Rules method for melanoma diagnosis(Goyal et al., 2019).*

- **c) 7-point checklist**

  It's a point system-based method (Argenziano et al., 1998, 2000). It is divided into major criteria (with 2 points for each) and minor criteria (with 1 point for each) as shown in the table 1.2. It is considered less complicated compared to the ABCD rule and pattern analysis. If the final score is equal to or greater than 3, it is diagnosed as melanoma with a sensitivity of 95% (Johr, 2002).

	Clinical features	**Score**
Major features	• Change in size • Irregular shape • Irregular colour	2 for each
Minor features	• Largest diameter 7 mm or more • Inflammation • Oozing • Change in sensation	1 for each

Table 1.2: *The 7-point checklist table where A score of 3 or more is regarded as suspicious (Hall, 2013)*

1.5.4 Imaging techniques

An accurate diagnosis is essential for efficient healthcare. It begins with a complete medical history and examination. If necessary, the patient is referred for medical imaging (e.g., X-rays, CT (computed tomography), and MRI (magnetic resonance imaging)) (UK Health Security Agency, 2022).

Dermatological doctors commonly confront circumstances wherein the potential for multiple diagnoses complexifies the diagnostic process and necessitates undertaking confirmatory investigations. This multiple diagnoses problem, appears because of visually based diagnosis (Sgouros et al., 2018).

To enhance visual diagnosis, the field has shifted from clinicopathologic diagnosis to an era of clinicoimaging diagnosis, throught out integrating imaging techniques with clinical evaluations (Robinson & Callen, 2005).

- **a) Macroscopy**

  Clinical or macroscopic images are dermatological images that represent what can be seen with the naked eye by dermatologists, whether these images are digital or non-digital, and demonstrate one or more skin lesions on the skin's surface, providing a brief description of the lesion's location on the human body (Jafari et al., 2016; Pathan et al., 2018).

  The capturing of these clinical images may be done under various scales and distances from the lesions, rendering the lesion analysis more difficult (Pathan et al., 2018).

- **b) Dermoscopy**

  Dermoscopy, also referred to in medical literature as epiluminescence microscopy, derma(to)scopy, and trichoscopy, It is a technique that allows dermatologists to better visualize and analyze the morphological features of skin lesions using skin examination with a smaller dermatoscope devices (Figure 1.7) that can be either simple handheld models or more advanced versions with integrated image capturing

Simpler Models	**Advanced Models**	**Contact Based Models**	**Non-Contact Based Models**
- lack a camera. -cannot connect to a PC.	- have high-resolution cameras. -can connect to a PC	-better illumination -better resolution	Avoid cross-infection between patients

Table 1.3: *Comparison of Dermoscopy Models.*

and analytical capabilities (Kaliyadan, 2016). The dermoscopy available models are explained in the Table 1.3.

When examining skin lesions using a dermatoscope, specific visual patterns and features are observed and extracted. These patterns are associated with different conditions, such as melanoma and non-melanoma skin cancers (Carli et al., 2000).

Each type of skin tumor has characteristic dermoscopic patterns that have been extensively studied and documented (Carli et al., 2000).

Figure 1.7: *A dermatologist, Dr. Jason Lee, examines a skin lesion on a patient's wrist using a dermatoscope (Abdelmalek, 2018).*

1.6 Datasets overview

The development of computer-aided diagnostic methods for classifying dermoscopic images has been prompted by the rising rate of melanoma cases (Mendonca et al., 2013). Recent research indicates that convolutional neural networks can classify melanoma images as accurately as dermatologists (Brinker et al., 2019). To accurately evaluate these deep learning models, public datasets may be utilized as benchmarks for assessing the efficiency of algorithms and overcoming barriers to collecting large-scale image datasets, such as technological capabilities, legal approvals, time, and financial constraints (S. M. Khan et al., 2021).

Melanoma is a dangerous, uncommon, and fatal form of tumors on the skin. Based on information from The (American Cancer Society, n.d.) melanoma skin cancer patients represent just 1% of total instances, however they have a higher fatality rate. Patients have several non-melanocytic lesions, resulting in a restricted number of diagnoses in the available datasets (Tschandl et al., 2018a). Where the previous study on automated cancer of the skin detection has concentrated on melanocytic lesions the lack of an accessible human benchmark limits the ability to compare these algorithms' performance and restricts innovation in technology in melanoma detection field (Brinker et al., 2019). In recent years, numerous public benchmarks have emerged:

1.6.1 ISIC Archive

ISIC, or the International Skin Imaging Collaboration, is a global repository that attempts to strengthen melanoma and skin disease diagnosis via the use of an extensive set of clinical images (Wen et al., 2022).

The ISIC archive hosts several skin image analysis competitions which aid with lesion segmentation and classification tasks, including :

1. **ISIC2016** (Gutman et al., 2016).
2. **ISIC2017** (N. Codella et al., 2017).
3. **ISIC2018** (N. C. Codella et al., 2018; Tschandl et al., 2018b).
4. **ISIC2019** (N. C. Codella et al., 2018; Combalia et al., 2019; Tschandl et al., 2018b).
5. **ISIC2020** (ISIC, 2020).

The competition included different tasks such as image segmentation and classifying the tumors. In order to unify the evaluation of the competitors' performance, a training database and a test database were provided in each announced competition. In this table 1.4, the announced competitions and the division of the images into the training and test databases are provided, as well as the number of classes to which the tumors belonged.

DataSets	Reference	Training Dataset	Testing Dataset	Class
ISIC-2016	(Gutman et al., 2016)	900	379	2
ISIC-2017	(N. Codella et al., 2017)	2,000	600	3
ISIC-2018	(N. C. Codella et al., 2018; Tschandl et al., 2018b)	10,015	1,512	7
ISIC-2019	(N. C. Codella et al., 2018; Combalia et al., 2019; Tschandl et al., 2018b)	25,331	8,238	8
ISIC-2020	(ISIC, 2020)	33,126	10,982	2

Table 1.4: *The number of images contained in the ISIC Archive.*

The distribution of images across classes is shown in the following table 1.5. It can be noted that there are few images in the melanoma class, while they are more numerous in the benign tumor class. This is what makes most systems that work with binary classification suffer from bias. which makes the ability to separate melanoma from non-melanoma a major challenge.

Class disease	**2016**	**2017**	**2018**	**2019**	**2020**
melanoma	173	374	1,113	4,522	584
melanocytic nevus	0	0	6,705	12,875	0
basal cell carcinoma	0	0	514	3,323	0
actinic keratosis	0	0	327	867	0
benign keratosis	0	0	1,099	2,624	0
dermatofibroma	0	0	115	239	0
vascular lesion	0	0	142	253	0
sqs cell carcinoma	0	0	0	628	0
benign	727	1,372	0	0	0
seborrheic keratosis	0	254	0	0	135
atypical melanocytic	0	0	0	0	1
cafe-au-lait macule	0	0	0	0	1
lentigo nos	0	0	0	0	44
lichenoid keratosis	0	0	0	0	37
nevus	0	0	0	0	5,193
solar lentigo	0	0	0	0	7
other/unknown	0	0	0	0	27,124
Total	**900**	**2,000**	**10,015**	**25,331**	**33,126**

Table 1.5: *ISIC Archive class distribution in the training datasets (Nugroho et al., 2024)*

The competitions collect data from a variety of sources (for example, ISIC2018 includes the HAM10000 dataset(Tschandl et al., 2018b)), and the benchmarks are used to train and evaluate dermatology-focused machine learning algorithms (Wen et al., 2022), we collected some of the used benchmarks information in the Table 1.6 .

Datasets		**Images informations**			
Name	**Ref**	**Format**	**Modality**	**N°.classes**	**N°.Images**
HAM 10,000	(Tschandl et al., 2018b)	JPG	Dermoscopic	8	10015
BCN 20,000	(Combalia et al., 2019)	JPG	Macroscopic	9	12413
MSK 1–5	(Yang et al., 2020)	JPG	Macroscopic	15	3918
2018 JID editorial images	(Kawahara et al., 2019)	JPG	Dermoscopic, Macroscopic	3	100
UDA 1–2	(De Faria et al., 2019)	JPG	Dermoscopic	7	617

Table 1.6: *skin cancer datasets that are included within the ISIC Archive.*

1.6.2 HAM10000

One of the most used databases since its launch in 2018, HAM10000, an abbreviation of Human Against Machine with 10,000 training images, is freely accessible via the ISIC repository. It contains distinct dermatoscopic images from various populations and modalities. It includes 8 classes and the distribution of the images in them is shown in the table 1.7 . Over half of the lesions were pathologically confirmed, with the remainder requiring follow-up, expert consensus, or in-vivo confocal imaging (Tschandl et al., 2018b).

Class	**Training Set**	**Testing Set**	**Validation Set**	**Total**
Actinic Keratoses (Solar Keratoses) (akeic)	222	92	13	327
Basal cell carcinoma (BCC)	339	150	25	514
Benign keratosis (BKL)	770	294	49	1113
Dermatofibroma (DF)	68	40	7	115
Melanoma (MEL)	4340	2027	338	6705
Melanocytic nevi (NV)	680	353	66	1099
Vascular skin lesions (VASC)	90	49	3	142
Total	**6509**	**3005**	**501**	**10,015**

Table 1.7: *HAM10000 Dataset Distribution*

1.6.3 PH²

The PH² dataset(Mendonça et al., 2013),comprises 200 dermoscopic images with a resolution of 768x560 pixels the images have been manually segmented, clinically diagnosed, and identified by experienced dermatologists and classed into 3 classes. It was created by the Dermatology Service of Hospital Pedro Hispano in collaboration with the University of Porto and the Institute of Electronics and Informatics Engineering of Aveiro Primarily used by researchers and practitioners in dermatology and medical imaging, it aids in the analysis and classification of pigmented skin lesions. And it is accessible at no cost for study and benchmarking.

1.6.4 Other Datasets

ISIC was not the only source of skin cancer images, but there are many open source databases that have also been used in evaluating skin cancer detection algorithms. These databases vary in their popularity and use according to the type of images and classes targeted in the study. Some of the most famous databases are summarized in the following table 1.8.

Datasets		**Images informations**			
Name	**Ref**	**Format**	**Modality**	**N°.classes**	**N°.Images**
MED-NODE	(Giotis et al., 2015)	JPG	Macroscopic	2	170
SD-260	(Yang et al., 2020)	JPG	Macroscopic	260	20 600
7-point criteria evaluation	(Kawahara et al., 2019)	JPG	Dermoscopic, Macroscopic	15	1011
SKINL2	(De Faria et al., 2019)	PNG	Light field photographs, dermoscopic photographs	8	814

Table 1.8: *skin cancer datasets that are not in the ISIC Archive.*

1.7 Conclusion

Melanoma is a fatal tumor that is difficult to diagnose since it requires years of experience. since of the recent increase in dissemination, early detection is required to save the lives of patients in the early stages of the disease. Since the diagnosis of the disease depends on the naked eye or certain imaging modalities, as well as the use of diagnostic algorithms, the use of artificial intelligence is becoming increasingly common.

This chapter provided a general overview of skin cancer and its different types. We then moved on to imaging techniques, and concluded the chapter with a brief review of accessible databases.

The next chapter examines the importance of using artificial intelligence as an early-stage diagnostic option for melanoma, as well as a comprehensive overview of published studies on the subject.

Chapter 2

Deep Learning Overview

2.1 Introduction

Traditional diagnostic methods for deadly melanoma rely heavily on visual assessment of the lesion by dermatologists, making the process subjective, error-prone, and requiring years of experience. To improve accuracy and efficiency, studies have increasingly turned to automated detection systems, focusing on deep learning techniques as an assistant to the dermatologist in decision-making.

Deep learning, is a part of machine learning as shown in the Figure 2.1, that provides an automatic features extraction (Puttagunta & Ravi, 2021), It has revolutionized many fields by enabling machines to learn and make decisions from large datasets. Among the various deep learning models, convolutional neural networks (CNNs) have proved the advantages of deep learning approaches for skin cancer detection, as evidenced by its effectiveness and efficacy in the early and effective detection of skin malignancies (Shah et al., 2023).

In this chapter, we provide an in-depth overview of deep learning. We begin the chapter by defining neural networks and their types. We then delve into the details of CNNs, discussing their layers and how they work. We then move on to transfer learning and the use of pre-trained models, and conclude the chapter with published work related to melanoma detection.

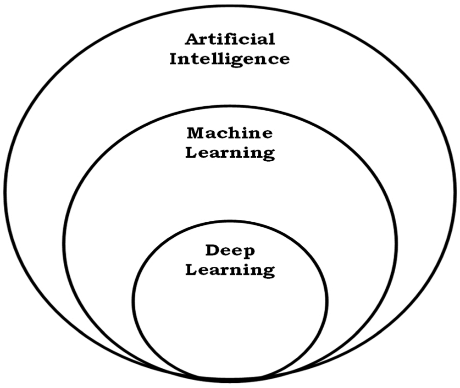


Figure 2.1: *Relation Between AI,ML and Deep Learning (Ahmed & Islam, 2020).*

2.2 Neural Network

A neural network is a category of machine learning algorithms modeled loosely on the human brain, consisting of interconnected nodes or "neurons" that process data in a manner inspired by biological neural networks. These algorithms are designed to recognize patterns, learn from data, and make decisions (Joshi et al., 2023).

The term (neural) refers to the (neuron) or the cells of nerves, this neuron is present in the brain and other parts within the human body, which constitute the basic functional component of the nervous system (Gu et al., 2015). The human brain is made up of around 86 billion neuron cells connected by axons (Schalkoff, 1997). The nerve cell communication system is distinguished by various inputs from sensory organs, which are received then accepted by dendrites and translated into electrical impulses that express the nerve message, as shown in the Figure 2.2.

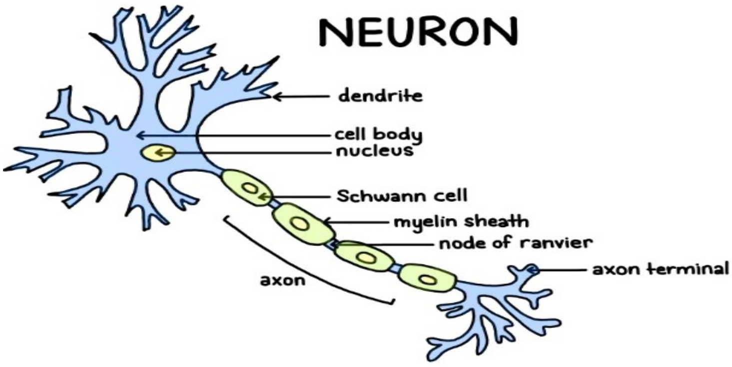


Figure 2.2: *The human Neuron (Guy-Evans, 2024).*

A neural network algorithms (Joshi et al., 2023) defined as a computational models comprises interconnected units known as neurons (building block), arranged in layers: input, hidden, and output layers, as shown in the Figure 2.3 illustrate. As information passes through these layers, each neuron processes the inputs by performing mathematical operations and generating an output. During the training process, the network adjusts the weights of the connections between neurons to improve accuracy. This involves using mathematical techniques to minimize the error between the predicted and actual outputs, its layers are defined as(Joshi et al., 2023) :

- **The Input Layer:** gets the initial input or characteristics that the neural network will interpret. Each of the neurons in the input layer indicates a particular component of the data being input.
- **The Hidden Layers:** Every neuron in a hidden layer receives input from neurons in the layer preceding it, applies an activation function, and then sends the result to neurons in the following layer.
- **The Output Layer:** The number of neurons in the output layer varies according to the sort of issue the neural network is tackling.

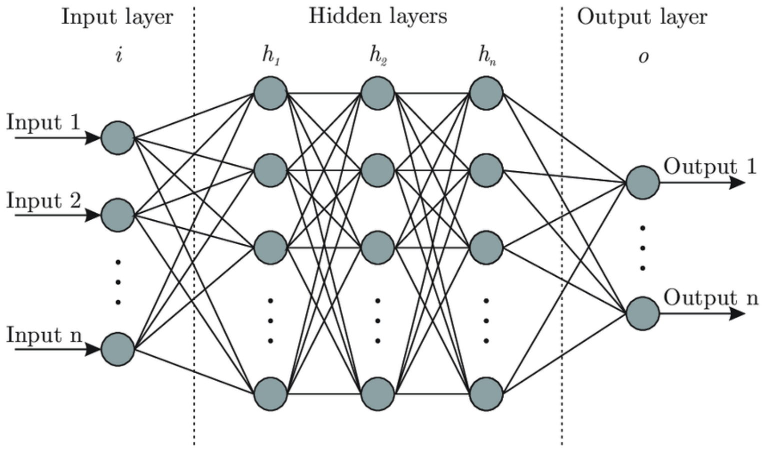


Figure 2.3: *Artificial neural network architecture (Bre et al., 2017).*

2.2.1 Neural Networks Types

Neural networks include many types that differ according to the methods of passing data through their neurons. In the following Table 2.1, we have summarized the most famous types of neural networks:

- **DNN:** Deep neural networks.
- **CNN:** Convolutional neural networks.
- **RNN:** Recurrent neural networks.
- **MLP:** Multi layer Perceptron.

Name	Description	Application Domain
MLP	A class of feedforward neural networks consisting of multiple layers of neurons, where each neuron is connected to every neuron in the previous and next layers.	Basic classification, regression analysis, simple pattern recognition (Popescu et al., 2009).
DNN	A type of artificial neural network (ANN) with multiple layers between the input and output layers. These networks can model complex non-linear relationships.	Natural language processing, speech recognition, image recognition,classification tasks (Awad & Khanna, 2015)
CNN	designed to process and analyze visual data. They use convolutional layers to automatically and adaptively learn spatial hierarchies of features.	spam detection, image classification, object detection, segmentation, face recognition (Arif et al., 2018).
RNN	A type of neural network designed for processing sequential data. RNNs have connections that form directed cycles, allowing information to persist. This makes them ideal for time series and sequence data.	Sequence recognition, NLP,audio modeling, speech recognition (Bianchi et al., 2017).

Table 2.1: *Main types of neural networks and their application domain.*

2.3 Convolutional Neural Network (CNN)

2.3.1 Definition

The CNN defined as an end-to-end deep learning algorithm with an architecture combines both feature extraction and classification, allowing the model to learn key features straight from data, making it ideal for difficult visual recognition(Alzubaidi et al., 2021).

The model has three layers that significantly impact its performance, in addition to the layers that handle input and output. These layers are as follows : Convolutional layers with Kernels (filters), Pooling layers, Fully connected layers. as shown in the Figure 2.4.

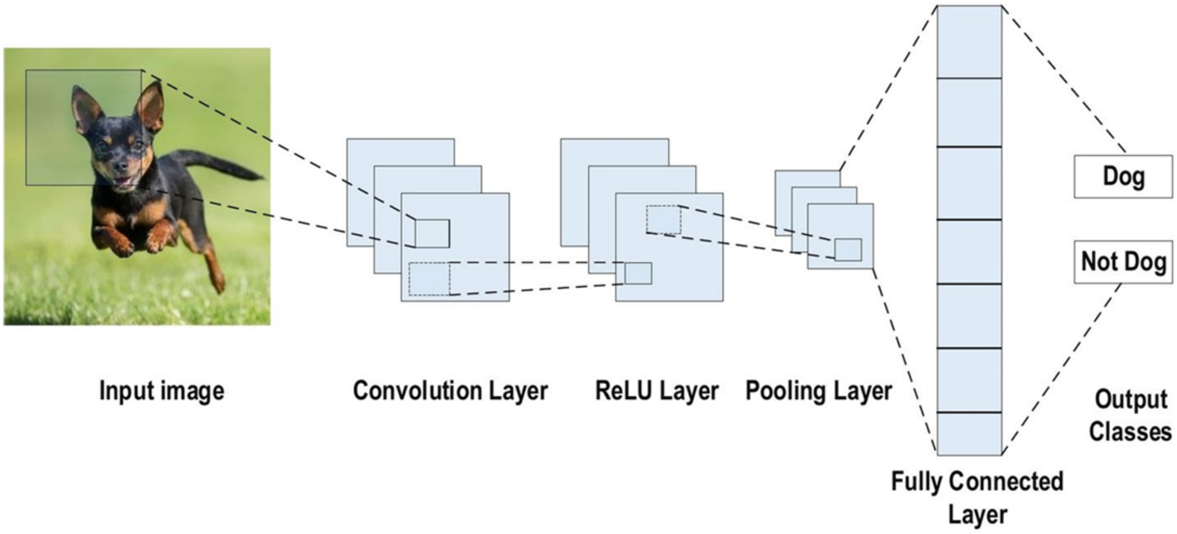


Figure 2.4: Diagram of a Convolutional Neural Network (CNN) for image classification (Alzubaidi et al., 2021).

2.3.2 CNN Layers

CNN generally consists of three types of building blocks: convolution, pooling, and fully connected layers (dense layers). The convolution and pooling layers are responsible for feature extraction, in contrast to the dense layers are in charge of classification. the layers are defined as following (Yamashita et al., 2018):

Convolutional layers

It is the initial phase to extract features from an input image using a mathematical operation between the pixels of the image matrix and a kernel, to determine the spatial relationships between pixels.

A convolved image is created by convolving a filter from top to bottom and from left to right, and then repeating the process to generate many output feature maps (Dumoulin & Visin, 2016). The process is shown in the Figure 2.5.

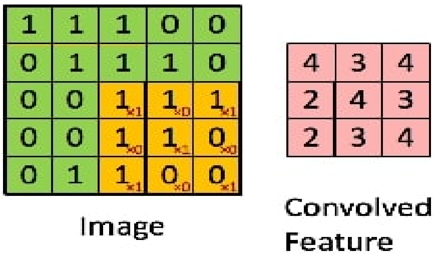


Figure 2.5: *Convolution operation: A kernel slides over the image, producing a feature map (Prabhu, 2018).*

Pooling Layer

The pooling layer downsamples the feature maps produced by the convolutional layers with the goal to reduce the total number of parameters that need to be processed in the layers that follow (see the example illustrated in Figure 2.6) . This helps to lower the total complexity of the network (A. Khan et al., 2020).

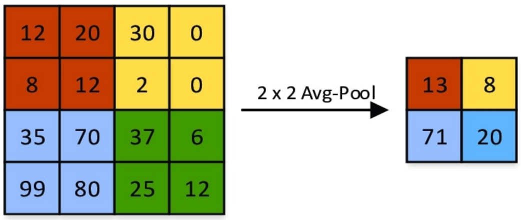


Figure 2.6: *Average pooling with a pool size of 2x2 (Prabhu, 2018).*

The most common types of pooling layer can be summarised as shown in Table 2.2:

Pooling type	**definition**	**Utility**	**weakness**
Average (AVG)	Extract the Average value in a region	reducing noise and over-fitting in the data	May overlook important features.
Max	Extract the Max value in a region	detecting features inthe input data	Reduces spatial information.

Table 2.2: *Pooling Types.*

Fully Connected Layers

In a fully connected (FC) layer, every neuron is connected to every other neuron. the input of this layer is the output from the feature extraction phase that is turned into a one-dimensional vector by the flatten layer. Then, this layer will combine all the features extracted in earlier layers in order to make final decisions.

2.3.3 CNN Hyper-parameters

Convolutional Neural Network (CNN) hyper-parameters, including learning rate, batch size, and filter size,have a significant impact on model performance and training efficiency. Meticulous tuning of these parameters promotes optimal convergence and mitigates over-fitting.

Training Hyper-parameters

Also known as performance hyper-parameters, are parameters that must be set to improve the performance and convergence of machine learning models, such as CNN architectures. These hyper-parameters are determined for the training process and include batch size, learning rate,loss function, and others.

The best settings (Kandel & Castelli, 2020) for these hyper-parameters frequently necessitate experimentation and modification based on the specific context and model architecture.

1. **Batch size :** It is the amount of data processed simultaneously during a single forward and backward pass training, its value needs to be adjusted correctly, A large batch size can slow down network convergence, while small batch sizes can destabilize networks and lead to poor results (Kandel & Castelli, 2020).
2. **Epochs :** Indicates the number of full forward and backward passes for a neural network during training (Khan Raiaan et al., 2024), it is established throughout the experiment with no pre-existing ideal parameters to use (Sharma et al., 2022).
3. **Learning rate (LR):** is the step size that governs the amount of change made to the model's weights during the optimization process. An incorrect learning rate setting might result in poor performance, and vice versa. (Georgakopoulos & Plagianakos, 2017; Smith, 2015).
4. **Optimizer** An optimizer can be defined as an algorithm for machine learning that adjusts a model's weights to reduce (minimize) the loss function and increase efficiency (Kingma & Ba, 2014). There are many optimizers, each for specific needs, We parencite these three popular ones in the table 2.3.

Optimizer	Reference	How It Works
Adam	*(Kingma & Ba, 2014)*	- Implements stochastic gradient descent with memory-efficient replacement optimization.
SGD (Stochastic Gradient Descent)	*(Bottou, 2012)*	- Picks one training sample randomly each iteration. Faster but less stable at convergence.
AdaGrad	*(Duchi et al., 2011)*	- Optimizes sparse gradients. Adjusts learning rates based on parameter frequency.(Kandel et al., 2020)

Table 2.3: *The Three Popular Optimizers.*

Layers Hyperparameters

are the parameters that need to be fine-tuned for optimal problem resolution (Yamashita et al., 2018). They include convolutional layer parameters (for kernel and convolution operation settings) and dense layer parameters (such as dropout, etc.).

1. **Number of kernels and their size :** The kernels affect the feature extraction phase, where the number of kernels determines the amount of data (convolutional layer output) in the feature map, Less data is indicated with a smaller value (Huang et al., 2021). Furthermore, the size of the kernel impacts the convolutional operation time and resource cost, Cost is lower for smaller sizes than for larger kernels (Gao et al., 2020).
2. **Activation function :** Defined as a decision-making function, it facilitates learning intricate patterns by determining a neuron's output through non-linear transformation of its input. An appropriate choice of activation function accelerates learning (A. Khan, Sohail, et al., 2019).

   For example: sigmoid, ReLU, Leaky ReLU, tanh.
3. **Padding :** Padding extends the input matrix size in order to fit the kernel(filter) and manage convolutional operations (Prusa & Khoshgoftaar, 2017).
4. **Dropout :** It's the process of randomly dropping units(hidden or visible) from the neural network to avoid overfitting caused by generating many parameters when working with numerous data points in neural networks (N. Srivastava et al., 2014).

2.4 Transfer learning

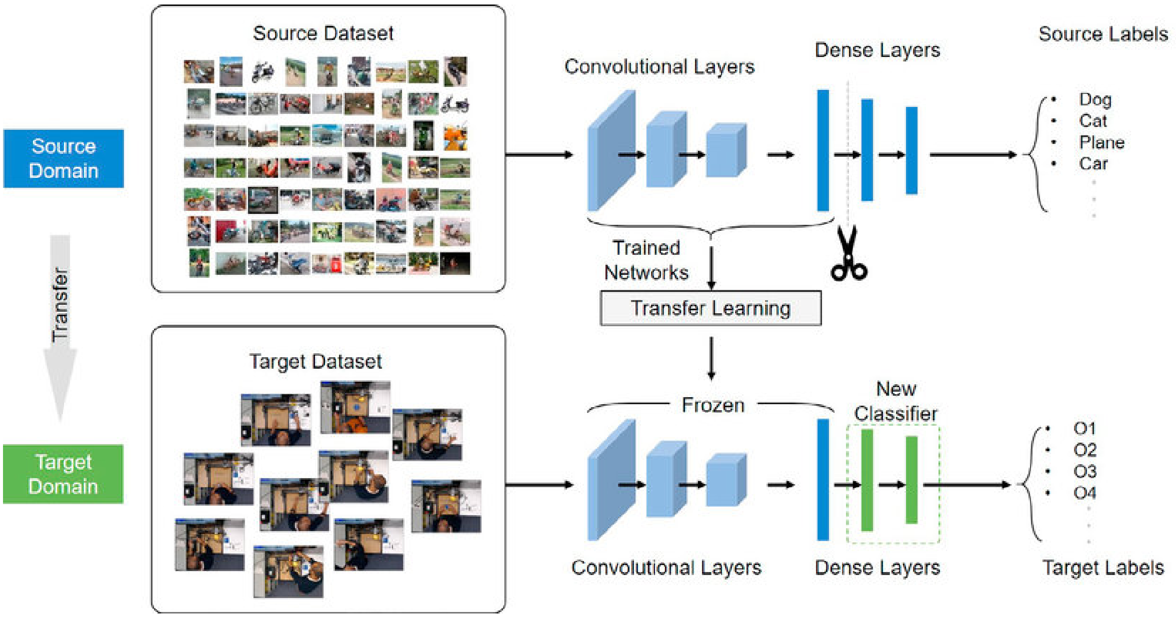


Figure 2.7: Transfer learning process (Tao et al., 2020).

Transfer Learning leverages knowledge from a source task to enhance performance on a target task by utilizing the generalization capabilities of pre-trained models. These models, trained on large datasets like ImageNet and COCO, learn to extract features that can be reused. Techniques include using these models as feature extractors by removing task-specific layers and using their output for new tasks, and fine-tuning (re-train some of the model's layers) , where the model's layers are updated with a lower learning rate to adapt to the new task (Hosna et al., 2022)(Olaoye & Potter, 2024).

It allows freezing early layers (layers that are closer to the input) to preserve general features and updating later layers (layers that are closer to the output) to capture task-specific features. The selection of layers for transfer depends on task similarity to avoid poor performance (Olaoye & Potter, 2024).

2.5 Most common CNN architectures

As a development of the convolutional neural network architecture, many architectures have emerged to improve and enhance the performance of convolutional neural networks. These architectures are trained on millions of images, such as ImageNet, and thus they save resources and time and can be used in many applications. Some of these architectures are :

2.5.1 VGG16

VGG16 is an abbreviation for Visual Geometry Group 16. It is a deep convolutional neural network model that expands on typical convolutional network architecture via the use of 16 layers of small 3x3 convolutions. The VGG16 pretrained model is trained on the ImageNet dataset, containing 14 million images broken down into 1,000 distinct classes, and then customized using the Keras framework(Simonyan & Zisserman, 2014). The VGG16 architecture is illustrated in the Figure 2.8.

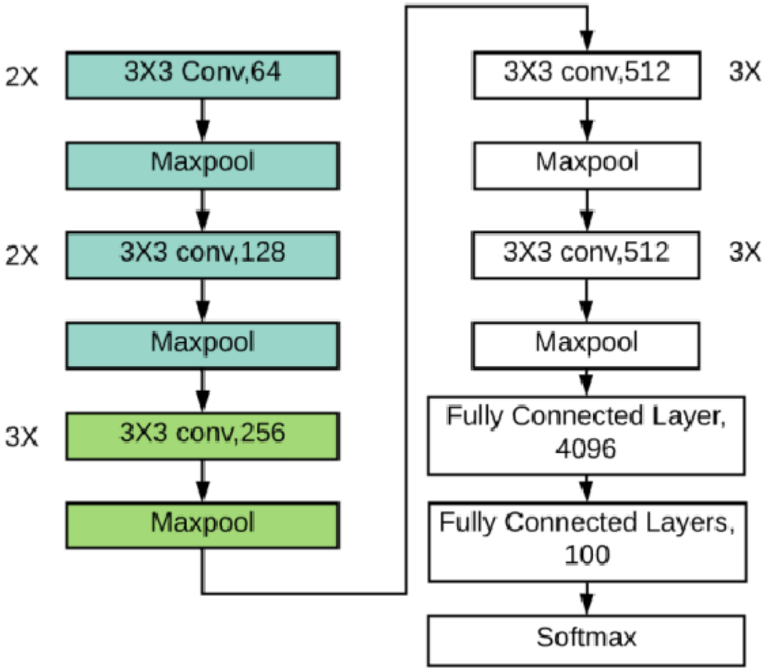


Figure 2.8: *The VGG16 architecture (Pal et al., 2020).*

2.5.2 ResNet50

ResNet-50 is the acronym for Residual Network. It is a deep residual learning framework intended for image classification tasks. It consists of fifty layers, including residual blocks. Each block has convolutional, pooling, and fully connected layers (K. He et al., 2015) as shown in the Figure 2.9.

This residual network employs the notion of skip connections, or shortcut connections, between one or more layers to speed up the training of extremely deep neural networks and enable the network to learn residual functions rather than direct mappings (K. He et al., 2015).

Pre-trained ResNet-50 model, such as those offered by Keras libraries, have been trained on a large benchmark dataset, such as ImageNet, and may be fine-tuned for a variety of computer vision tasks, reducing computing resources as well as time (Choe et al., 2020).

Residual Connection Types

Residual connections are classified into two types based on the dimensions of the residual block input and output. The connections are called Identity Shortcuts and Projection Shortcuts. The ResNet paper defines these as follows (K. He et al., 2015):

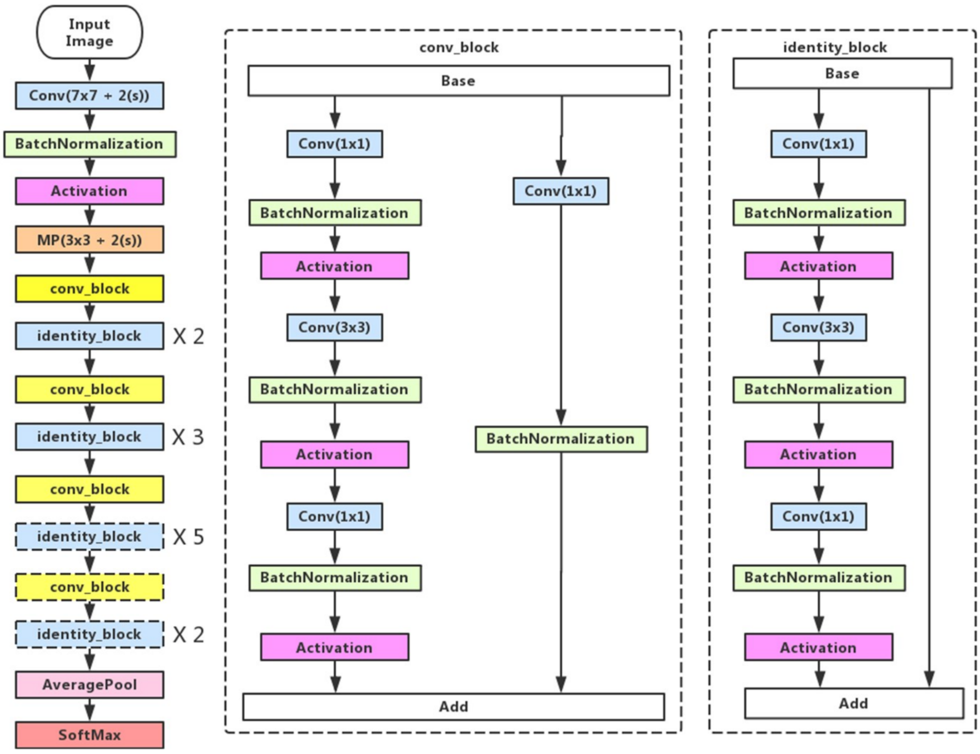


Figure 2.9: *The ResNet50 architecture (Ji et al., 2019).*

1. Identity Shortcuts :

Employed whenever the input and output feature maps have identical dimensions. This form of shortcut maintains the input data in its original dimensions and adds it directly to the output.

If X is the input to a residual block and the output after performing the block's layers (e.g., convolution, batch normalization, and activation) is $F(X)$, therefore the block with an identity shortcut has the following output 2.1:

$$Y = F(X) + X \tag{2.1}$$

2. Projection Shortcuts :

Occurs when the input feature map's dimensions differ from those of the output feature map. To address the non-matching issue, the 1x1 convolutional layer adjusts the number of channels (depth) and occasionally the spatial dimensions (height and width) of the input feature map so it matches the output.

If the input to a residual block is X and the output after applying the layers within the block is $F(X)$, and a 1x1 convolutional layer, W, has been selected for the projection shortcut, then the block's output is 2.2:

$$Y = F(X) + WX \tag{2.2}$$

The impact of residual connection is illustrated in the following Figure 2.10.

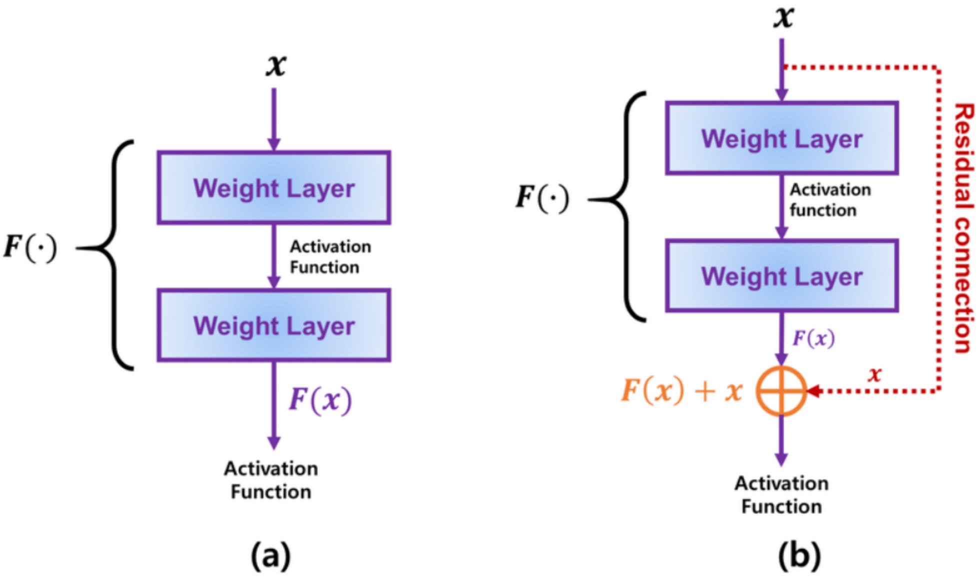


Figure 2.10: *Comparison of general and residual connections in ResNet. (a) shows input x through weight layers in standard CNNs. (b) shows x through both weight layers with residual connections in ResNet (K.-S. Kim & Choi, 2021).*

ResNet-50 Advantages:

ResNet exceeded the deep network (Ebrahimi & Abadi, 2018), by resolving these two challenges that deep networks face:

- During training, neurons may die due to disappearing or exploding gradients (Bengio et al., 1994; Glorot & Bengio, 2010).
- Adding extra parameters complicates optimization, resulting in higher training errors, not only overfitting (R. K. Srivastava et al., 2015).

2.5.3 Xception

The acronym Xception (Chollet, 2017) refers to Extreme Inception, It is a pretrained deep convolutional neural network design that advances the Inception architecture by implementing depthwise separable convolutions. Xception preptrained model is trained on ImageNet dataset, with 14 million images organized into 1,000 different classes and is fine-tuned via the Keras framework. Its architecture shown in the Figure 2.11.

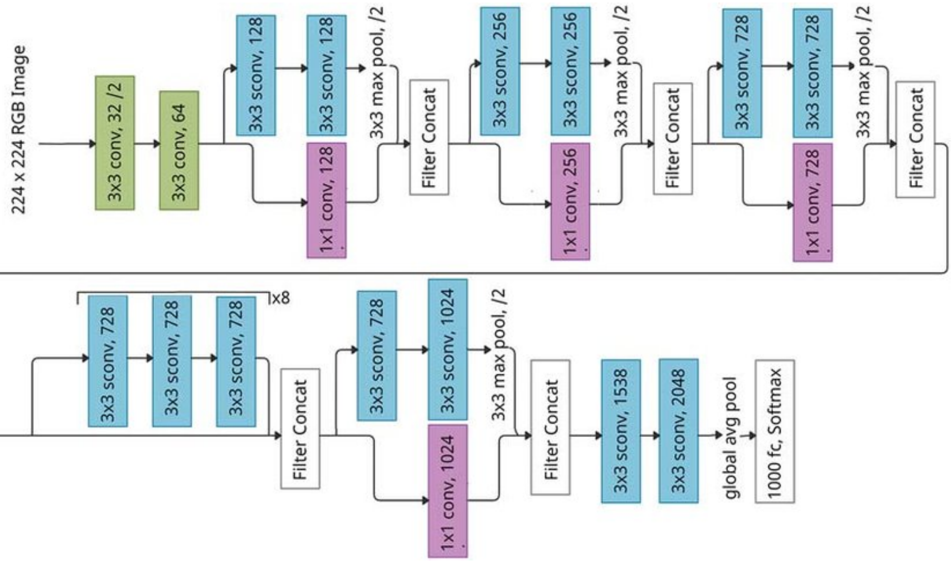


Figure 2.11: *The architecture of the Xception (Srinivasan & et al., 2021).*

2.6 literature review

Melanoma is one of the most dangerous and deadly types of skin cancer, as it has the highest number of deaths annually compared to other types according to statistics from the World Health Organization, as this type, as mentioned in the first chapter, is characterized by the risk of spreading to all parts of the body, and therefore controlling or limiting it becomes ineffective when it reaches that stage.

Diagnosis of melanoma depends on visual examination by a specialist, and visual examination has a confidence level of 65 percent (Lee, 2001), which makes it a very low percentage despite the emergence of dermoscopic devices that image the lesion with high quality, as the accuracy of the diagnosis is always subject to the standard of objectivity and the educational background of the specialist. Which calls for the necessity of automatic diagnosis, which makes the trend towards artificial intelligence as it is the most popular technology recently, so research began with machine learning techniques going to deep learning techniques.

The review is divided into a section to examine machine learning techniques where the features of the lesion are extracted using algorithms that simulate the diagnostic techniques approved by doctors. Then we move to the section on deep learning where it was Automatic feature extraction by techniques such as convolutional neural networks. Finally We conclude the review with a table containing published works that deal with melanoma diagnosis.

Melanoma detection using Machine Learning techniques

Radiologists typically use pattern analysis, the 7-point checklist, and the ABCD rule of dermoscopy to determine the difference between malignant and benign skin lesions. Several studies have used machine learning to implement these guidelines as a features extraction methode for the melanoma diagnosis systems :

(Chatterjee et al., 2021)'s work used the ABCD rule to aid the dermatological expert system (DerESy) in diagnosing melanoma with irregular forms. This algorithm has been improved by taking into account the spatial features of dermoscopic structures in order to recognize malignant lesions more accurately. They employed the ISIC challenge datasets from 2016, 2017, and 2018, as well as the PH2 dataset. Preprocessing was performed to eliminate noise with a median filter, remove hair, and segment lesions using the lesion segmentation module (LSM). The system had a diagnostic accuracy of 97.86%, sensitivity of 97.69%, and specificity of 97.97%.

For the melanoma classification phase, (L. Singh et al., 2020) utilized a three-layer Support Vector Machine (SVM) classifier, with a Preprocessing involved eliminating noise (e.g., bubbles, hair), segmenting lesion images, and extracting features. The ABCD rule was used to extract characteristics relating to the lesion's asymmetry, color, border, and diameter, which were then passed on to the SVM classifier, which classified the images into one of three categories.The system was tested using the PH2 dataset, which includes 200 dermoscopy images where 160 are non melanoma and 40 are melanoma lesions.The proposed approach scored 92.5% accuracy, 87.5% specificity, and 100% selectivity.

In another study by (Senan & Jadhav, 2021), the ABCD criteria were also applied for feature extraction to create an effective system for early diagnosis of skin cancer. The suggested approach was evaluated using the PH2 dataset, which includes three types of skin diseases: atypical nevi, melanoma, and common nevus. The model is divided into two phases: preprocessing (using a median filter to decrease noise) and extraction of the Region of Interest (RoI). In the next stage, features retrieved using ABCD rules are processed using the Total Dermoscopic Score (TDS). Because of the high-resolution images, the system obtained an 84% accuracy rate.

(Shetty et al., 2022)compared machine learning and CNN approaches for classifying skin lesion images in the HAM10000 dataset. Thier system organization included arranging the dataset, preprocessing, and data augmentation followed by the classification. During the classification phase, the adapted CNN surpassed the machine learning algorithms, with these ML techniques achieving accuracies of 58% (Logistic Regression), 57% (Linear Discriminant Analysis), 48% (K-Nearest Neighbors), 68% (Decision Tree), 87% (Random Forest), 36% (Naive Bayes), and 53% (Support Vector Machine).

(Imtiaz et al., 2021) employed ABCD rules for feature extraction and Adaboost for skin cancer classification on the PH^2 dataset. Despite the contrast issue and for the aim to increase the model accuracy, they used morphological operations for noise reduction, Otsu thresholding and image subtraction to recover the region of interest, as a preprocessing to eliminate noise from dermoscopic images. The PH^2 dataset was balanced using SMOTE. Utilizing Adaboost, the technique produced segmentation accuracy of 90.25% and classification accuracy of 98.13% for melanoma and non-melanoma.

These studies are summarized in the Table 2.4, for greater clarity of the results, the databases used, and the machine learning techniques.

Study	Method	Dataset	Results
(Chatterjee et al., 2021)	ABCD rule, spatial features	ISIC 2016, 2017, 2018; PH2	Accuracy: 97.86%, Sensitivity: 97.69%, Specificity: 97.97%
(L. Singh et al., 2020)	Three-layer SVM classifier, ABCD rule	PH2 (200 images)	Accuracy: 92.5%, Specificity: 87.5%, Selectivity: 100%
(Senan & Jadhav, 2021)	ABCD rule, TDS score	PH2 (atypical nevi, melanoma, common nevus)	Accuracy: 84%
(Shetty et al., 2022)	Comparison of ML and CNN	HAM10000	CNN outperformed ML, ML accuracies: 36%-87%
(Imtiaz et al., 2021)	ABCD rules, Adaboost	PH2	Segmentation accuracy: 90.25%, Classification accuracy: 98.13%

Table 2.4: *Summary of Studies on Skin Lesion Classification using machine learning.*

Melanoma detection using Deep Learning techniques

The issue of dataset imbalance, where the proportion of positive cases is significantly lower than negative cases, is a common problem in many clinical datasets and can introduce a bias that results in an unacceptably high rate of false negatives when using the model in real-world clinical applications, diminishing the number of informative data points and reducing the accuracy of many machine learning techniques (H. He & Garcia, 2009).

The latest technology used recently to solve the problem of database imbalance due to the lack of availability of images is ensemble Learning (Polikar, 2012).

Using this technique, the authors of this study (Qureshi & Roos, 2023) proposed a structure consisting of six CNN-based models, two of which were pre-trained on other balanced datasets. The ensemble was trained on ISIC2020, which consists of dermoscopic images along with meta-data about the patients, such as age and tumor location. Later on, the group's decisions were collected and unified as a final classification by a support vector machine in order to achieve higher accuracy. The proposed technique achieves average F1, AUC-PR, and AUC-ROC values of 0.23, 0.16, and 0.87, respectively. However, the limited generalizability due to dataset bias and the need for further validation on diverse datasets were noted as study limitations.

(Rahman et al., 2021) present a weighted average ensemble learning-based model for multi-classification skin lesions. They designed the ensemble based on five deep neural network models: ResNeXt, SeResNeXt, ResNet, Xception, along with DenseNet. They trained then evaluated their models using the HAM10000 and ISIC 2019 datasets, employing class balance, noise reduction, and data augmentation approaches. The five models performed quite well throughout examination, with macro-average recall scores of 88%, 89%, 91%, 88%, and 84% for ResNeXt, SeResNeXt, DenseNet, Xception, and ResNet, respectfully.

Several recently published papers have offered data augmentation as a method for data preparation for the aim to increase the number of images in the dataset to solve the imbalence :

For that purpose, (Bozkurt, 2023). proposed a hybrid CNN-based model, which combines Inception and ResNet with affine transformation techniques to diversify and augment images in the dataset. The model was evaluated on the HAM10000 dataset both with and without data augmentation. The results showed a 95.09% accuracy for the Inception-ResNet-v2 model with augmentation, and 83.59% without it. However, these results were limited by an imbalanced dataset, highlighting the need for high computational resources and the potential for overfitting.

Where the published work of (Swain & et al., 2021). focused the light on the effectiveness of data augmentation on the accuracy of their proposed 13-layer Deep Convolutional Neural Network (DCNN) model. The model was trained and evaluated using the PH^2 dataset, achieving an accuracy of 78% on the real dataset and 81% on the augmented dataset using geometric transformation. However, the study was limited by class imbalance, the effectiveness of which depends on the choice of augmentation methods, and potential overfitting.

(Maqsood & Damaševičius, 2023). suggested a computer-aided design methodology for skin cancer segmentation and classification. Dermoscopic images were preprocessed and segmented using a customized 26-layer CNN. Four previously trained CNN models (Xception, ResNet-50, ResNet-101, and VGG16) were adjusted and trained with transfer learning, Deep feature vectors were extracted and merged using convolutional sparse image decomposition. The best features were chosen using univariate measurement and the Poisson distribution, before being classified using a multi-class SVM. The system obtained 93.7% accuracy on the ISIC 2019 dataset. Limitations included poor contrast lesions, unbalanced datasets, excessive memory complexity, and duplicate feature extraction.

Study	**Method**	**Dataset**	**Results**
(Qureshi & Roos, 2023)	CNN ensemble unified by SVM	ISIC2020	F1: 0.23, AUC-PR: 0.16, AUC-ROC: 0.87
(Rahman et al., 2021)	DenseNet	HAM10000, ISIC 2019	Macro-average recall: 91%
(Bozkurt, 2023)	Inception-ResNet-v2 with augmentation	HAM10000	Accuracy: 95.09%
(Swain & et al., 2021)	13-layer DCNN with augmentation	PH^2	Accuracy: 81%
(Maqsood & Damaševičius, 2023)	26-layer CNN with transfer learning	ISIC 2019	Accuracy: 93.7%

Table 2.5: *Summary of studies on skin lesion classification methods using deep learning*

Some additional studies where their classification task is the same as ours binary classification. These studies are shown in the following table 2.6.

Study	**Method**	**Classification type**	**DataSet**	**ACC**
(Mehra et al., 2021)	ResNet-50	multi-classification	HAM 10000	84.87%
(Barbosa et al., 2019)	ResNet	Binary-classification	ISIC 2017	69%
(Hekler & Kather, 2020a)	CNN model	Binary-classification	804 images for training, 384 images for testing.	75.03%
(Hekler & Kather, 2020b)	CNN	Binary-classification	HAM10000, ISIC (not specified)	75.03%
(Adegun & Viriri, 2020)	Enhanced encoder-decoder network with multi-stage and multi-scale approach.	Binary-classification	ISIC2017,PH^2	95%
(Cassidy et al., 2022)	VGG19	Binary-classification	ISIC2020	AUC:80%
(Cassidy et al., 2022)	VGG19	Binary-classification	ISIC2017	56%

Table 2.6: *Some related work*

2.7 Conclusion

Due to the rapid technological development in recent years, the spotlight has been on the automation of disease diagnosis in the medical field to raise the accuracy rates of diagnoses and thus increase survival rates, specifically in skin cancer, where melanoma has received great attention due to being a deadly tumor. The application of deep learning techniques, especially those specific to image processing such as convolutional neural networks (CNNs), has become an essential solution.

This chapter offers an in-depth exploration of deep learning techniques. It begins with a presentation of the foundations of neural networks, before focusing on convolutional neural networks (CNNs) and the architectures derived from them and based on pre-trained models. It concludes with a review of research into melanoma diagnosis.

Chapter 3

Conception

3.1 Introduction

The least common but deadliest form of this killer disease is melanoma which is especially difficult to diagnose medically. Because if it develops, the prognosis is very bad, and its early detection is important. Nonetheless, melanoma is incrementally diagnostically challenging, requiring significant specialisation as a result of its intricate and often subtly occurring symptomatology. The major step in the process of melanoma detection is the feature extraction phase which is responsible for determining relevant information in the skin lesion images that help in differentiating malignant from benign tumors.

Over the past few years, deep learning, and especially Convolutional Neural Networks (CNNs), have changed the way a data scientist thinks about feature extraction. CNNs are good at learning complex patterns directly from image data, can automatically learn features without manual extraction, and greatly improve the diagnostic accuracy. Yet data is massively skewed towards images of healthy skin, exacerbated by the few number of melanoma images and the significant privacy constraints when it comes to patient data. Such an imbalance creates an urgent challenge: machine learning models trained on uneven datasets are subject to bias, which severely undermines their generalization and robustness.

In this chapter, we provide an overview of the procedure we followed for dividing the dataset using the proposed balancing technique. Next, we present the architecture of our melanoma detection system, which comprises three phases, starting from the dataset preparation to the final phase where the diagnosis decision is made. presenting pre-trained model used as a good practice for feature learning which will also improve the classification.

3.2 Objective

Our objective is to handle the imbalance in the dataset, as it can affect the reliability of a diagnostic system. To address this problem, we introduce a new method for distributing data more fairly. By dividing the dataset so that the ratio of each class is balanced, we aim to improve accuracy and achieve more unbiased classification. Furthermore, this method benefits the entire existing CNN system for diagnosis.

3.3 System architecture

In this work, we propose an approach for melanoma detection using a deep learning approach. Our system aims to determine whether a given lesion is melanoma or not. It incorporates a crucial feature extraction phase essential for any classification process, a proposed method to balance the dataset, and a classifier to ensure a final accurate classification.

Our system architecture is illustrated in the following Figure 3.1, This system for melanoma detection consists of three phases: dataset preparation, the training phase, and finally, the decision and classification phase. Each phase is responsible for a specific task to ensure the reliability of the entire system, and they are as follows:

- **Dataset preparation phase:**

  This phase includes all the required steps for preparing the dataset to be fed into the deep learning technique in the next phase. The process involves dividing the dataset into k subsets based on the proposed balancing method, followed by the splitting and preprocessing of these k subsets, which includes normalization and augmentation.

- **Training phase:**

  The input for this phase is the output of the first phase, which consists of the prepared k subsets. Each subset is then used for a training phase with a CNN-based pretrained model that offers automatic feature extraction and classification. After undergoing rigorous training, each model will be able to accurately detect and classify melanomas.

- **Decision and classification phase:**

  This final step depends on the results of the deep learning CNN-based methods, where each model learns a specific subset. Together, these models form a unified decision. This step consolidates the solutions through decision fusion. The decisions are combined to diagnose the lesion using a classifier that determines the final output probability.

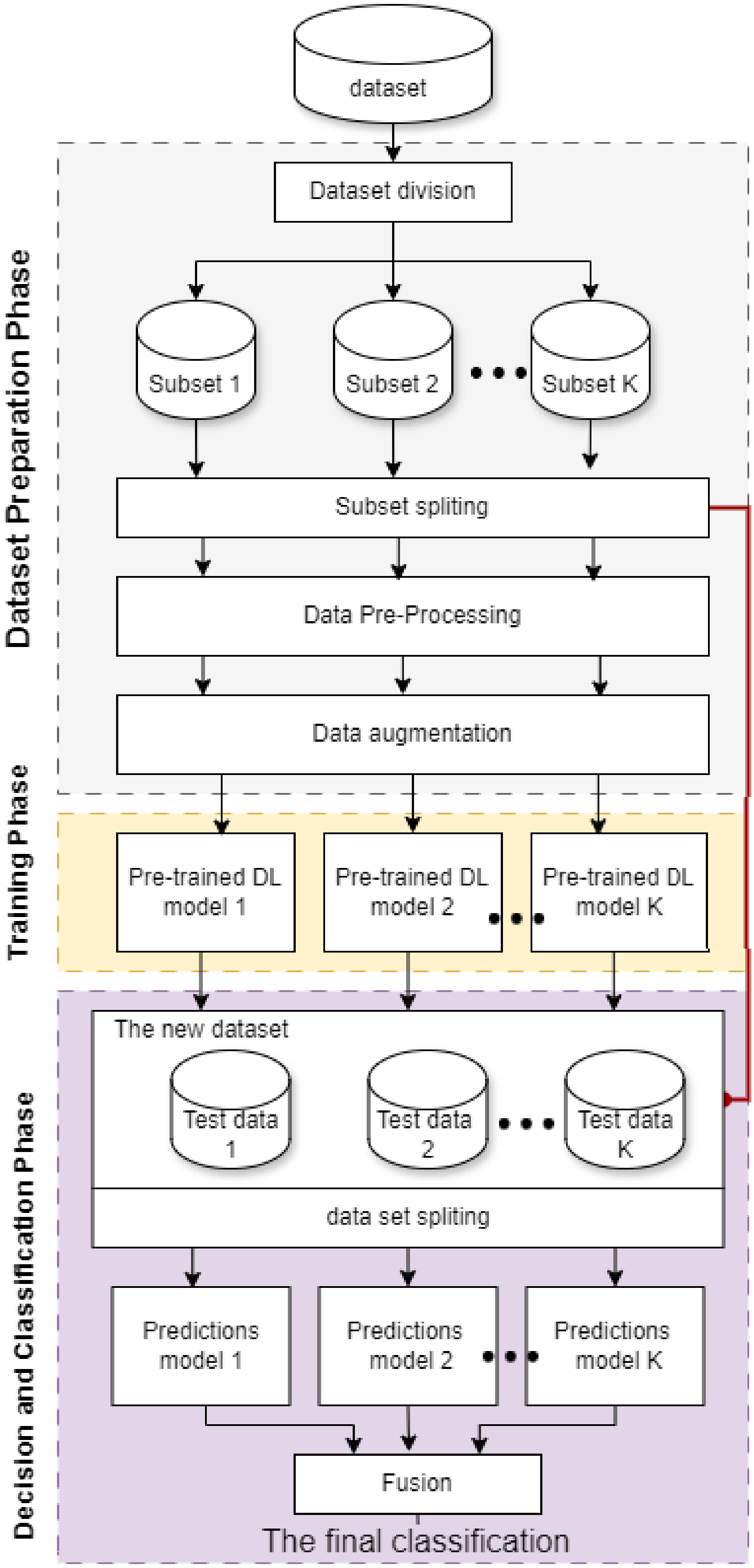


Figure 3.1: *Melanoma Classification System Architecture*

3.4 Dataset preparation phase

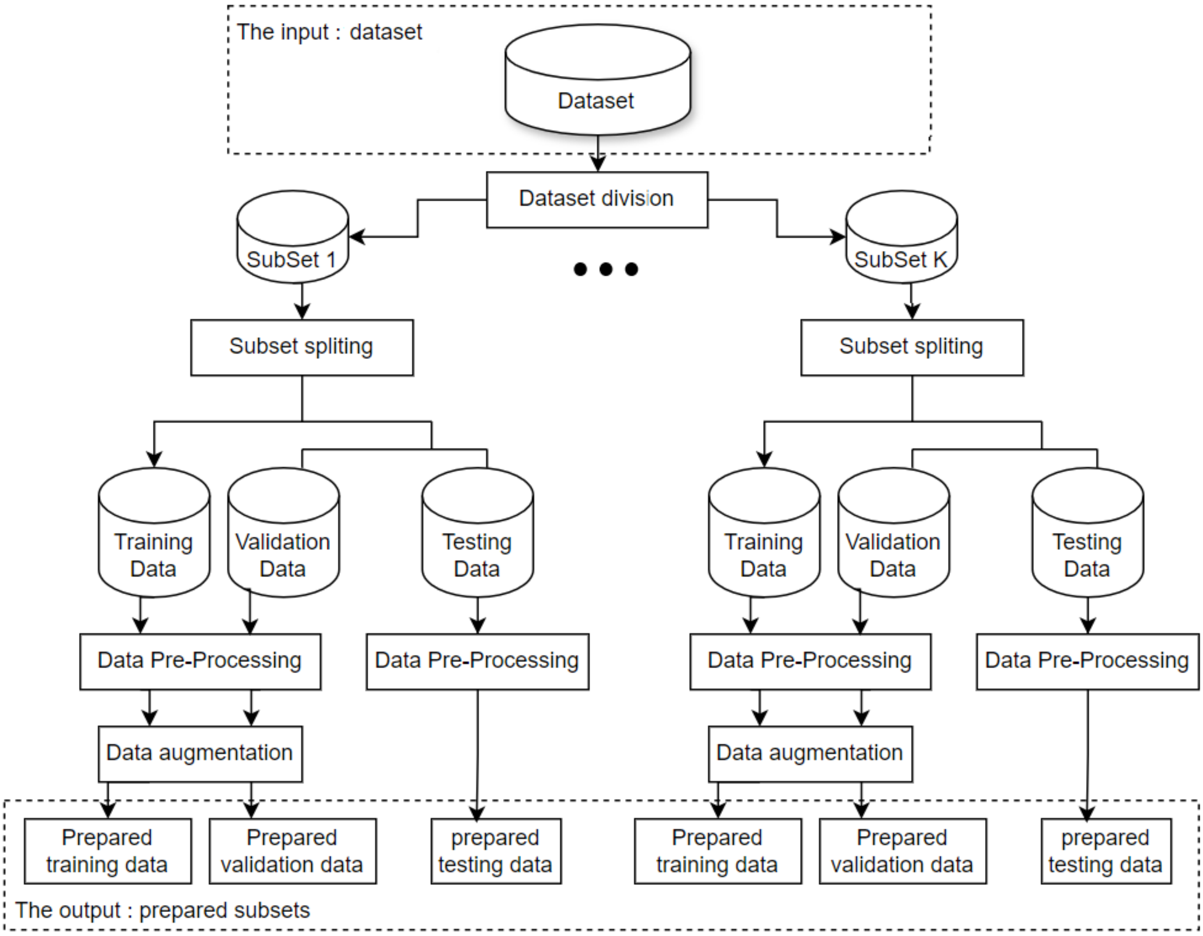


Figure 3.2: *Dataset preparation phase architecture*

This phase includes all the operations that will be applied to the database from applying the proposed balancing technique as dataset division method, to the pre-processing. The input of this phase is the binary database while the outputs are a k prepared subsets that is pre-processed and splitted and ready to train deep learning models on it.

3.4.1 Proposed balancing technique

Dealing with class imbalance is a critical challenge in binary classification problems, especially in medical datasets where one class (e.g., "non-melanoma") is vastly over represented compared to the other (e.g., "melanoma"). In the context of a dataset consisting of medical images labelled with a cancer diagnosis, we introduce an alternative approach to divide the majority class to ensure balanced training and test sets, which is crucial for the performance and reliability of deep learning models.

Our objective is to divide the majority class into k balanced subsets while ensuring that each subset will contain a proportional number of samples from both classes (melanoma, non-melanoma).

This alternative method for dividing the majority class in a dataset involves the following steps:

1. Define k (The number of subsets into which the majority class samples will be divided), based on the size of minority class (melanoma).
2. Define a partition function F to distribute the majority class samples across k subsets.
3. Ensure that the majority class samples are uniformly distributed.
4. Define the subsets according to the function F.

1. Problem definition

Let D be the binary dataset, D consisting of medical images, with labels "non-Melanoma" (C0), "Melanoma" (C1). When the dataset is highly unbalanced, the majority category (C0) is considerably more frequent than the minority category (C1). Let be :

- n: Total number of samples.
- $D = \{(x_i, y_i)\}_{i=1}^{n}$: The dataset where x_i is an image and y_i is the label.
- N_0: Number of samples labeled as "non-melanoma".
- N_1: Number of samples labeled as "melanoma".
- k: The number of subsets into which the majority class samples will be divided.

2. Definition of the partition function F

The partition function F distributes the samples of the majority class into k balanced subsets (k being defined in equation 3.1), by mapping the indices of the samples of the majority class to one of the k subsets, the partition function is defined in equation 3.2.

$$k = (N_1 Div N_2) + 1 \tag{3.1}$$

$$F(i) = ((i-1) \mod k) + 1 \quad \text{for } i \in \{1, \ldots, N_0\}. \tag{3.2}$$

The function F assigns each sample (i) in the majority class to one of the k subsets. The aim is to ensure that the samples are uniformly distributed between these subsets. Mathematically, This should achieve the following equality 3.3, Where the length (total number of samples) should be equal to the total number of samples in the dataset divided by k (the number of subsets).

$$\left|F^{-1}(s)\right| \approx \frac{N_0}{k} \quad \forall s \in \{1, \ldots, k\} \tag{3.3}$$

where :

- $F^{-1}(s)$: This represents the inverse mapping of a subset label s. It identifies the indices of samples assigned to subset s.
- k: The number of subsets.
- N_0: Number of samples labeled as "non-melanoma".

- $|F^{-1}(s)|$: The size of the subset s, meaning the number of samples assigned to subset s.

3. Subset expression

Once the equation for dividing the majority class into k subsets based on the total number of samples in the minority class has been defined, we can associate each subset to one of two classes: Non-Melanoma and Melanoma. since Melanoma is the minority class, it remains consistent across all the subsets. The formula for defining the subset it is as follows 3.4:

$$S_s = \left\{(x_j, y_j) \mid y_j = 0, j \in F^{-1}(s)\right\} \cup \{(x_j, y_j) \mid y_j = 1\} \tag{3.4}$$

where :

- Non-melanoma samples: $\{(x_j, y_j) \mid y_j = 0, j \in F^{-1}(s)\}$
- Melanoma samples: $\{(x_j, y_j) \mid y_j = 1\}$
- (x_j, y_j): A sample where x denotes the image and y denotes the label.
- $y_j = 0$: Non-melanoma label.
- $y_j = 1$: Melanoma label.
- $F^{-1}(s)$: Indices of samples assigned to subset s.

3.4.2 Data preprocessing

In the preprocessing phase, we ensure the appropriate input for the proposed architecture by applying resizing, image normalization, and preprocessing steps. These steps are elucidated as follows:

1. Image normalization

Before the dataset images are passed on to a pretrained model, the pixel values of the images are standardized to a specified range, commonly [0, 1] or [-1, 1], depending on the model's requirements.

This is done to normalize the pixel values of the image in a similar distribution and further to enable the model's structure and optimisation process to work well and achieve an improved convergence plus overall performance (Patro & Sahu, 2015).

2. Resizing

As a crucial pre-processing step, and acknowledging the dataset's diversity in image sizes (ranging from high resolution to various smaller dimensions), the dataset samples needs to uniformly resized to a standardized dimension. This resizing aligns with the common input size requirement for many pre-trained models, ensuring compatibility and optimal performance during model training and inference.

3.4.3 Data augmentation

Data augmentation is an important process for enhancing the generalization capabilities of deep learning model by making it see different transformations on images and learn to detect target features in different conditions, resulting in increased testing accuracy and reduced over-fitting (Shorten & Khoshgoftaar, 2019).

One of the data augmentation methods we used involves applying random transformations to each image during training without increasing the number of samples, ensuring that a diverse set of inputs is passed to the model. The transofmations are :

a) Rotation

Rotation in data augmentation involves rotating the image by a certain angle between 1degree to 359 degree, selected randomly or predefined, an example is provided in the Figure 3.3.

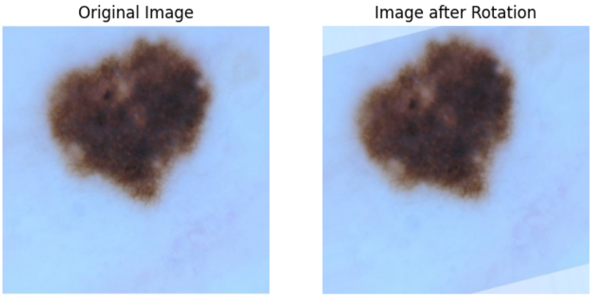


Figure 3.3: *Rotation of an an image from dataset.*

b) Flipping

Flipping involves mirroring the image horizontally or vertically, where Vertical axis flipping is far less prevalent than horizontal axis flipping.an example is provided in the Figure 3.4.

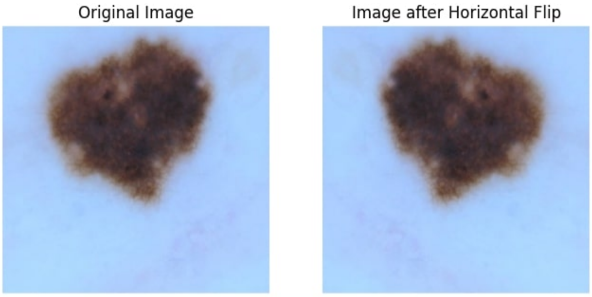


Figure 3.4: *fliping.*

c) Translation

Shifting images up, down, left, or right requires translating the image in a particular direction. Depending on the settings, the remaining area is filled with a constant value, random or Gaussian noise. This method is crucial since it helps the model develop the ability to identify objects in a variety of locations, eliminating positional bias and boosting its capacity to generalize beyond perfectly centered pictures.

In our case, we give width and height adjustments within a small range to prevent losing significant characteristics while accounting for the existence of non-centered images in the dataset.

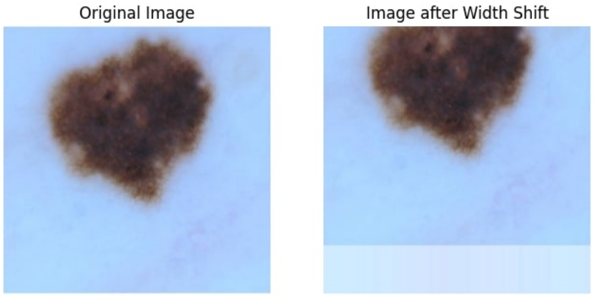


Figure 3.5: *shifting.*

3.5 Training phase

This is the phase in which the dataset preparation stage will obviously take the outputs of the preparation phase (prepared subsets) as inputs for the pre-trained deep learning models.

This phase provides each subset (s) with its own pre-trained model (Ms), which will go through a learning stage shown in the Figure 3.6. After all models go through the learning and evaluation process, the result is trained models ready to provide predictions.

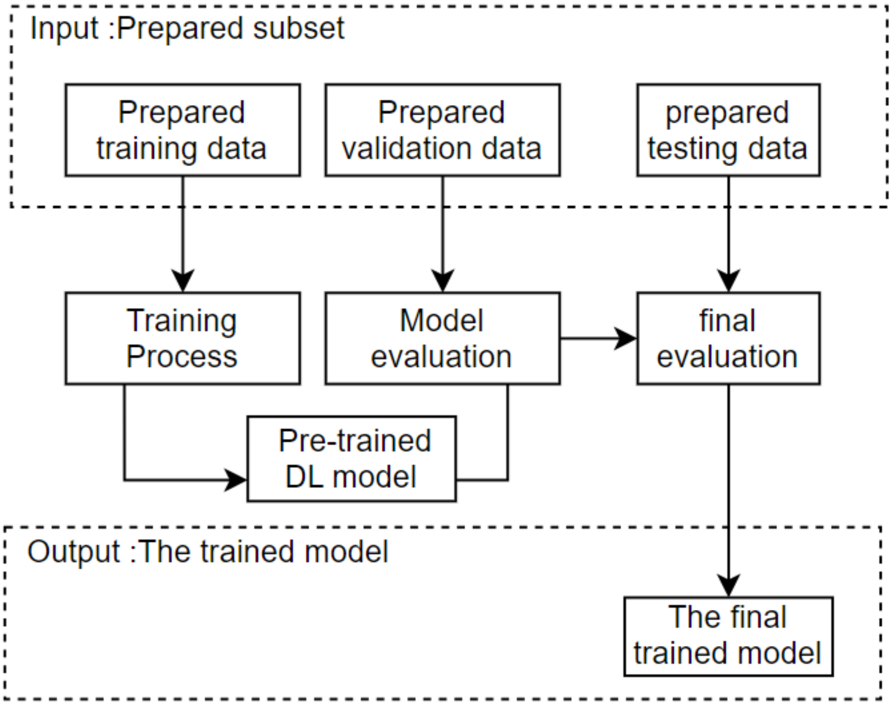


Figure 3.6: *the training phase architecture.*

3.5.1 Model architecture

We use pre-trained models in our Melanoma detection system, all of which share the ResNet50 architecture. We retained the base part of ResNet50 for feature extraction leveraging its pre-trained weights on ImageNet for transfer learning. Further we added our customized top layers to adapt the pre-trained model to our specific task.

The architecture of ResNet50, as detailed in Chapter Two, is characterized by its 50-layer depth and the innovative use of residual connections. These residual connections come in two primary forms: the identity shortcut connection and the projection shortcut connection. After removing its top layers and replacing them with our customized top layers, the detailed architecture can be seen in Figure 3.7.

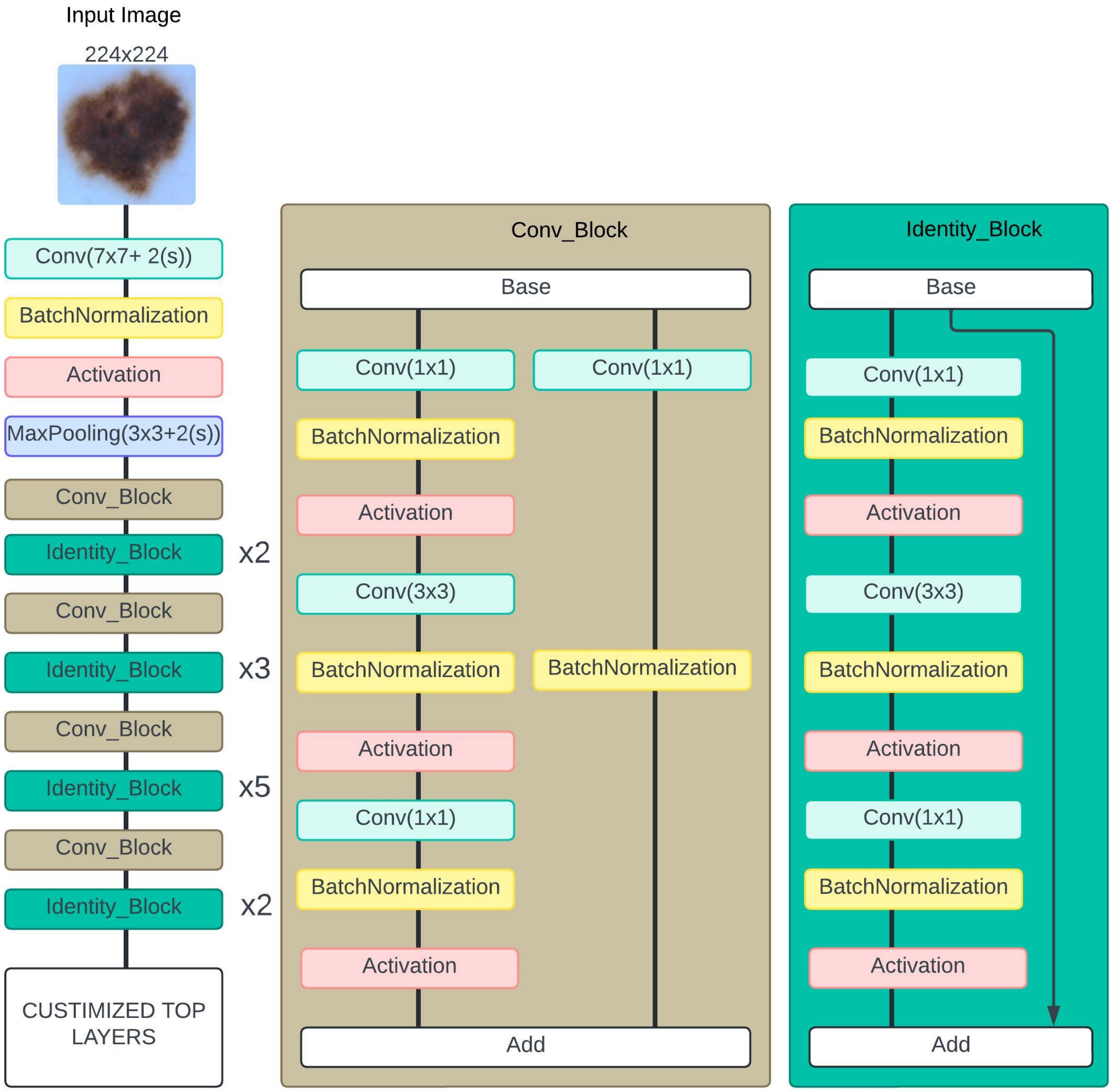


Figure 3.7: The architecture of the used ResNet50 pre-trained model in ours system.

Customized top layers

According to the ResNet50 architecture, following the residual blocks, there is an average pooling layer followed by an output layer with a softmax activation function that fits the categorical classification of ImageNet. These top layers were removed and replaced by our customized top layers, which were chosen through experimentation and tailored to the binary classification task for melanoma detection. These layers are illustrated in the Figure 3.8, and explained in order as follows:

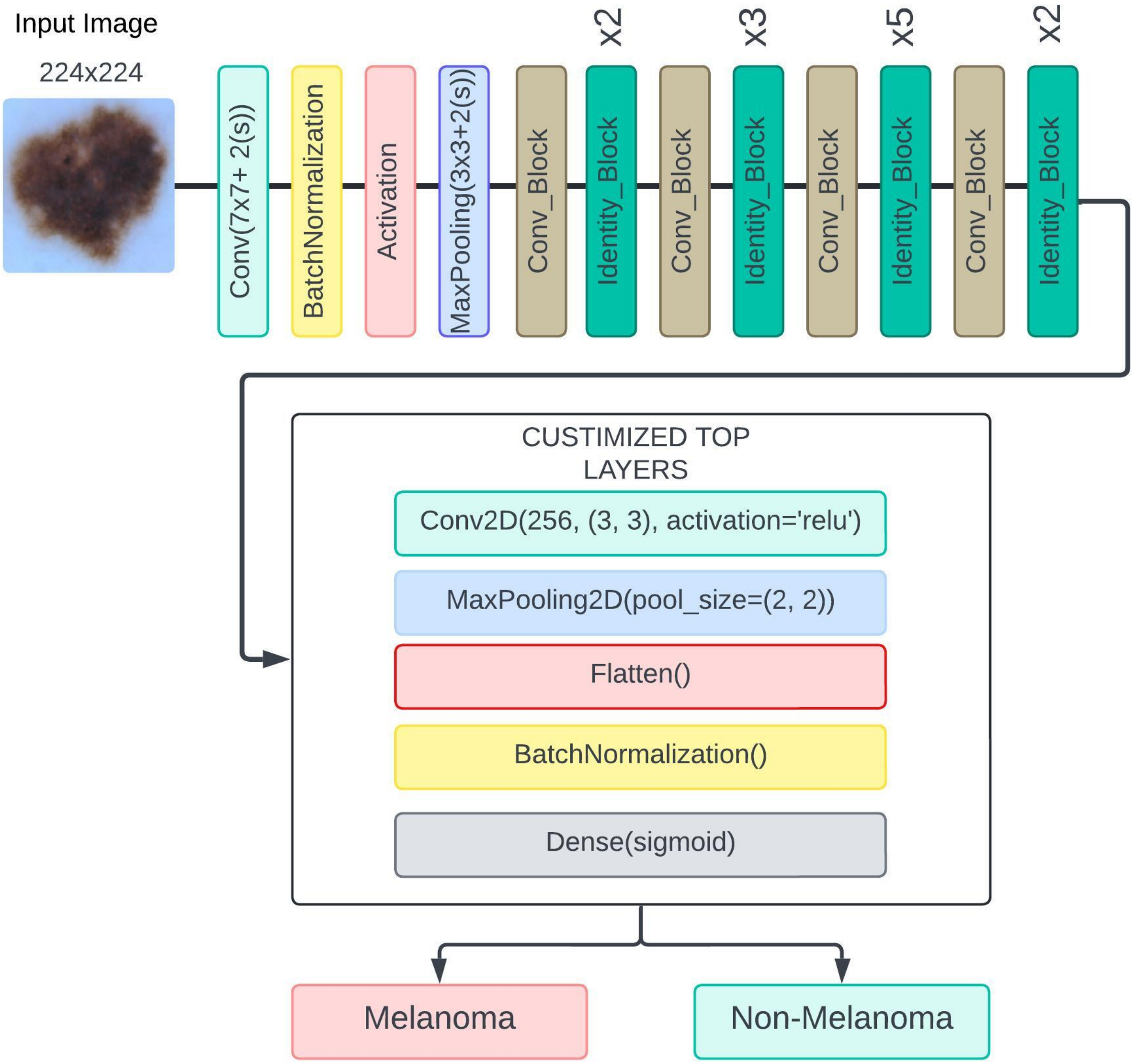


Figure 3.8: *Our Customized Top Layers.*

1. Convolutional layer

A convolutional layer uses a kernel (in simple terms, a square matrix that contains weights), performs a convolution operation on the input feature map given that kernel and then generates feature maps as a result of that operation highlighting certain aspects of that input, This layer is the first in the top layers following the pretrained Resnet50 model. Placing this layer after a pretrained model refines the extracted features further.

Key parameters include the number of kernels and their size (e.g. NxN,3x3) , with the activation function adding essential non-linearity. For our layer, we chose 256 kernels of size 3x3 and ReLU as the activation function, The stride, or step size, determines how the kernel moves across the input and is set by default to 1 in our case.

ReLU (Rectified Linear Unit) activation function outputs the input directly if it is positive; otherwise, it outputs zero. Mathematically, it's defined as 3.5:

$$f(x) = \max(0, x) \tag{3.5}$$

2. Max Pooling layer

A Max-Pooling layer performs a pooling operation on the input feature map with a specified pooling size (for example, 3x3). This process minimizes the size of the feature maps by picking the largest value from each pool, allowing significant characteristics to be retained while lowering spatial dimensions. This layer comes soon after the convolutional layer and processes the improved features.

Key options are pooling size and stride, which affect how the pooling window advances over the input. For our layer, we used a pooling size of 2x2. The stride is normally set to the same number as the pooling size, which in this case is two.

3. Flatten layer

It make flatten the input into one dimension. This is done by flattening the input data, i.e. squashing all dimensions except of the batch size. In short, no math operations (convolutions or pooling, for example) are carried. However, it only changes the form of input data to be suitable to feed to the fully connected layers in the neural network.

The input shape,dimension of feature map,flatten-length that could be calculate as multiplication of all dimension. Given an input to the flatten layer with dimensions of height H, width W, and depth D, the output will be a 1D vector with the following shape: $H \times W \times D$.

This flatten layer in our configuration is after max-pool layer which converts the multi-dimension of the data to 1d array format so that it could be fed into fully connected layers easily.

4. Batch Normalization layer

Batch normalization (BN) layers normalize the input data (by fitting scaling and shifting across each feature channel). It reduces internal covariate shift and helps the model to train faster providing the boost to the learning process. The normalization has accomplish by calculate the mean and variance for each feature channel across the batch, then normalized the input based on this statistics.

In BN layer, the essential parameters are the epsilon to avoid the zero division and the gamma/beta which are learnable during the training to scale/shift the normalized output.

Mathematically, for a given input x, the output y of a batch normalization layer can be expressed in the following expression 3.6:

$$y = \gamma \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \tag{3.6}$$

where μ is the mean of the input, σ^2 is the variance, γ and β are learned scaling and shift parameters, and ϵ is a small constant for numerical stability.

5. Output layer

finally we add a dense output layer with sigmoid activation function, which will return a single output for each neuron at the final layer, and often used in binary classification tasks. In order to be able to do this only the sigmoid activation function can map the output Y, to a value in between 0 and 1 and there way be used for modeling all level of probabilities, according to the sigmoid equation 3.7.

$$y = \frac{1}{1 + e^{-z}} \tag{3.7}$$

where z defined as shown in the Equation 3.8 :

$$z = \sum_i w_i x_i + b \tag{3.8}$$

where w_i are the weights, x_i are the inputs, and b is the bias.

The major hyperparameters are the regularization techniques L1 and L2 which is used to avoid overfitting in the training process. L1 regularization is 0.01, L2 regularization is 0.01 for our layer.

There addition of L1 and L2 regularization techniques in machine learning models, mainly in neural networks, for the aim to avoid over fitting. The regularization is added to the loss function of a model training by introducing a penalty that favours simpler models and discourages complex and non-generalizing models. encourage the model to generalize well to new data.

a) L1 Regularization:

L1 regularization or Lasso regularization, adds a penalty proportional to the absolute value of the model's weights. It can lead to sparsity in the weights, effectively setting some of them to zero, simplifying the model. The cost function for L1 regularization can be expressed in this Equation 3.9:

$$\text{Loss} = \text{Original Loss} + \lambda \sum |w_i| \tag{3.9}$$

where λ is a regularization strength that controls the degree of regularization.

b) L2 Regularization:

L2 regularization or Ridge regularization, adds a penalty proportional to the square of the model's weights. It encourages smaller weights and helps to maintain the model's generalization ability. It doesn't induce sparsity. The cost function for L2 regularization can be expressed in the following Equation 3.10:

$$\text{Loss} = \text{Original Loss} + \lambda \sum w_i^2 \tag{3.10}$$

where λ is a regularization strength that controls the degree of regularization.

3.6 Classification phase

In the context of our project, the classification phase is crucial because it consolidates the experiences and knowledge gained from training multiple models on different subsets of the data. The goal is to integrate these insights into a unified decision-making process. To achieve this, a fusion method needs to be applied to combine the predictions (probabilities) from the k models into one final decision. The classification phase is illustrated in Figure 3.9:

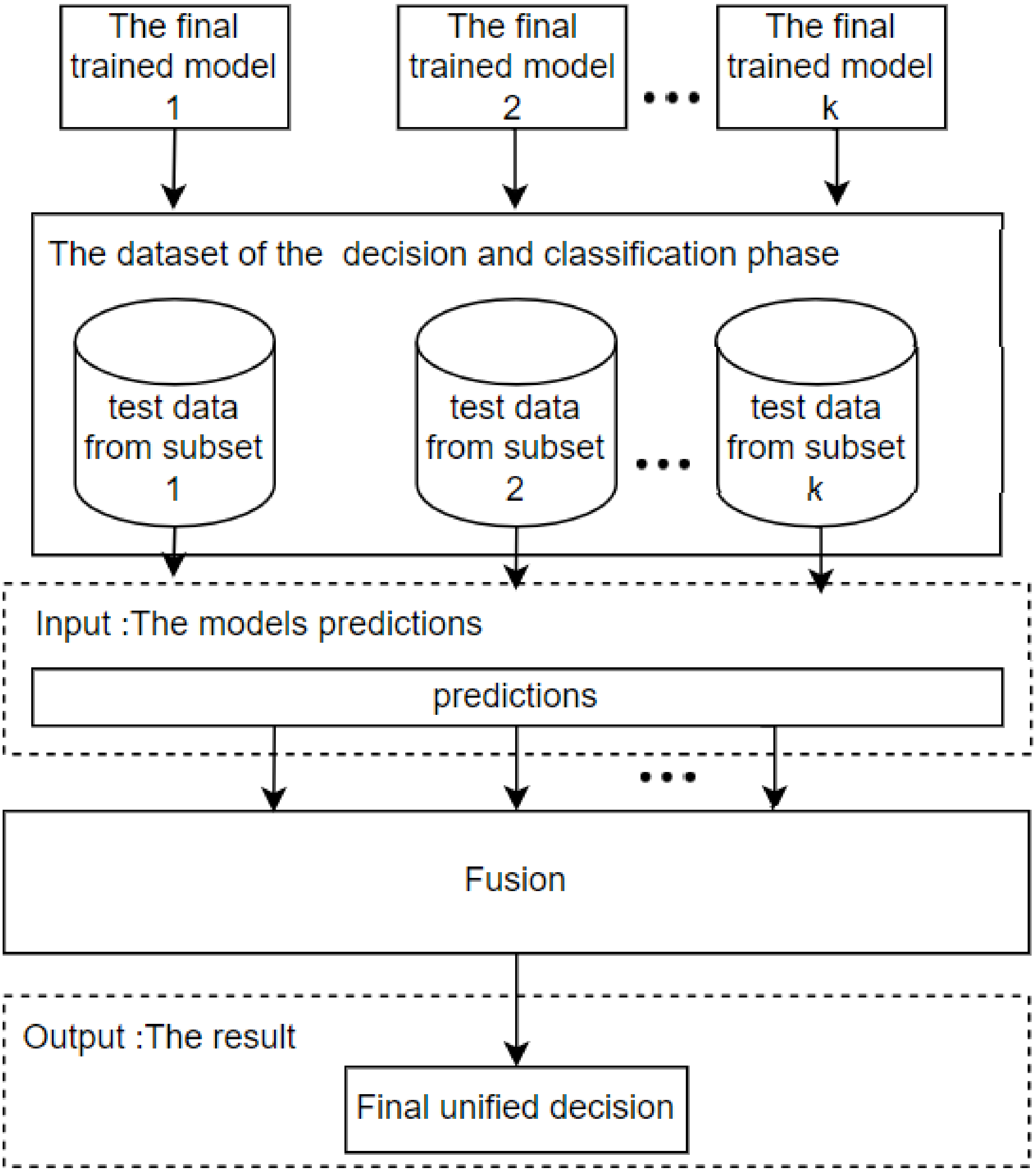


Figure 3.9: *The classification phase.*

In order to achieve this fusion, we used 3 techniques that fall within the machine learning classifiers, and for comparison, we added two techniques from the voting techniques.

3.6.1 Machine learning classifiers

The classifiers are used to identify patterns in the predictions that can enhance the unified decision of the k models. To enable the classifiers to have this ability, a training process needs to be applied. The predictions will then be split into train and test datasets to serve the training process. Finally, the classifiers will be ready to lead the final decision. The fusion performed by the machine learning classifiers is illustrated in the Figure 3.10 ,The classifiers used in this work are:

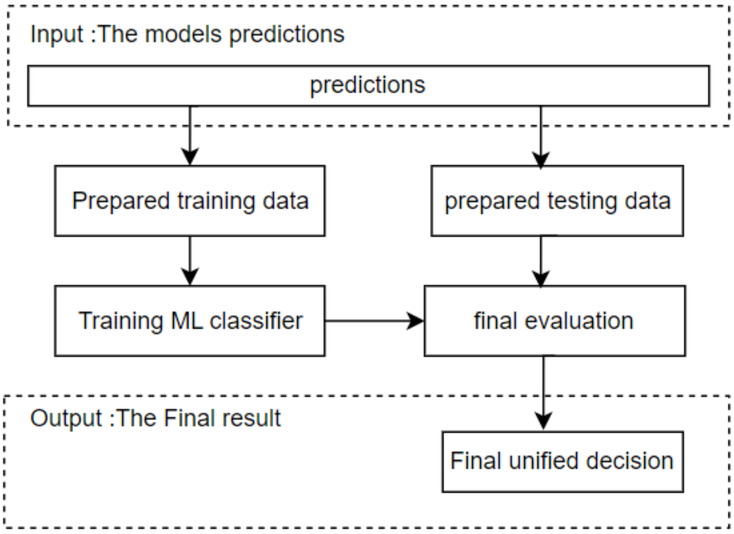


Figure 3.10: *The fusion using the ML classifiers.*

1. Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a type of supervised learning algorithm primarily used for binary classification, which means it outputs one of two possible categories. Despite this, SVMs can be adapted for multi-class problems and regression tasks. SVMs are effective in handling non-linear data. They are versatile tools applicable in various fields, such as pattern recognition, medical diagnosis, and marketing (Urso et al., 2019).

The core principle of SVM is to identify a hyperplane that best separates data points into distinct classes based on their features. Among the numerous potential hyperplanes, the optimal one is selected to differentiate the data into two classes. The position and orientation of this hyperplane are determined by a subset of data points known as support vectors. These support vectors are critical because their removal can alter the hyperplane's position and orientation (K. R. Singh & Dash, 2023).

2. XGBoost

XGBoost, short for Extreme Gradient Boosting, is an advanced ensemble supervised machine learning algorithm created by Tianqi Chen. It can be utilized for both classification and regression tasks. As a variation of gradient boosting, XGBoost distinguishes itself from traditional gradient boosting models in several key ways. Additionally, XGBoost is particularly effective for decision-making processes involving numerous independent variables, as it excels in handling complex datasets and identifying the most significant predictors (Belyadi & Haghighat, 2021).

3. Logistic Regression

Logistic regression is a statistical method designed to assess how multiple independent variables impact a categorical dependent variable, which is limited to a set number of possible values. This technique is especially effective for pinpointing the most significant variables in a dataset with numerous factors, where the outcome is restricted to specific categories, such as yes/no responses or different types of medical conditions (Fraccaro & et al., 2015).

3.6.2 Voting techniques

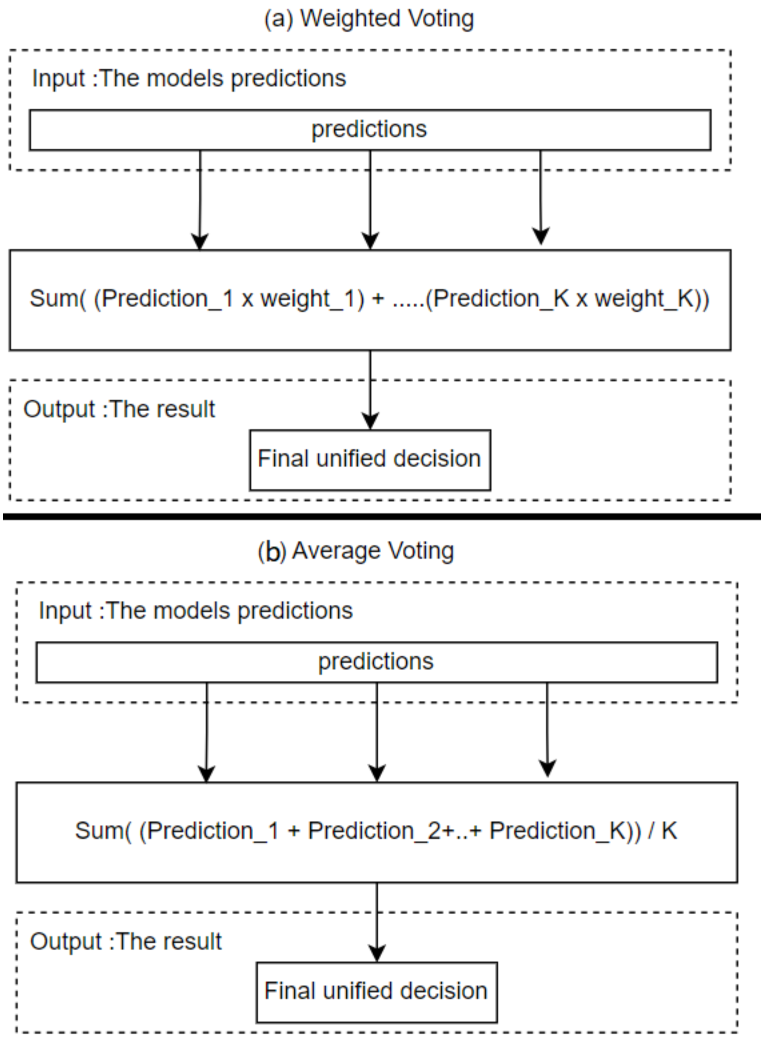


Figure 3.11: *The fusion using the voting techniques.*

In our approach, we utilized two techniques: average voting and weighted voting. Both techniques do not require training and can be applied directly to the predictions of the models. the voting processes is illustrated in the Figure 3.11. the two voting technique are explained as follow (Ganaie et al., 2022):

1. Average voting

Average voting involves taking the arithmetic mean of the predictions from multiple models. This method leverages the consensus among models to produce a final prediction, reducing the impact of any single model's errors.

2. Weighted voting

Weighted voting assigns different weights to the predictions of each model based on their performance. Models with higher accuracy are given more influence in the final prediction, ensuring that stronger models have a greater impact on the outcome.

3.7 Conclusion

Melanoma, the rarest but most deadly skin cancer, is hard to spot early because it shows few clear signs. But catching it fast is key for patience life saving. Convolutional Neural Networks (CNNs) have changed the situation by getting better at pulling out details from pictures of skin lesions, helping with more accurate picks. Yet, we face a big hurdle: there's a lot more data on normal skin than on melanoma, leading to a bias in our learning systems that can mess up their performance.

In this chapter, we tackle these problems by presenting a dataset balancing technique,then we explaining the architecture of our melanoma detection system,explaining each phase of the system, and the use of pre-trained models for feature learning. By addressing these aspects, we aim to improve the classification and diagnosis of melanoma.

Chapter 4

Implementation and realization

4.1 Introduction

This chapter details the implementation of a melanoma detection system based on deep learning and convolutional neural networks (CNN). The aim is to improve early diagnosis of melanoma through accurate analysis of dermoscopic images.

The chapter describes each stage of the process, from the tools and environments used to the development of the user interface. It also presents the data pre-processing techniques, the selection and adjustment of the model's hyperparameters, the experiments carried out and an in-depth analysis of the system's results and limitations.

4.2 Environment and tools

4.2.1 Programming language

Python :

It is a high-level scripting programming language. It has been around for 33 years since it was made by Guido van Rossum and came out in 1991.

Python is a very popular language because it supports various programming paradigms such as procedural, object-oriented and functional programming styles and has a simple syntax that is easy to read and great for rapid application development, it has a rich library and many add-on packages making it suitable for a wide range of applications, including data analysis, and artificial intelligence tasks such as image processing (Python Software Foundation, 2024).

4.2.2 Environment

Kaggle :

It is an online platform for data science competitions that was founded in April 2010, now it has grown to become a community of data scientists and machine learning practitioners.

Kaggle offers a range of resources, including the discovery and the uploading of datasets and pretrained models, alongside the programming environment, giving the programmers the opportunity to construct their own models and participate in competitions ("Kaggle Documentation", 2024).

4.2.3 Model construction tools

Tensorflow :

TensorFlow (tf), is a scientific computing library developed by Google specifically for deep learning workloads, It executes all operations using tensor objects, with complicated neural network algorithms consisting primarily of fundamental tensor operations such as multiplication and addition, which facilitates the construction of artificial intelligence (AI) models even by novice Python users (TensorFlow, 2024).

It has a wide range of applications, including handwritten digit identification, word embeddings, facial recognition and classification tasks (TensorFlow, 2024).

Keras :

Keras is an open-source neural network computing framework, sometimes known as a deep learning API, containing Python frontend and backend interfaces. It can be run on top of frameworks such as TensorFlow or PyTorch ("About Keras", 2024).

It is built as a highly modular and extendable high-level neural network interface, making it a powerful and adaptable API (utilized in industrial-strength applications such as NASA) allowing programmers to easely create and train models (Long & Zeng, 2022).

Keras is regarded as the second most popular deep learning framework, due to its high level of abstraction and ease of use ("About Keras", 2024).

4.2.4 Pre-processing tools

Pandas :

Pandas is a quick, powerful, flexible, and user-friendly open-source data analysis and manipulation framework, built on the Python computing language ("pandas", 2024). It offers an in-memory NoSQL database that includes SQL-like syntax, rudimentary statistical and analytic tools, and graphing features (Harrison, 2016).

Pandas is commonly used to replace Excel, execute ETL tasks, process tabular data, load CSV or JSON files, and so on (Harrison, 2016).

Numpy :

NumPy is the abbreviation for "Numerical Python," known as a Python library and the fundamental package for scientific computing that was created in 2005 by Travis Oliphant (NumPy Developers, 2023).

It offers a multidimensional array objects and is intended to speed up array operations like as manipulation, sorting, selecting, basic linear algebra, and basic statistical calculations, among others, according to its documentation (NumPy Developers, 2023).

OpenCv :

OpenCV stands for "Open Source Computer Vision Framework." It's an open-source library based on Python, created to offer a common infrastructure for computer vision applications and to accelerate the use of machine perception in commercial products (OpenCV Developers, 2024).

With its 2,500 optimized algorithms, OpenCV facilitates several computer vision tasks: face recognition, object identification, human actions classification in videos, and tracking tasks such as camera and object movements (OpenCV Developers, 2024).

SkLearn :

Scikit-learn is an abbreviation for "SciPy Toolkit-Learn." It is a free, open-source machine learning library that is part of the SciPy ecosystem, designed for scientific and technical computing in Python (scikit-learn Developers, 2024).

Developed by David Cournapeau, scikit-learn aims to be simple and efficient for predictive data analysis tasks (scikit-learn Developers, 2024).

4.2.5 Plotting tools

MatPlotLib :

Matplotlib is a Python data visualization package developed by Michael Droettboom et al. and first released in 2003. It features an object-oriented API to construct high-quality, publishable plots and graphs (Matplotlib Developers, 2024).

This framework can handle a variety of plot types, such as line graphs, scatter plots, bar charts, histograms, and more. It also supports customizing visual styles, layout, and exporting to a variety of file types. Matplotlib is extensively used in the scientific computing and data analysis fields, and it is simple to integrate with Jupyter notebooks and graphical user interfaces (Matplotlib Developers, 2024).

4.2.6 Performance evaluation tools

In medical applications, particularly in binary classification tasks such as melanoma detection, overall accuracy and error rate are insufficient for a full evaluation. Our study focuses on evaluating the behavior of our CNN based classifiers using a variety of performance indicators to acquire a better understanding of our system's efficacy, (Baratloo et al., 2015):

Confusion matrix

A confusion matrix (as shown in table 4.1) , used to evaluates the model performance by comparing true labels (melanoma or not melanoma) and predicted labels. The intersections between these are: TP: True Positive (melanoma predicted as melanoma), TN: True Negative (not melanoma predicted as not melanoma), FP: False Positive (not melanoma predicted as melanoma), and FN: False Negative (melanoma predicted as not melanoma).

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 4.1: *Confusion Matrix*

Accuracy

As defined by the equation 4.1, it measures the proportion of true results (both true positives and true negatives) among the total number of cases examined.

In melanoma detection, it represents the proportion of correctly identified melanoma and non-melanoma cases out of all samples.It can be misleading in imbalanced datasets where non-melanoma cases are much more prevalent.

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

Precision

Also known as Positive Predictive Value, It refers to the fraction of accurately recognized melanoma cases among all cases predicted as melanoma.

Defined by the equation 4.2:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.2}$$

Recall

Also called as Sensitivity or True Positive Rate, quantifies the fraction of true positive instances properly detected by the model. In melanoma detection, recall reveals how many real melanoma cases were accurately recognized. It defined as 4.3:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.3}$$

F1-score

It is also occasionally referred to as the F-score or F-measure. The F1 score is established as a weighted average of precision and recall. Its most prominent and poorest values are 1 and 0, respectively.

the F1-score (Eq :4.4), Ensures that the model not only identifies as many melanoma cases as possible (high recall) but also minimizes the number of false positives (high precision).

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.4}$$

ROC curve

It Refers to the Area Under the Receiver Operating Characteristic (ROC) Curve, which is a graphical representation of the true positive rate (recall) against the false positive rate.

In melanoma detection, it provides an aggregate measure of the model's performance across all classification thresholds, ensuring the model maintains high true positive rates while keeping false positive rates low. The ROC curve itself is plotted with:

$$y\text{-axis: True Positive Rate (Recall)}, \quad x\text{-axis: False Positive Rate (FPR} = \frac{FP}{FP+TN}) \tag{4.5}$$

4.3 DataSet preparation

4.3.1 Used dataset description

The ISIC 2019 dataset was made available in 2019, This dataset contains a total of 33,569 images—25,331 training images and 8,238 test images,provided by researchers.

Similar to ISIC 2018, they offered ground truth data for the training set only, identifying the classes of dermatofibroma (DF), vascular lesions (VASC), actinic keratosis (AK), benign keratosis (BK), melanoma (MEL), and melanocytic nevus , in addition to basal cell carcinoma (BCC) and (SCC), as shown in the figure 4.1.

A notable feature of the ISIC 2019 dataset is the presence of multiplets of single lesions, which are variations of the same lesion at various zoom levels. These variances might offer distinctive features that are significant at certain magnification settings.

Both training and testing sets include patient metadata. For the training set, metadata indicates the patient's approximate age, anatomical site, lesion ID, and gender. Lesion ID is specified for 23,247 images and unspecified for 2,084 images, with 11,848 unique IDs from a total of 25,331 images. The testing set consists of nine classes: eight classes as in the training set, plus an additional unknown class. The testing metadata indicates the patient's approximate age, anatomical site, and gender.

The ISIC2019 does not release the ground truth for the testing set, so we used the training set as our Dataset, which includes the ground truth, for our analysis. This dataset comprises 25,331 lesion images, and the class distribution is shown in the table 4.2. This approach ensures that we have labeled data for accurate evaluation.

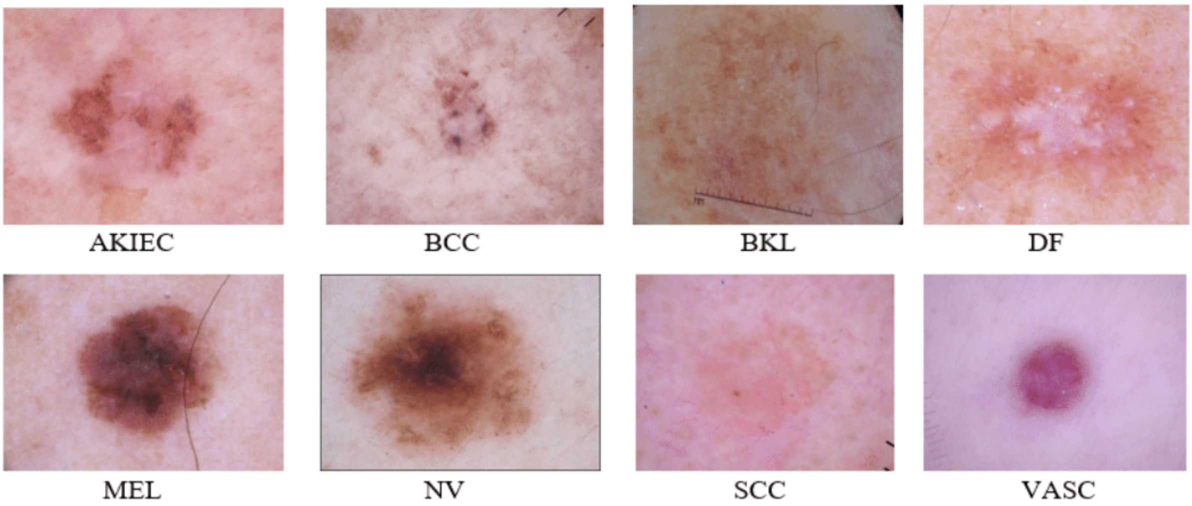


Figure 4.1: *Samples from the ISIC2019 skin cancer dataset.*

4.3.2 Dataset restructuring

In order to adapt the ISIC database, initially composed of eight classes, to our binary classification task (melanoma vs non-melanoma), we restructured it by grouping the seven remaining classes into a single category called "non-melanoma". This reorganisation simplified the classification and better matched the requirements of our machine learning model. As shown in Table 4.2, the restructured database now consists of two classes:

- Melanoma: 4,522 images
- Non-melanoma: 20,809 images

This new breakdown makes it easier to interpret the results and validate the performance of the melanoma detection system.

Class	ISIC2019	Binary Classification	Total
Melanoma	4522	Melanoma	4522
Melanocytic nevus	12,875	Non-Melanoma	20,809
Basal cell carcinoma	3323		
Actinic keratosis	867		
Benign keratosis	2624		
Dermatofibroma	239		
Vascular lesion	253		
Squamous cell carcinoma	628		

Table 4.2: *ISIC2019 Classes Distribution and Our Binary Classification Distribution*

4.3.3 Applying the proposed balancing technique

The ISIC2019 skin cancer dataset, initially consisting of 25,331 dermoscopic images (see Table 4.2), was pre-processed to satisfy the requirements of binary melanoma classification. The images were reorganised and labelled, by assigning a value of 1 to melanoma images (4,522 images) and a value of 0 to non-melanoma images (20,809 images). The distribution of classes after reorganisation is shown in figure 4.2, highlighting a significant imbalance between the majority (non-melanoma) and minority (melanoma) classes.

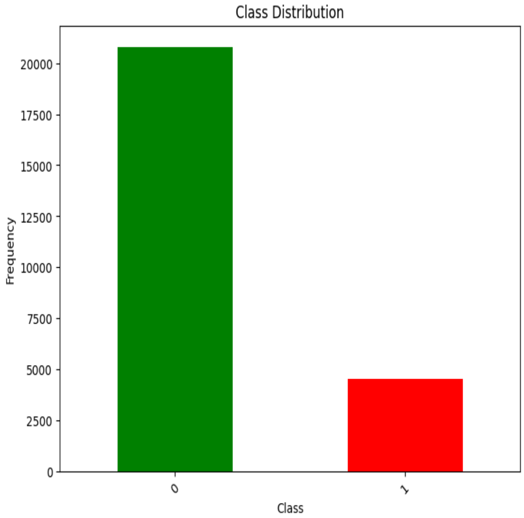


Figure 4.2: *The Classes distribution in our binary classification task (class 0 on the left) illustrates the non-melanoma, (class 1 on the right) illustrates the melanoma.*

To overcome this imbalance, a class balancing technique was applied, generating five balanced subsets. These subsets are presented in Figure 4.3 and their distribution in terms of samples is detailed in Table 4.3.

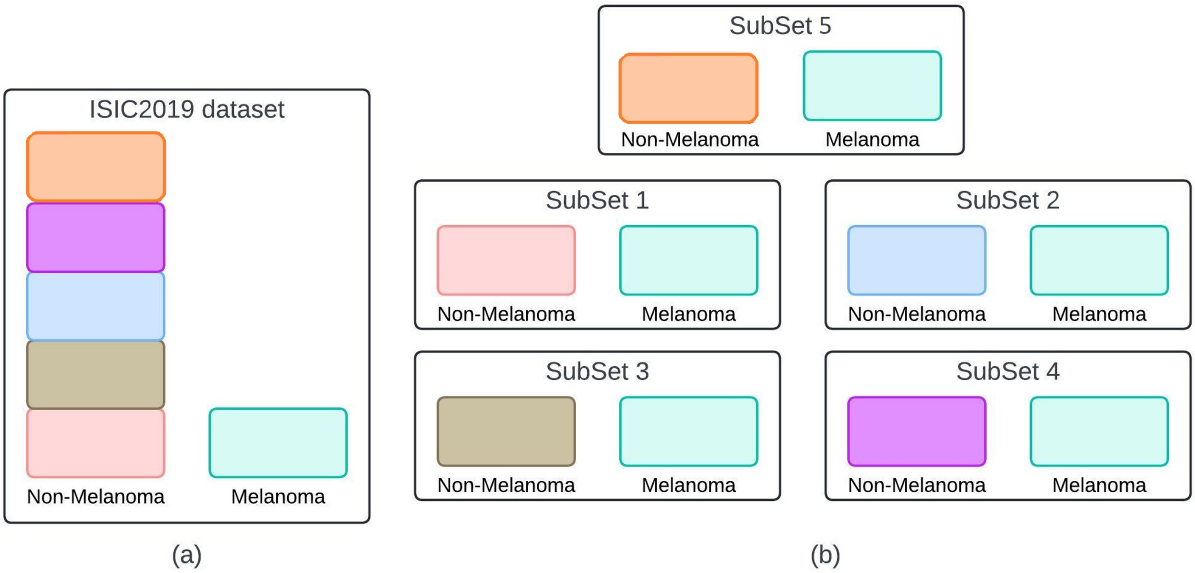


Figure 4.3: *The Subsets extraction process: (a) illustrates the ISIC2019 dataset with the imbalance issue, (b) illustrates the subsets extracted after undersampling.*

Subsets	**Number of Samples**		**Total**
	Melanoma	**Non-Melanoma**	
subset 1	4522	4162	**8684**
subset 2		4162	**8684**
subset 3		4162	**8684**
subset 4		4162	**8684**
subset 5		4161	**8683**
Total	**4522**	**20809**	**25331**

Table 4.3: *Distribution of Melanoma and Non-Melanoma Samples.*

4.3.4 SubSets spliting

In order to evaluate the performance of our binary melanoma classification model, we divided each data subset (melanoma and non-melanoma) into training, validation and test sets. Data allocation followed the 70/20/10 rule, where 70% of the data was allocated to model training, 20% to validation and 10% to testing. Specifically, 90% of each subset was devoted to training, from which 20% was then extracted to form the validation set. This careful division of data ensured a rigorous evaluation of the model, making sure that it did not over-learn the training data and that it generalised well to the unknown data. The data splitting process is illustrated in Figure 4.4.

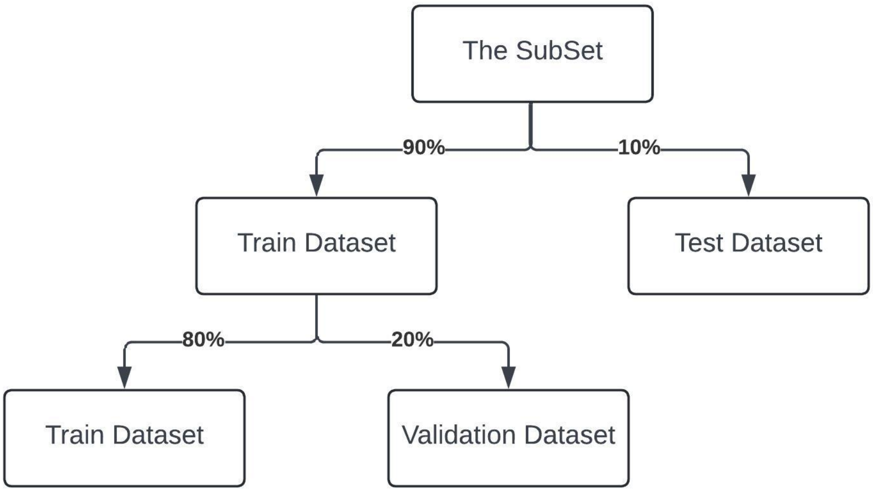


Figure 4.4: *The Subsets Spliting Distribution.*

For each subset, which has 8,684 samples, the distribution of the train, validation, and test datasets will be as shown in the following Table 4.4. Note that within each class (melanoma and non-melanoma), the samples are equally distributed across the train, validation, and test datasets.

	Train Dataset (70%)	Validation DataSet (20%)	Test Dataset (10%)
Sample	6252	1563	869

Table 4.4: *Distribution of Train, Validation, and Test Datasets for Each Subset*

4.3.5 data preprocessing

In our work, we adjusted the images to a size of 224x224 pixels, following the specifications of the ResNet50 model that had been pre trained. Ensuring alignment, with the models expected processing steps we standardized the images to fall within the [0, 1] range. By utilizing Keras frameworks $preprocess_input$ function each image underwent preparation to fulfill ResNet50s input criteria. This function upholds the color consistency in the ImageNet dataset which's essential for utilizing the pre existing weights. The pre processing phase results in a float type array where RGB image data undergoes conversion, to BGR format, where each color channel was calculated by subtracting mean of all pixels present in ImageNet dataset without dividing ("resnet.preprocess_input", 2024). an example is presented in the Figure 4.5.

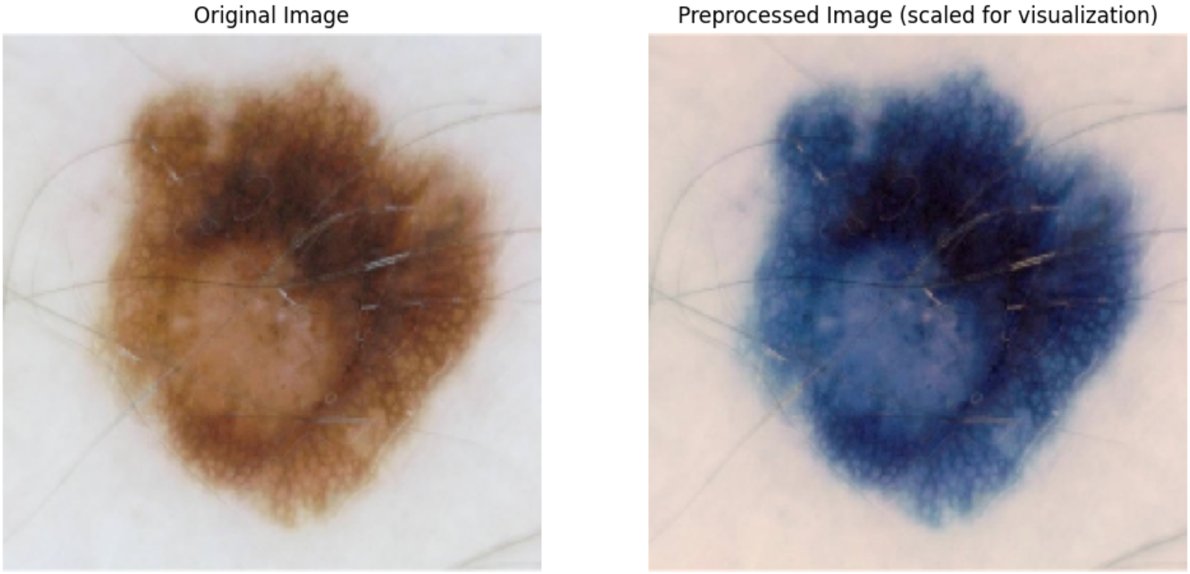


Figure 4.5: *preprocess input function before and after.*

4.3.6 Data augmentation

In order to enrich the training dataset and improve the generalisability of the model, we exploited the ImageDataGenerator class of the TensorFlow Keras library to apply data augmentation transformations. Details of the transformations applied are shown in Table 4.5.

Augmentation parameters	**Range**
Rotation range	20
Width shift range	0.2
Height shift range	0.2
Shear range	0.2
Zoom range	0.2
Horizontal flip	True
Fill_mode	nearest
Rescale	1./255

Table 4.5: *Image Augmentation Parameters and Their Ranges*

It is important to note that these transformations were only applied during the model training phase. This application has several advantages:

- Improved model generalisation.
- Prevention of overlearning.
- Management of input image variability.

4.4 Training phase

4.4.1 Model Hyper-parameters

There are no ideal hyperparameters available for direct application in deep learning. The selection of hyperparameters is critical since it has a major impact on model performance and convergence. Because of its relevance, we used a hyperparameter optimizer to automatically investigate a wide range of hyperparameter alternatives.

The Optuna framework (Akiba et al., 2019), was specifically used to allow for an efficient and methodical exploration of the hyperparameter space, ultimately improving the model's performance and generalization potential.

Parameters

We optimized certain hyperparameters (the hyper-parameters of the top layers) using Optuna, while keeping others fixed (the base model) for stability and consistency. This approach allows us to focus on optimizing key parameters without overwhelming the model. The table 4.6 summarizes the hyperparameters optimized and the fixed parameters used in our experiments.

Parameter	**Type**	**Range / Value**
Hyperparameters Optimized by Optuna		
Filter Size	Integer	[3, 5]
Dropout Rate	Float	[0.0, 0.4]
L1 Regularization Value	Log-uniform	[1e-5, 1e-2]
L2 Regularization Value	Log-uniform	[1e-5, 1e-2]
Number of Kernels	Categorical	{64, 128, 256}
Fixed Parameters		
Base Model	Fixed	ResNet50 (pre-trained on ImageNet)
Input Shape	Fixed	(224, 224, 3)
Optimizer	Fixed	Adam
Learning Rate	Fixed	0.001
Loss Function	Fixed	Binary Crossentropy
Metrics	Fixed	Accuracy
Epochs	Fixed	20
Callbacks	Fixed	EarlyStopping, ModelCheckpoint

Table 4.6: *Summary of Hyperparameters and Fixed Parameters for Optuna Optimization*

We applied hyper-parameter tuning on one of the subsets, with 10 trials and 20 epochs for each trial. The entire tuning process took around 2 hours and 30 minutes, with each trial taking approximately 20 minutes. The results are presented in tables 4.7 and 4.8, where it can be seen that the best hyperparameters are observed in trial 6.

Metric	**Trial 0**	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**
Value	0.64	0.61	0.61	0.68	0.59
Dropout Rate	0.21	0.39	0.16	0.36	0.16
Filter Size	5	3	4	5	3
L1 Value	0.0001	0.0024	0.0048	0.0004	0.0000
L2 Value	0.0059	0.0001	0.0022	0.0043	0.0014
Num Kernel	64	256	256	128	128
Precision	0.62	0.71	0.78	0.58	0.65
Recall	0.89	0.64	0.48	0.93	0.82
F1 Score	0.73	0.67	0.59	0.72	0.73
Accuracy	0.65	0.68	0.66	0.62	0.68

Table 4.7: *Summary of Optuna Trials and their Results (Trials 0 to 4)*

Metric	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9
Val Loss	0.60	**0.47**	0.59	0.62	0.62
Dropout Rate	0.15	**0.00**	0.15	0.26	0.20
Filter Size	3	**3**	4	5	4
L1 Value	0.001	**0.01**	0.0001	0.0030	0.0048
L2 Value	0.0003	**0.01**	0.0026	0.0000	0.0068
Num Kernel	256	**256**	64	256	64
Precision	0.75	**0.79**	0.69	0.79	0.70
Recall	0.58	**0.77**	0.79	0.49	0.71
F1 Score	0.66	**0.78**	0.73	0.61	0.70
Accuracy	0.68	**0.78**	0.70	0.67	0.69

Table 4.8: *Summary of Optuna Trials and their Results (Trials 5 to 9)*

Table 4.9, which details the hyperparameters of the model refined by our fine-tuning process, giving a complete overview of its optimised configuration.

Parameter	Type	Range / Value
Hyperparameters Optimized by Optuna		
Filter Size	Integer	3
L1 Regularization Value	Log-uniform	0.01
L2 Regularization Value	Log-uniform	0.01
Number of Kernels	Categorical	256
Fixed Parameters		
Base Model	Fixed	ResNet50 (pre-trained on ImageNet)
Input Shape	Fixed	(224, 224, 3)
Optimizer	Fixed	Adam
Learning Rate	Fixed	0.001
Loss Function	Fixed	Binary Crossentropy
Metrics	Fixed	Accuracy
Epochs	Fixed	20
Callbacks	Fixed	EarlyStopping, ModelCheckpoint

Table 4.9: *results hyperparameters for the resnet50 preptrained model.*

4.4.2 Results and discussions

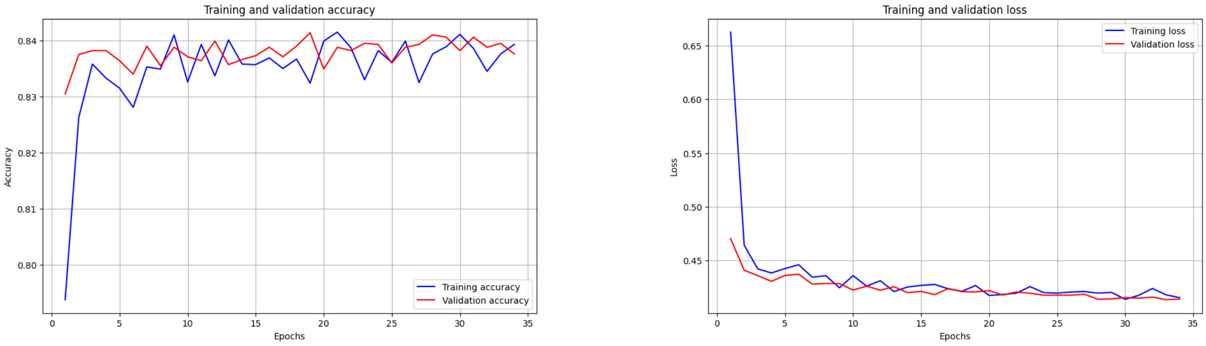
Experiment 1: Basic performance evaluation with ResNet50

In order to establish a baseline performance for melanoma detection, we use a pre-trained ResNet50 model with custom top layers, which we apply to the ISIC2019 dataset without prior modifications. Table 4.10 summarises the main training parameters used, including details of the dataset, preprocessing methods. training took approximately 3 hours.

Setting	**Details**
Dataset	ISIC2019
Data Augmentation	None
Normalization	Rescale
Class Imbalance	Highly Imbalanced (2 classes)
Epochs	34
Batch Size	32
Dataset Split	Train: 70%, Validation: 20%, Test: 10%
Training Images	18237
Validation Images	4560
Test Images	2534

Table 4.10: *Training Settings for the First Experiment*

The validation loss and precision curves are shown in Figure 4.6, while the Confusion Matrix is shown in Figure 4.7, and finally, Table 4.11 details the following performance measures: Precision, recall, F1 score and accuracy.



(a) *training and validation accuracy graph* ***(b)*** *Training and validation loss graph*

Figure 4.6: *The training curves*

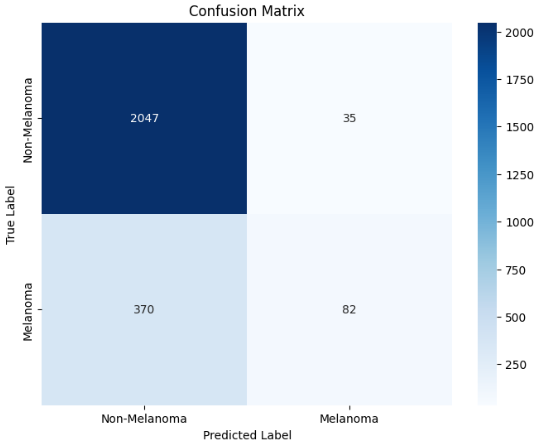


Figure 4.7: *The confusion matrix on the test datasets*

Metric	**Value**
Validation Loss	0.4185
Validation Accuracy	0.8402
Validation AUC	0.7721
Precision	0.7009
Recall	0.1814
F1 Score	0.2882
Overall Accuracy	0.8402

Table 4.11: *Results metrics of the First Experiment*

The training and validation curves show a promising performance of the model on the validation data. The training and validation accuracies (Figure 4.6(a)) reach around 84% respectively and stabilise after around 10 epochs. This indicates that the model converges rapidly and generalises well on the validation data. The loss curve (Figure 4.6(b)) follows a rapid downward trajectory initially before stabilising at around 0.4, suggesting that the model is effectively minimising error during training. The ROC curve shows an area under the curve (AUC) of 0.7721, indicating good discriminative ability of the model between melanoma and non-melanoma classes.

However, the confusion matrix (Figure 4.7(b)) reveals that the model correctly identifies 2047 non-melanoma images and 82 melanoma images, but it also commits 35 false positives and 370 false negatives. This imbalance suggests that the model is biased towards the majority class (non-melanoma), resulting in low recall for the minority class (melanoma).

The overall performance metrics reported in Table 4.10 highlight a good generalisation of the model. The AUC of 0.7721 confirms the discriminating ability observed in the ROC curve. However, precision (0.7009) and recall (0.1814) reveal a significant disparity in the model's performance between classes. This low recall indicates that only 18.14% of melanomas are correctly identified. These results highlight the need to improve the model

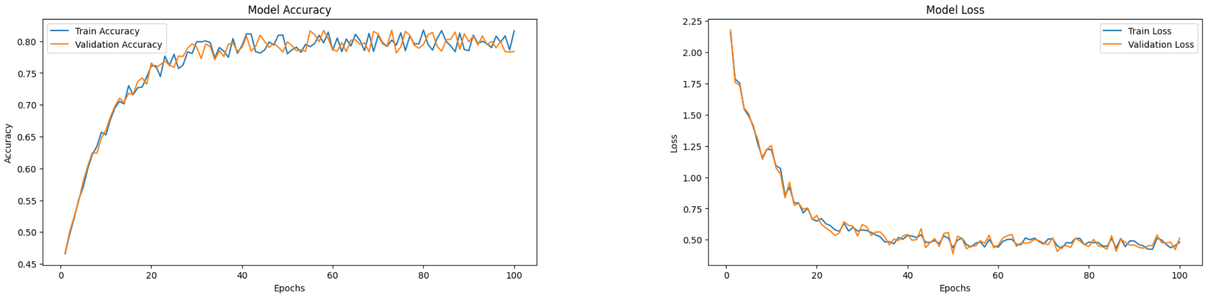
Experiment 2: Performance evaluation with ResNet50 and the proposed balancing technique

As explained in Chapter 3, we propose to reorganise the dataset e into several subsets in order to address the problem of data set imbalance. For this purpose, the ISIC 2019 dataset is restructured into 5 subsets, each of which is used to train its corresponding deep learning model (Resnet50). The training parameters are the same for all subsets and are shown in Table 4.12 below.

Training Detail	**Description**
Number of Subsets	5
Model Architecture	ResNet-50 (5 models)
Data Augmentation	Applied to training dataset
Normalization	Applied to all data
Epochs	100 epochs
Early Stopping	Patience of 20 epochs
Checkpointing	Save the best model
Metrics	Validation loss
Training Time per Model	Approximately 4 hours
Total Training Time	Approximately 20 hours
Batch Size	32
Data Split	70% training, 20% validation, 10% test

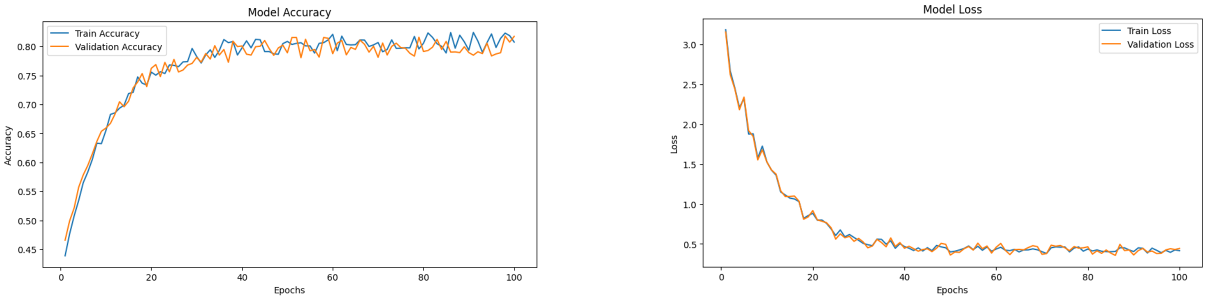
Table 4.12: *Training Details Summary*

Figures 4.8 to 4.12 show the training curves for each model respectively, while Figures 4.13 to 4.15 show the ROC curves and confusion matrices for each model. Finally, the evaluation measures for all models are summarised in Table 4.13.



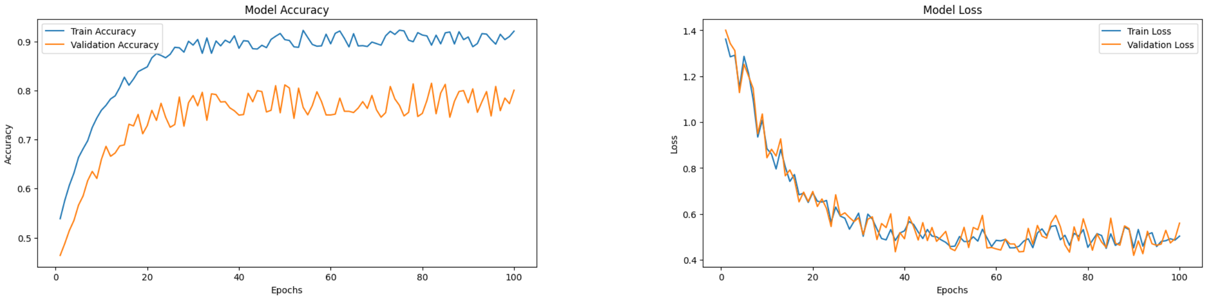
(a) *training and validation accuracy graph* ***(b)*** *Training and validation loss graph*

Figure 4.8: *The training curves of the first model (100 epoch)*



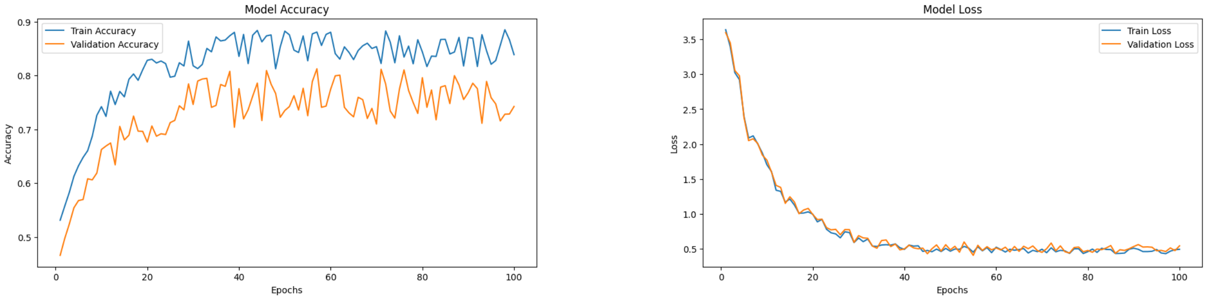
(a) *training and validation accuracy graph* ***(b)*** *Training and validation loss graph*

Figure 4.9: *The training curves of the second model (100 epoch)*



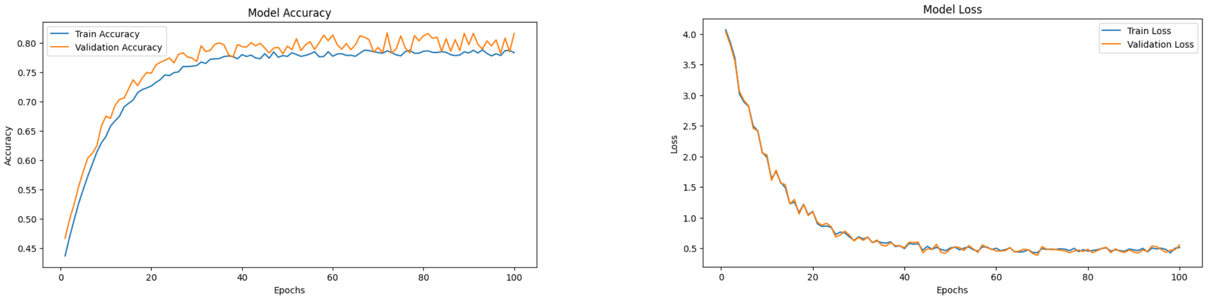
(a) *training and validation accuracy graph* ***(b)*** *Training and validation loss graph*

Figure 4.10: *The training curves of the third model (100 epoch)*



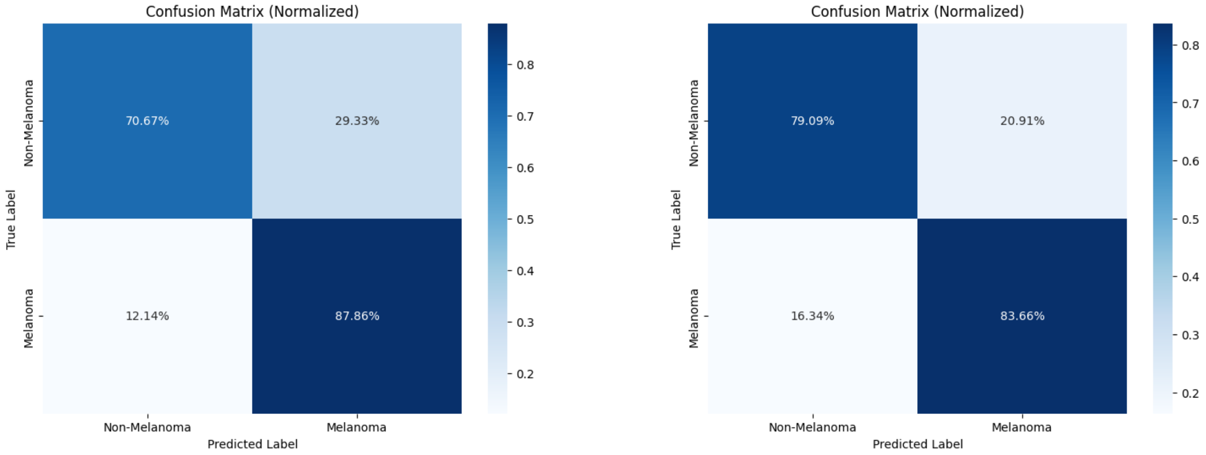
(a) *training and validation accuracy graph* ***(b)*** *Training and validation loss graph*

Figure 4.11: *The training curves of the third model (100 epoch)*



(a) *training and validation accuracy graph* ***(b)*** *Training and validation loss graph*

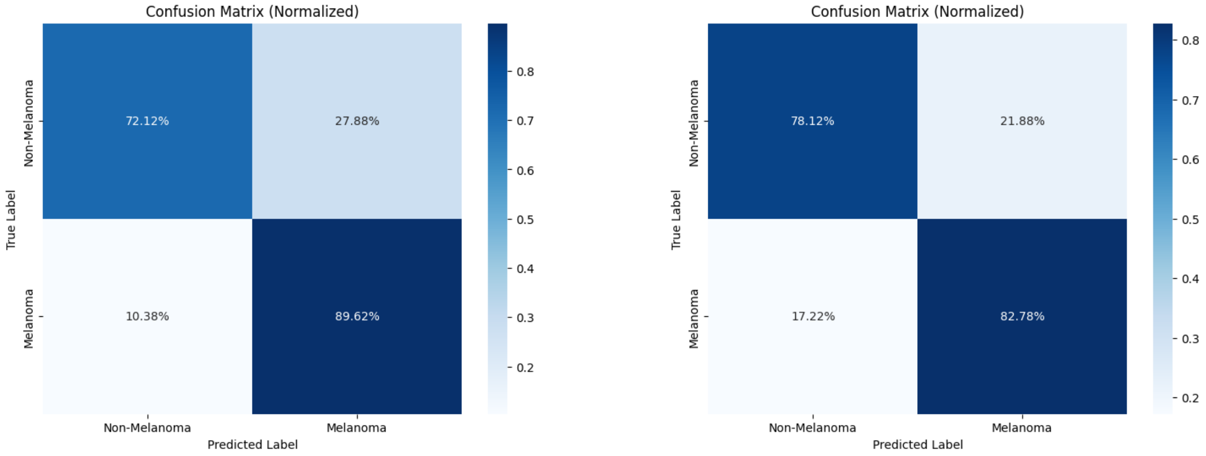
Figure 4.12: *The training curves of the fifth model (100 epoch)*



(a) The confusion matrix of the first model

(b) The confusion matrix of the second model

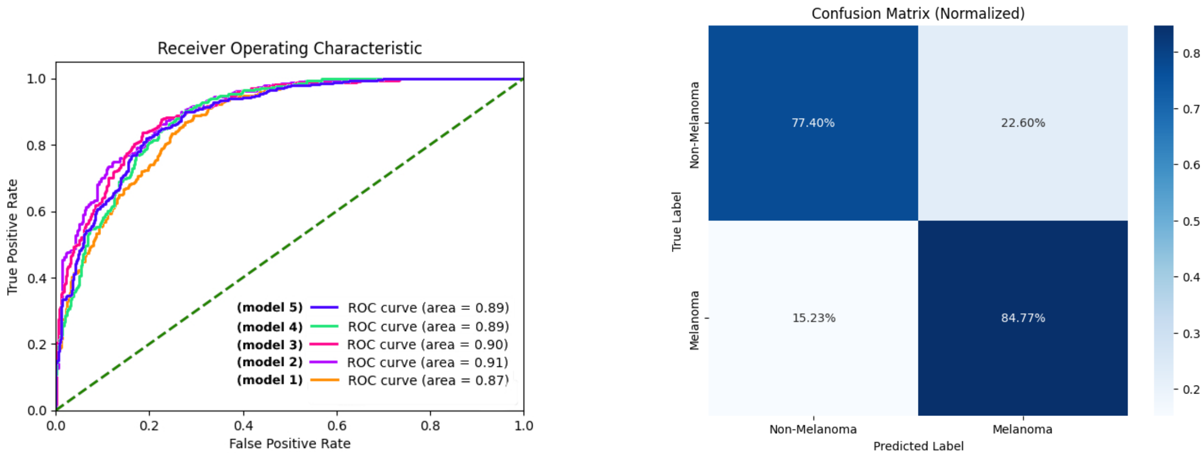
Figure 4.13: *The confusion matrix of the first and the second models*



(a) The confusion matrix of the third model

(b) The confusion matrix of the fourth model

Figure 4.14: *The confusion matrix of the third and the fourth models*



(a) *The ROC graph* ***(b)*** *The confusion matrix fifth model*

Figure 4.15: *The evaluation ROC curve and confusion matrix of the fifth model*

Metric	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
Loss	49.15	42.95	44.70	45.95	45.20
Accuracy (%)	79.63	81.47	81.24	80.55	81.24
AUC (%)	87.42	90.63	90.04	88.65	88.86
Precision (%)	76.54	81.33	77.78	80.47	80.33
Recall (%)	87.86	83.66	89.62	82.78	84.77
F1 Score (%)	81.81	82.48	83.28	81.61	82.49

Table 4.13: *Evaluation Metrics of Five Models*

Each of the five models (trained on subsets of the ISIC 2019 dataset) shows promising training and validation performance. The accuracy and loss graphs for the models (Figures 4.8 to 4.12) reveal consistent trends. The accuracies range from 79.63% to 81.47%, where all models' training and validation accuracies stabilize after around 10 epochs. The accuracy curves give good results, along with the loss curves, which decrease rapidly during the initial epochs and stabilize afterward, with results ranging from 42.95% to 49.15%.

The ROC curves and AUC values shown in Figure 4.15 provide an assessment of the discriminatory abilities of the models; all models show good discriminatory abilities between the melanoma and non-melanoma classes. Although model 2 stands out with an AUC of 0.9063, indicating an excellent ability to distinguish melanoma from non-melanoma. The other models had AUCs ranging from 0.8742 to 0.8911, also reflecting good performance. A high AUC demonstrates an increased ability of the model to avoid false positives and false negatives.

The confusion matrices for each model, shown in figures 4.13 to 4.15 , reveal the distribution of correct and incorrect predictions: Our models performed well in predicting non-melanoma (majority class) and melanoma (minority class). However, false positives (non-melanoma classified as melanoma) and false negatives (melanoma classified as non-melanoma) indicate areas for improvement, particularly in the detection of melanoma.

The proposed balancing technique effectively deals with data imbalance, resulting in improved recall and precision measures compared to the first experiment. Recall, which indicates the sensitivity of the models to melanoma detection, increased from 82.78% to 89.62%. This improvement suggests a reduction in the bias of the models, enabling them to detect a greater proportion of real melanomas.

All trained models were saved with their test dataset for future use and classification.

4.5 Classification phase

To leverage the best of our five models, we have adopted a decision fusion strategy. This strategy aims to combine the strengths of each model to achieve even more accurate melanoma detection performance. The process was as follows:

- Combining the test sets of the models: The test datasets generated by each of the five models were merged to form a new, larger dataset for the final classification phase. This process is illustrated in Table 4.14a.

- Splitting the new dataset: The new combined dataset was then split into separate training and test sets, as shown in Table 4.14b.

- Prediction on the combined test set: Each model was used to predict melanoma probabilities on the combined test set. This process generated five distinct probability predictions, presented visually in Figure 4.16.

Using this decision fusion strategy, we can exploit the strengths of each model and potentially improve the overall accuracy of melanoma detection.

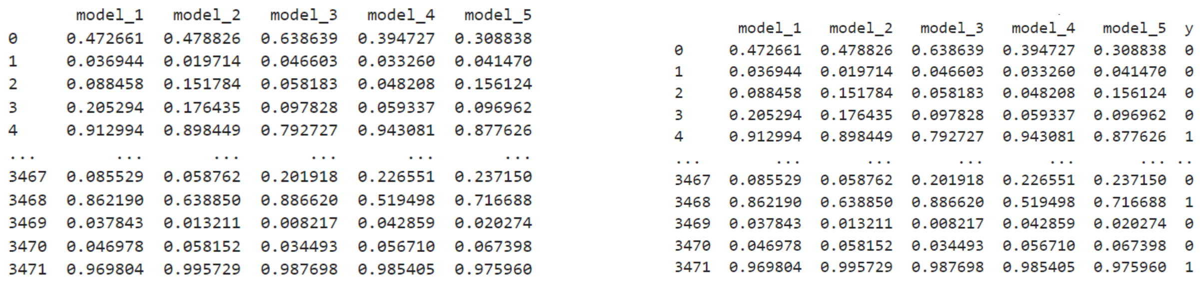
Subset	Number of Samples
Subset 1 Test Set	869
Subset 2 Test Set	869
Subset 3 Test Set	869
Subset 4 Test Set	869
Subset 5 Test Set	869
Total	**4345**

(a) Combined Dataset from the 5 Subsets Test Sets

Split	Number of Samples
Training Set (80%)	3477
Test Set (20%)	868
Total	**4345**

(b) Dataset Split into Training and Test Sets

Table 4.14: Summary of Combined Dataset and its Split



```
      model_1   model_2   model_3   model_4   model_5
0     0.472661  0.478826  0.638639  0.394727  0.308838
1     0.036944  0.019714  0.046603  0.033260  0.041470
2     0.088458  0.151784  0.058183  0.048208  0.156124
3     0.205294  0.176435  0.097828  0.059337  0.096962
4     0.912994  0.898449  0.792727  0.943081  0.877626
...        ...       ...       ...       ...       ...
3467  0.085529  0.058762  0.201918  0.226551  0.237150
3468  0.862190  0.638850  0.886620  0.519498  0.716688
3469  0.037843  0.013211  0.008217  0.042859  0.020274
3470  0.046978  0.058152  0.034493  0.056710  0.067398
3471  0.969804  0.995729  0.987698  0.985405  0.975960
```

(a) Prediction of the train dataset, used for voting methods

```
      model_1   model_2   model_3   model_4   model_5  y
0     0.472661  0.478826  0.638639  0.394727  0.308838  0
1     0.036944  0.019714  0.046603  0.033260  0.041470  0
2     0.088458  0.151784  0.058183  0.048208  0.156124  0
3     0.205294  0.176435  0.097828  0.059337  0.096962  0
4     0.912994  0.898449  0.792727  0.943081  0.877626  1
...        ...       ...       ...       ...       ... ..
3467  0.085529  0.058762  0.201918  0.226551  0.237150  0
3468  0.862190  0.638850  0.886620  0.519498  0.716688  1
3469  0.037843  0.013211  0.008217  0.042859  0.020274  0
3470  0.046978  0.058152  0.034493  0.056710  0.067398  0
3471  0.969804  0.995729  0.987698  0.985405  0.975960  1
```

(b) Prediction of the train dataset along with the true labels as Y column for the training of the classifiers

Figure 4.16: *The Train dataset prediction and their usage in both voting methods and ml classifiers*

4.5.1 Machine learning classifiers

To identify the best-performing fusion approach, we carried out extensive experimentation involving three well-known machine learning classifiers: SVM (Support Vector Machine), XGBoost and Logistic Regression. These classifiers were trained on the combined predictions of the models, with the aim of learning to make a final decision on the basis of this aggregated information. Table 4.15 summarises the training parameters used for each classifier.

Classifier	**Parameter**	**Value**
XGBoost	n_estimators max_depth learning_rate	100 5 0.1
SVM (RBF kernel)	kernel C gamma	rbf 10.0 0.01
Logistic Regression (with polynomial features)	degree (PolynomialFeatures) max_iter (LogisticRegression)	2 10000

Table 4.15: *Parameters of the Classifiers*

The classifier is then evaluated on the test dataset. This evaluation aims to measure its performance in fusing model predictions and making the final decision for melanoma detection. To do this, we used four key performance indicators: precision, accuracy, recall and F-score. The detailed results of this evaluation are presented in Table 4.16.

Metric	**XGBoost**	**SVM**	**Logistic Regression**
Accuracy (%)	86.64	87.10	87.33
Precision (%)	85.18	85.31	86.16
Recall (%)	88.71	89.63	88.94
F1 Score (%)	86.91	87.42	87.53

Table 4.16: *Evaluation Metrics for XGBoost, SVM, and Logistic Regression*

The machine learning classifiers show remarkable enhancement in performance, wherein the accuracy varies between 86.64% to 87.33% where Logistic Regression obtained the highest. Herein, precision values fall within 85.18% to 86.16% with Logistic Regression performing better than other classifiers once again. The recall values lie between 88.71% and 89.63% hence the best return being by SVM. With Logistic Regression leading, F1 Score stands between 86.91% and 87.53% balancing between recall and precision.Overall, the fusion approach enhances accuracy, precision, recall, and F1 Score compared to individual models (accuracy below 82%), resulting in more robust and reliable melanoma detection.

4.5.2 Voting methods

Besides using machine learning classifiers, we explored voting methods to combine the model predictions. Specifically, we implemented average voting and weighted voting techniques. These methods were applied directly to the training set since they do not require additional training.

We computed the mean of the predictions using the NumPy library. This straightforward approach provides an aggregated decision by averaging the outputs of the five models. Weighted voting, on the other hand, involved fine-tuning the weights assigned to each model's predictions to achieve the best performance with Possible Weights Range : 0.0 to 1.0 (step 0.1). The result of the fine-tuning are shown in the table 4.17.

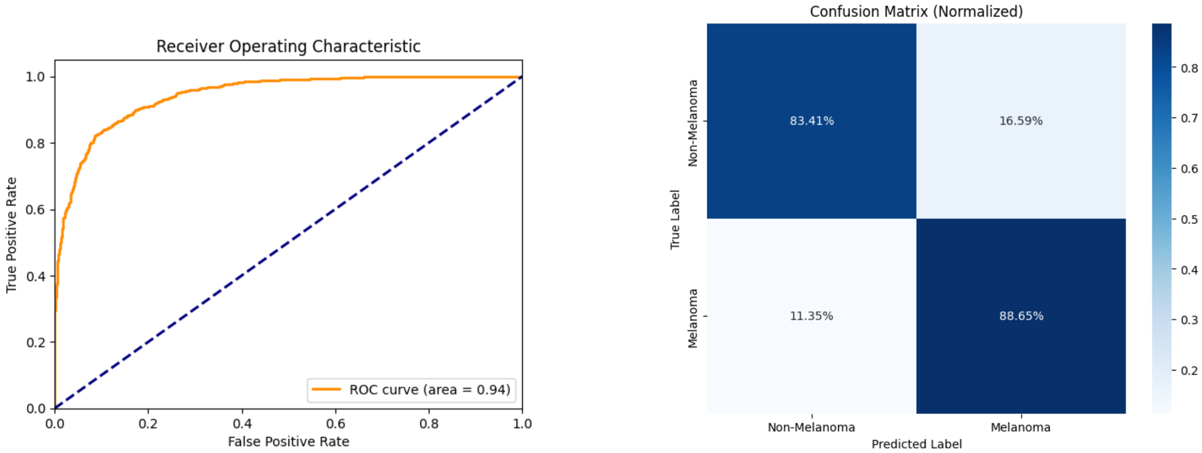
Model	**Weight After Fine-Tuning**
Model 1	0.600
Model 2	0.100
Model 3	0.100
Model 4	0.200
Model 5	0.000

Table 4.17: *Weights After Fine-Tuning*

The results of these voting methods are presented in Figures 4.17 and 4.18 for average voting and weighted voting respectively. The evaluation matrices, shown in Table 4.18.

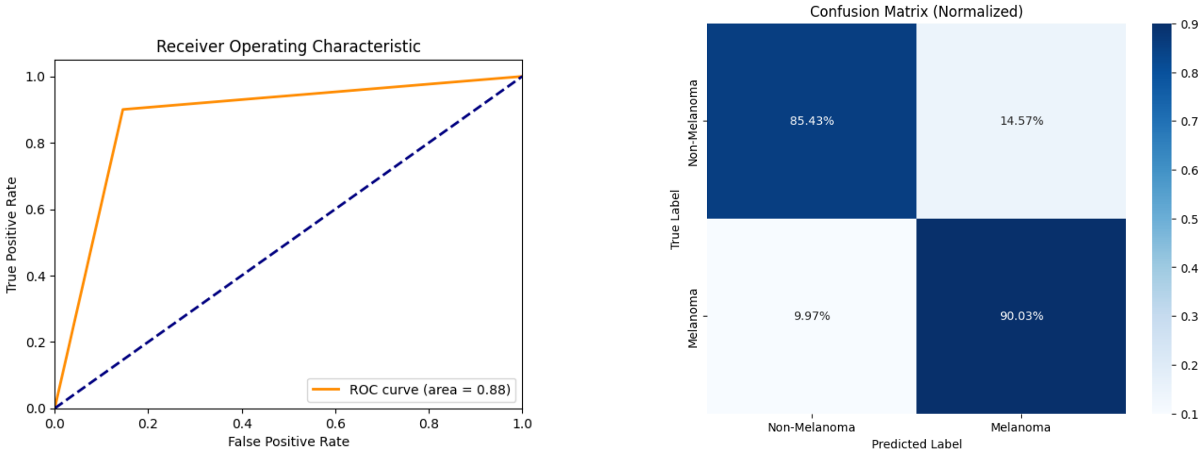
Metric	Weighted Voting (%)	Average Voting (%)
Best Accuracy	87.73	86.03
Precision	86.07	84.24
Recall	90.03	88.65
F1 Score	88.01	86.39

Table 4.18: *Comparison of Weighted Voting and Average Voting Results*



(a) *The ROC graph* ***(b)*** *The confusion matrix*

Figure 4.17: *The evaluation ROC curve and confusion matrix of the average voting*



(a) *The ROC graph* ***(b)*** *The confusion matrix*

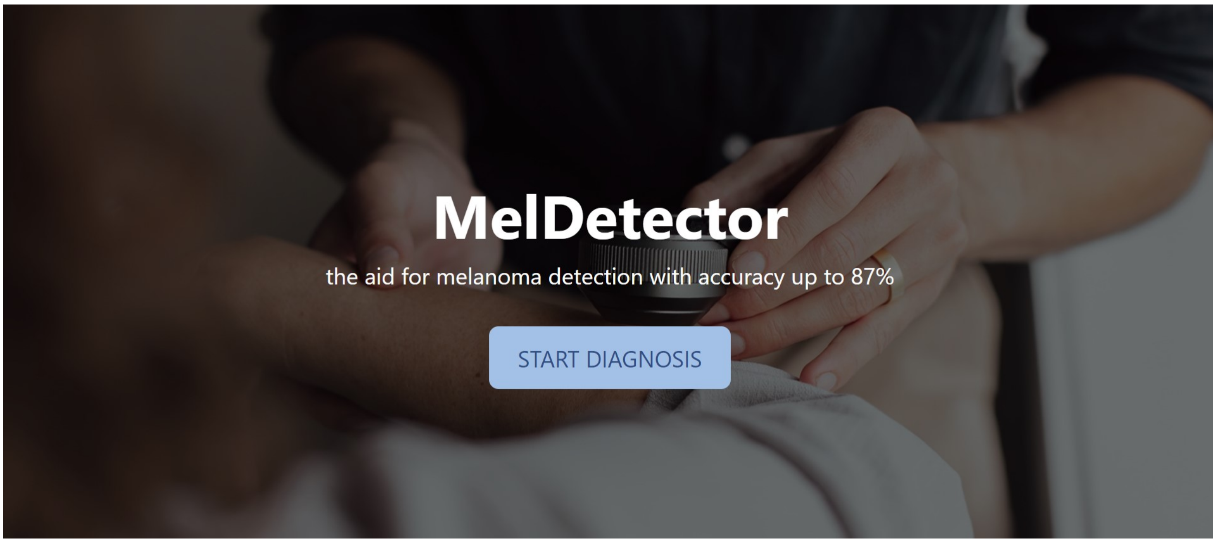
Figure 4.18: *The evaluation ROC curve and confusion matrix of the weighted voting*

Weighted Voting outperforms Average Voting in aspects. It achieves results in terms of accuracy (87.73%, vs. 86.03%) precision (86.07% vs. 84.24%) recall (90.03% vs. 88.65%) and F1 Score (88.01% vs. 86.39%). This suggests that giving weights to individual model predictions enhances the accuracy of the predictions made.

When compared to classifiers like XGBoost, SVM and Logistic Regression Weighted Voting also demonstrates equivalent performance, particularly in recall and F1 Score making it more balanced and robust in its outcomes. Therefore Weighted Voting emerges as the approach, for precise and dependable melanoma detection.

4.6 The system Interface

For user-friendly usability, we provide our melanoma detection system with a simple, friendly interface that can be used to evaluate skin lesions and return the final prediction. The interface has a simple home page that leads the user through the "Start Diagnosis" button to the diagnosis page, as shown in the figure.

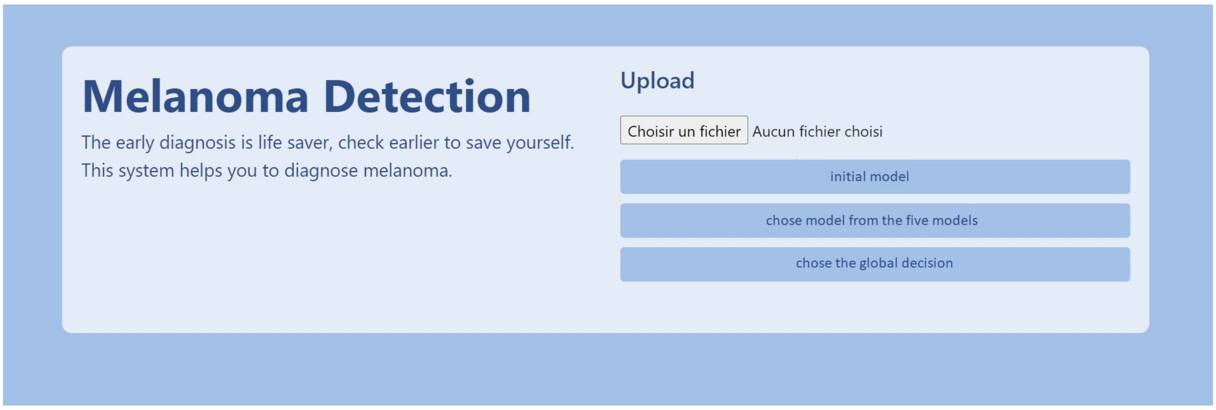


MelDetector

the aid for melanoma detection with accuracy up to 87%

Figure 4.19: The system home page.

After the user reaches the diagnosis page, they can upload the lesion image through the "Upload Image" button, followed by selecting a model to use. The diagnosis page is shown in the figure.



Melanoma Detection

The early diagnosis is life saver, check earlier to save yourself.
This system helps you to diagnose melanoma.

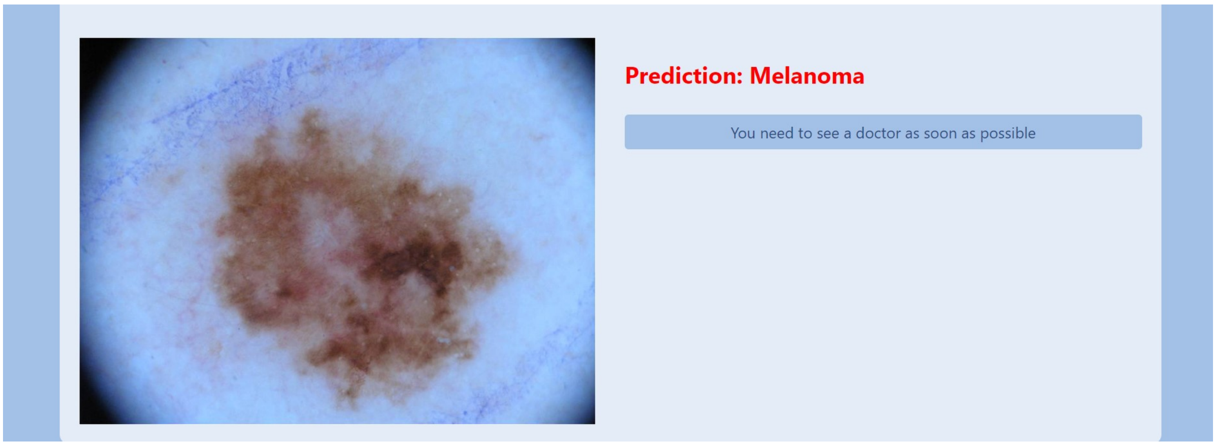
Figure 4.20: The diagnosis page.

After the image upload and model selection, the page will return the diagnosis. The results are shown in the figure 4.22 for a lesion that is not melanoma and in the figure for a lesion that is melanoma 4.22.

Prediction: No Melanoma

You are safe.

__Figure 4.21:__ The results page where the diagnosis indicates Non melanoma.



Prediction: Melanoma

You need to see a doctor as soon as possible

__Figure 4.22:__ The results page where the diagnosis indicates melanoma.

4.7 Final discussion

The comparative analysis of the two experiments highlights the crucial importance of data rebalancing in classification tasks where the classes are unbalanced, such as melanoma detection.

While the first experiment obtained satisfactory results in terms of overall precision, it nevertheless revealed notable weaknesses in terms of recall and melanoma detection. Indeed, the imbalance of classes led to a neglect of actual melanoma cases, which raises concerns for an application as critical as cancer detection. In contrast, the second experiment, using balanced subsets, overcame these limitations. In addition to improving recall and F1 scores, it also maintained high precision and discriminative ability. This result demonstrates the effectiveness of rebalancing the data to compensate for the biases introduced by the unequal distribution of classes.

On the other hand, a comparison with published works dealing with melanoma detection reveals that our approach is a very respectable competitor, as shown by the results in table 4.19. Our system achieves comparable or even superior performance to some recent state-of-the-art methods, while proving simpler to implement and more robust in the face of data variations. These results confirm the great potential of our approach for detecting melanoma from dermoscopic images.

Study	**ACC**
(Mehra et al., 2021)	84.87%
(Barbosa et al., 2019)	69%
(Hekler & Kather, 2020a)	75.03%
(Hekler & Kather, 2020b)	75.03%
(Adegun & Viriri, 2020)	95%
(Cassidy et al., 2022)	AUC:80%
(Cassidy et al., 2022)	56%
Our proposed system	87.77%

Table 4.19: *Comparison with the Related Works*

The proposed partitioning is a promising idea and does not require database inflation, as it ensures diversity and sound learning without creating bias towards the majority class, but training each subset on its own deep learning model will create pressure on training resources. also further reprocessing can be useful to avoid the variance in this dataset.

4.8 Conclusion

In conclusion, we have meticulously detailed the implementation of our melanoma detection system, leveraging the power of Convolutional Neural Networks (CNN) within the realm of deep learning. This system aims to enhance the early detection and diagnosis of melanoma through accurate interpretation of dermoscopic images.

We thoroughly discuss each phase of the implementation process, starting with the tools and environments utilized. The preprocessing techniques applied to the data are elaborated upon, followed by an in-depth examination of the selection and fine-tuning of model hyperparameters. We also outline the series of experiments conducted to refine our model, describe the development of the user interface, and provide a comprehensive analysis of the results along with an acknowledgment of the system's limitations.

Conclusion and Future Works

Melanoma, although a rarer form of skin cancer, is also one of the most deadly. Early diagnosis is crucial to improving the chances of survival. However, traditional diagnostic approaches based on visual examination and dermatoscopic imaging are often insufficient to detect the disease at an early stage.

Automating melanoma diagnosis using deep learning and advanced imaging techniques is now essential. Convolutional neural networks (CNNs) have shown promising potential in medical image analysis due to their ability to identify complex patterns. However, the imbalance of the data sets represents a major challenge for their training. Images of non-melanoma lesions far outnumber melanoma images. This imbalance often leads to biased models that are unable to detect melanoma accurately.

Our approach addresses the data imbalance by dividing the dataset into balanced subsets. This reduces the numerical imbalance between non-melanoma classes, allowing deep learning models pre-trained on these subsets to perform better. We used ResNet50 as a pre-trained model, associating it with each subset and combining its decisions by mean and weighted voting. Classifiers such as SVM, XGBoost and logistic regression were also applied.

Our results were promising. Individual models achieved up to 81% accuracy. By fusing the model decisions, we further enhanced performance. The weighted voting method produced the best results: 87.73% accuracy, 86.07% precision, 90% recall, and 88.01% F1 score. This significant improvement highlights the importance of addressing dataset imbalance and integrating multiple models to strengthen the overall detection system.

The results obtained in this study open up a number of promising avenues for future work, aimed at further improving the performance of melanoma detection models. Some key perspectives are as follows:

1. Improve weighted prediction fusion techniques to maximise performance by assigning optimal weights based on individual model performance.
2. Use GANs to generate synthetic images of melanoma.
3. Combine several model architectures (e.g. ResNet, Inception, EfficientNet) to take advantage of their respective strengths, using ensemble techniques.

Bibliography

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

Ahmed, M., & Islam, N. (2020). Deep learning: hope or hype. *Annals of Data Science*, *7*. https://doi.org/10.1007/s40745-019-00237-0

Akiba, T., Sano, S., & Yanase, T. e. a. (2019). Optuna: a next-generation hyperparameter optimization framework, 2623–2631. https://doi.org/10.1145/3292500.3330701

Alzubaidi, L., Zhang, J., & Humaidi, A. J. e. a. (2021). Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Argenziano, G., Fabbrocini, G., Carli, P., et al. (1998). Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of Dermatology*, *134*, 1563–1570.

Argenziano, G., Soyer, H. P., De Giorgi, V., et al. (2000). *Interactive cd of dermoscopy*. EDRA Medical Publishing; New Media.

Arif, R. B., Siddique, M. A., & et al., K. (2018). Study and observation of the variations of accuracies for handwritten digits recognition with various hidden layers and epochs using convolutional neural network.

Baratloo, A., Hosseini, M., Negida, A., & El Ashal, G. (2015). Part 1: simple definition and calculation of accuracy, sensitivity and specificity. *Emergency (Tehran, Iran)*, *3*(2), 48–49.

Barbosa, J., de, M., & Baleiras, G. (2019). Melanoma detection using deep learning methods. https://api.semanticscholar.org/CorpusID:216077247

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166.

Bottou, L. (2012). Stochastic gradient descent tricks. *Neural Networks*. https://api.semanticscholar.org/CorpusID:121049

Bozkurt, F. (2023). Skin lesion classification on dermatoscopic images using effective data augmentation and pre-trained deep learning approach. *Multimedia Tools and Applications*, *82*, 18985–19003. https://doi.org/10.1007/s11042-022-14095-1

Bre, F., Gimenez, J., & Fachinotti, V. (2017). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, *158*. https://doi.org/10.1016/j.enbuild.2017.11.045

Brinker, T. J., Hekler, A., & Hauschild, A. e. a. (2019). Comparing artificial intelligence algorithms to 157 german dermatologists: the melanoma classification benchmark.

European Journal of Cancer, *111*, 30–37. https://doi.org/10.1016/j.ejca.2018.12.016

Carli, P., Giorgi, V. D., & et al, D. M. (2000). The role of pattern analysis and the abcd rule of dermoscopy in the detection of histological atypia in melanocytic naevi. *British Journal of Dermatology*, *143*(2), 290–297.

Cassidy, B., Kendrick, C., & Brodzicki, A. e. a. (2022). Analysis of the isic image datasets: usage, benchmarks and recommendations. *Medical Image Analysis*, *75*, 102305. https://doi.org/10.1016/j.media.2021.102305

Chatterjee, S., Dey, D., Munshi, S., & Gorai, S. (2021). Dermatological expert system implementing the abcd rule of dermoscopy for skin disease identification. *Expert Systems with Applications*, *167*, 114204.

Chen, T., Kuo, T., & Chan, H. (2000). Dermatofibroma is a clonal proliferative disease. *J Cutan Pathol*, *27*(1), 36–9. https://doi.org/10.1034/j.1600-0560.2000.027001036.x

Choe, D., Choi, E., & Kim, D. K. (2020). The real-time mobile application for classifying of endangered parrot species using the cnn models based on transfer learning. *Mobile Information Systems*, *2020*, 1475164. https://doi.org/10.1155/2020/1475164

Choi, A., & Oakley, A. (2021). A retrospective review of cutaneous vascular lesions referred to a teledermatology clinic. *Journal of Primary Health Care*, *13*(1), 70–74. https://doi.org/10.1071/HC20046

Chollet, F. (2017). Xception: deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258. https://doi.org/10.48550/arXiv.1610.02357

Codella, N., Gutman, D., & Celebi, M. e. 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). *arXiv:1710.05006 [cs.CV]*. https://arxiv.org/abs/1710.05006

De Faria, S., Filipe, J., & et al., P. (2019). Light field image dataset of skin lesions. *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2019*, 3905–3908. https://doi.org/10.1109/EMBC.2019.8856578

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(61), 2121–2159. http://jmlr.org/papers/v12/duchi11a.html

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*. https://doi.org/10.48550/arXiv.1603.07285

Dwivedi, A., Tripathi, A., & Ray, R. S. e. a. (2022). *Skin cancer: pathogenesis and diagnosis*. Springer Singapore. https://doi.org/10.1007/978-981-16-0364-8

Elder, C., David E. M.B., Jucovy, P. M. M., & Tuthill, R. J. M. e. a. (1980). The classification of malignant melanoma. *The American Journal of Dermatopathology*, *2*(4), 315–320. https://doi.org/10.1097/00000372-198000240-00005

Fraccaro, P., & et al., D. O. (2015). Behind the screens: clinical decision support methodologies – a review. *Health Policy and Technology*, *4*(1), 29–38. https://doi.org/10.1016/j.hlpt.2014.10.001

Frederick O. Stephens, K. R. A. (2015). *Basics of oncology*. Springer Cham. https://doi.org/10.1007/978-3-319-23368-0

Ganaie, M., Hu, M., Malik, A., Tanveer, M., & Suganthan, P. (2022). Ensemble deep learning: a review. *Engineering Applications of Artificial Intelligence*, *115*, 105151. https://doi.org/10.1016/j.engappai.2022.105151

Gao, Z., Luo, Z., Zhang, W., Lv, Z., & Xu, Y. (2020). Deep learning application in plant stress imaging: a review. *AgriEngineering*. https://doi.org/10.3390/agriengineering2030029

Georgakopoulos, S. V., & Plagianakos, V. P. (2017). A novel adaptive learning rate algorithm for convolutional neural network training. *International Conference on Engineering Applications of Neural Networks*. https://api.semanticscholar.org/CorpusID:37553399

Giotis, I., Molders, N., Land, S., & et al., B. (2015). Med-node: a computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Systems with Applications*, *42*. https://doi.org/10.1016/j.eswa.2015.04.034

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 249–256.

Gu, J., Wang, Z., & Kuen, J. e. a. (2015). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*. https://doi.org/10.1016/j.patcog.2017.10.013

Gutman, D., Codella, N. C. F., & Celebi, E. e. 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). *eprint arXiv:1605.01397*. https://arxiv.org/abs/1605.01397

Hall, C. (2013). New technologies to aid in the diagnosis of malignant melanoma. *Dermatology in Practice*, *19*, 21.

Harrison, M. (2016). *Learning the pandas library: python tools for data munging, analysis, and visual*. CreateSpace Independent Publishing Platform.

Hasegawa, K., Fujiwara, R., & Sato, K. e. a. (2015). Possible involvement of keratinocyte growth factor in the persistence of hyperpigmentation in both human facial solar lentigines and melasma. *Ann Dermatol*, *27*(5), 626–629. https://doi.org/10.5021/ad.2015.27.5.626

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://api.semanticscholar.org/CorpusID:206594692

Hekler, A., & Kather, J. N. e. a. (2020a). Effects of label noise on deep learning-based skin cancer classification. *Frontiers in Medicine*, *7*. https://doi.org/10.3389/fmed.2020.00177

Hekler, A., & Kather, J. N. e. a. (2020b). Effects of label noise on deep learning-based skin cancer classification. *Frontiers in Medicine*, *7*. https://doi.org/10.3389/fmed.2020.00177

Higgins, H. 2., Lee, K., & Galan, A. e. a. (2015). Melanoma in situ: part i. epidemiology, screening, and clinical features. *J Am Acad Dermatol*, *73*(2), 181–90, quiz 191–2. https://doi.org/10.1016/j.jaad.2015.04.014

Higgins, J., Maher, M., & Douglas, M. (2015). Diagnosing common benign skin tumors. *Am Fam Physician*, *92*(7), 601–7.

Hodgson, C. (1963). Senile lentigo. *Arch Dermatol*, *87*, 197–207. https://doi.org/10.1001/archderm.1963.01590140059010

Hosna, A., Merry, E., Gyalmo, J., et al. (2022). Transfer learning: a friendly introduction. *Journal of Big Data*, *9*(1), 102. https://doi.org/10.1186/s40537-022-00652-w

Huang, L., Diao, J., Nie, et al. (2021). Memristor based binary convolutional neural network architecture with configurable neurons. *Frontiers in Neuroscience*, *15*, 639526. https://doi.org/10.3389/fnins.2021.639526

Imtiaz, I., Ahmed, I., Jeon, G., & Muramatsu, S. (2021). An efficient image processing and machine learning based technique for skin lesion segmentation and classification. *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1499–1505. https://api.semanticscholar.org/CorpusID:246533718

Jafari, M., Karimi, N., & Nasr-Esfahani, E. e. a. (2016). Skin lesion segmentation in clinical images using deep learning. *2016 23rd International Conference on Pattern Recognition (ICPR)*, 337–342. https://doi.org/10.1109/ICPR.2016.7899656

Ji, Q., Huang, J., He, W., & Sun, Y. (2019). Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images. *Algorithms*, *12*(3), 51. https://doi.org/10.3390/a12030051

Johr, R. H. (2002). Dermoscopy: alternative melanocytic algorithms-the abcd rule of dermatoscopy, menzies scoring method, and 7-point checklist. *Clinical Dermatology*, *20*(3), 240–247. https://doi.org/10.1016/s0738-081x(02)00236-5

Joshi, S., Kumar, V., & Venkataramanan, V. e. a. (2023). A review on neural networks and its applications. *Journal of Computer Technology & Applications*, *14*, 2023. https://doi.org/10.37591/jocta.v14i2.1062

Kaliyadan, F. (2016). The scope of the dermoscope. *Indian Dermatology Online Journal*, *7*(5), 359–363. https://doi.org/10.4103/2229-5178.190496

Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of convolutional neural networks on a histopathology dataset. *ICT Express*, *6*, 312–315. https://api.semanticscholar.org/CorpusID:218937426

Kandel, I., Castelli, M., & Popovi, A. (2020). Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images. *Journal of Imaging*, *6*. https://api.semanticscholar.org/CorpusID:225228297

Kawahara, J., Daneshvar, S., Argenziano, G., & Hamarneh, G. (2019). 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, A., Sohail, A., et al. (2019). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*, 5455–5516. https://doi.org/10.48550/arXiv.1901.06032

Khan, A., Sohail, A., & Zahoora, U. e. a. (2020). A survey of the recent architectures of deep convolutional neural networks [Number of Pages: 70, Number of Figures: 11, Number of Tables: 11]. *Artificial Intelligence Review*, *53*, 5455–5516. https://doi.org/10.1007/s10462-020-09825-6

Khan, S. M., Liu, X., & Nath, S. e. a. (2021). A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet Digital Health*, *3*(1), e51–e66. https://doi.org/10.1016/S2589-7500(20)30240-5

Khan Raiaan, M. A., Sakib, S., & Fahad, N. M. e. a. (2024). A systematic review of hyperparameter optimization techniques in convolutional neural networks. *Decision Analytics Journal*, *11*, 100470. https://doi.org/https://doi.org/10.1016/j.dajour.2024.100470

Kim, J. C., Park, T. J., & Kang, H. Y. (2022). Skin-aging pigmentation: who is the real enemy? *Cells*, *11*(16), 2541. https://doi.org/10.3390/cells11162541

Kim, K.-S., & Choi, Y.-S. (2021). Hyadamc: a new adam-based hybrid optimization algorithm for convolution neural networks. *Sensors*, *21*, 4054. https://doi.org/10.3390/s21124054

Kingma, D. P., & Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://doi.org/10.48550/arXiv.1412.6980

Long, L., & Zeng, X. (2022). *Beginning deep learning with tensorflow: work with keras, mnist data sets, and advanced neural networks* (1st ed.). Apress Berkeley, CA. https://doi.org/https://doi.org/10.1007/978-1-4842-7915-1

Luzar, B., & Calonje, E. (2010). Cutaneous fibrohistiocytic tumours - an update. *Histopathology*, *56*(1), 148–165. https://doi.org/10.1111/j.1365-2559.2009.03447.x

Maqsood, S., & Damaševičius, R. (2023). Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare [Epub 2023 Jan 24]. *Neural Networks*, *160*, 238–258. https://doi.org/10.1016/j.neunet.2023.01.022

Mendonca, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., & Rozeira, J. (2013). Ph² - a dermoscopic image database for research and benchmarking. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5437–5440. https://doi.org/10.1109/EMBC.2013.6610779

Mendonça, T., Ferreira, P. M., & et al., M. (2013). Ph2 - a dermoscopic image database for research and benchmarking. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5437–5440. https://doi.org/10.1109/EMBC.2013.6610779

Mothoneos, J. (2023). *Understanding melanoma*. Cancer Council Australia.

Nachbar, F., Stolz, W., Merkle, T., & et al., C. (1994). The abcd rule of dermatoscopy. high prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, *30*(4), 551–559. https://doi.org/10.1016/s0190-9622(94)70061-3

Nugroho, E., Ardiyanto, I., & Nugroho, H. A. (2024). Addressing imbalance isic-2019 dataset in dermoscopic pigmented skin lesion classification. *18*, 563–573. https://doi.org/10.24507/icicel.18.06.563

Olaoye, F., & Potter, K. (2024). Transfer learning and its role in machine learning. *Technology*.

Pal, A., Kar, S., Taneja, A., & Jadoun, V. (2020). Image captioning and comparison of different encoders. *Journal of Physics: Conference Series*, *1478*, 012004. https://doi.org/10.1088/1742-6596/1478/1/012004

Patel, A. M., Chou, E. L., & Findeiss, L. e. a. (2012). The horizon for treating cutaneous vascular lesions. *Seminars in Cutaneous Medicine and Surgery*, *31*(2), 98–104. https://doi.org/10.1016/j.sder.2012.02.001

Pathan, S., Prabhu, K. G., & Siddalingaswamy, P. C. (2018). Techniques and algorithms for computer aided diagnosis of pigmented skin lesions - a review. *Biomed. Signal Process. Control.*, *39*, 237–262. https://api.semanticscholar.org/CorpusID:41025652

Patro, S. G., & Sahu, D.-K. K. (2015). Normalization: a preprocessing stage. *IARJSET*. https://doi.org/10.17148/IARJSET.2015.2305

Pehamberger, H., Steiner, A., & Wolff, K. (1987). In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions. *Journal of* *the American Academy of Dermatology*, *17*(4), 571–583. https://doi.org/10.1016/s0190-9622(87)70239-4

Penman, I. D., Ralston, S. H., & Strachan, M. W. J. e. a. (2022). *Davidson's principles and practice of medicine*. Health Sciences.

Popescu, M.-C., Balas, V., & Perescu-Popescu, L. e. a. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, *8*.

Prusa, J. D., & Khoshgoftaar, T. M. (2017). Improving deep neural network design with new text data representations. *Journal of Big Data*, *4*(1), 7. https://doi.org/10.1186/s40537-017-0065-8

Puttagunta, M., & Ravi, S. (2021). Medical image analysis based on deep learning approach. *Multimedia Tools and Applications*, *80*, 24365–24398. https://doi.org/https://doi.org/10.1007/s11042-021-10707-4

Qureshi, A. S., & Roos, T. (2023). Transfer learning with ensembles of deep neural networks for skin cancer detection in imbalanced data sets. *Neural Processing Letters*, *55*(5), 4461–4479. https://doi.org/10.1007/s11063-022-11049-4

Rahman, Z., Hossain, M. S., & et al., I. (2021). An approach for multiclass skin lesion classification based on ensemble learning. *Informatics in Medicine Unlocked*, *25*, 1–9. https://doi.org/10.1016/j.imu.2021.100659

Riker, A. I. (2018). *Melanoma - a modern multidisciplinary approach*. Springer Cham. https://doi.org/10.1007/978-3-319-78310-9

Robinson, J. K., & Callen, J. P. (2005). Biotechnology succeeds in revolutionizing medical science. *Archives of Dermatology*, *141*(2), 133–134. https://doi.org/10.1001/archderm.141.2.133

Schalkoff, R. J. (1997). *Artificial neural networks* (illustrated) [Original from the University of Michigan, Digitized 17 Nov 2007]. McGraw-Hill.

Senan, E. M., & Jadhav, M. E. (2021). Analysis of dermoscopy images by using abcd rule for early detection of skin cancer [1st International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE - 2020)]. *Global Transitions Proceedings*, *2*(1), 1–7. https://doi.org/https://doi.org/10.1016/j.gltp.2021.01.001

Sgouros, D., Apalla, Z., & Ioannides, D. e. a. (2018). Dermoscopy of common inflammatory disorders [Epub 2018 Jul 31]. *Dermatologic Clinics*, *36*(4), 359–368. https://doi.org/10.1016/j.det.2018.05.003

Shah, A., Shah, M., & Pandya, A. e. a. (2023). A comprehensive study on skin cancer detection using artificial neural network (ann) and convolutional neural network (cnn). *Clinical eHealth*, *6*, 76–84. https://doi.org/https://doi.org/10.1016/j.ceh.2023.08.002

Sharma, S., Gupta, S., & Gupta, D. e. a. (2022). Performance evaluation of the deep learning based convolutional neural network approach for the recognition of chest x-ray images. *Frontiers in Oncology*, *12*, 932496. https://doi.org/10.3389/fonc.2022.932496

Sheard, R. (2021). *Understanding skin cancer*. Cancer Council Australia.

Shetty, B., Fernandes, R., & et al., R. (2022). Author correction: skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, *12*.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(60). https://doi.org/10.1186/s40537-019-0197-0

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition [Focus to learn more]. *arXiv preprint arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

Singh, L., Janghel, R., & Sahu, S. (2020). Designing a retrieval-based diagnostic aid using effective features to classify skin lesion in dermoscopic images. *Procedia Computer Science*, *167*, 2172–2180. https://doi.org/https://doi.org/10.1016/j.procs.2020.03.267

Smith, L. N. (2015). Cyclical learning rates for training neural networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 464–472. https://api.semanticscholar.org/CorpusID:15247298

Srinivasan, K., & et al., G. (2021). Performance comparison of deep cnn models for detecting driver's distraction. *Cmc -Tech Science Press-*, *68*, 4109–4124. https://doi.org/10.32604/cmc.2021.016736

Srivastava, N., Hinton, G., & et al., A. K. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. https://doi.org/10.48550/arXiv.1505.00387

Swain, B. K., & et al., R. (2021). Effectiveness of data augmentation for classification of melanoma using deep convolutional neural network. *2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, 1–5. https://doi.org/10.1109/APSIT52773.2021.9641346

Tao, W., Al-Amin, M., & Chen, H. e. a. (2020). Real-time assembly operation recognition with fog computing and transfer learning for human-centered intelligent manufacturing. *Procedia Manufacturing*, *48*, 926–931. https://doi.org/10.1016/j.promfg.2020.05.131

Tschandl, P., Rosendahl, C., & Kittler, H. (2018a). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, *5*, 180161. https://doi.org/10.1038/sdata.2018.161

Wen, D., Khan, S. M., & Xu, A. J. e. a. (2022). Characteristics of publicly available skin cancer image datasets: a systematic review. *REVIEW*, *4*(1), e64–e74. https://doi.org/https://doi.org/10.1016/S2589-7500(21)00252-1

Yamashita, R., Nishio, M., Do, R. K., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Yang, J., Wu, X., Liang, J., & et al., S. (2020). Self-paced balance learning for clinical skin disease recognition. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(8), 2832–2846. https://doi.org/10.1109/TNNLS.2019.2917524

Webography

Abdelmalek, D. M. (2018). Detecting skin cancer with a handheld magnifying tool vs. the naked eye. Retrieved February 22, 2024, from https://abcnews.go.com/Health/detecting-skin-cancer-handheld-magnifying-tool-naked-eye/story?id=59905561

About keras. (2024). *Keras website*. Retrieved May 18, 2024, from https://keras.io/about/

American Cancer Society. (n.d.). Key statistics for melanoma skin cancer. Retrieved February 8, 2024, from https://www.cancer.org/content/dam/CRC/PDF/Public/8823.00.pdf

Ebrahimi, M. S., & Abadi, H. K. (2018). Study of residual networks for image recognition. Retrieved May 18, 2024, from https://doi.org/10.48550/arXiv.1805.00325

Guy-Evans, O. (2024). Neuron structure and function [Reviewed by Saul Mcleod, PhD]. Retrieved April 23, 2024, from https://www.simplypsychology.org/neuron.html

Kaggle documentation. (2024). Retrieved May 16, 2024, from https://www.kaggle.com/docs

Matplotlib Developers. (2024). Matplotlib: visualization with python. Retrieved May 18, 2024, from https://matplotlib.org/

MedlinePlus. (2024). Images from medlineplus: skin cancer. Retrieved June 16, 2024, from https://medlineplus.gov/ency/imagepages/19679.htm%20and%20https://medlineplus.gov/ency/imagepages/8912.htm

Moscow Clinical Scientific Center. (2024). Neoplasms of the skin: diagnosis and treatment. Retrieved June 17, 2024, from https://mknc.ru/uslugi-centra/11756/novoobrazovaniya-koji-diagnostika-i-lechenie/2/?kt_lang=en

NumPy Developers. (2023). What is numpy? Retrieved May 18, 2024, from https://numpy.org/doc/stable/user/whatisnumpy.html

OpenCV Developers. (2024). About opencv. Retrieved May 18, 2024, from https://opencv.org/about/

Pandas. (2024). *Pandas*. Retrieved May 18, 2024, from https://pandas.pydata.org/

Prabhu, R. (2018, March). Understanding of convolutional neural network (cnn) – deep learning [Accessed: 2024-05-23].

Python Software Foundation. (2024). Python documentation. Retrieved May 17, 2024, from https://www.python.org/doc/

Resnet.preprocess_input. (2024). Retrieved May 18, 2024, from https://www.tensorflow.org/api_docs/python/tf/keras/applications/resnet/preprocess_input

Science Photo Library. (n.d.). Science photo library. Retrieved February 22, 2024, from https://www.sciencephoto.com/

scikit-learn Developers. (2024). Scikit-learn: machine learning in python [Accessed: 2024-05-18]. https://scikit-learn.org/stable/

TensorFlow. (2024). Tensorflow: documentation. Retrieved May 18, 2024, from https://www.tensorflow.org/

UK Health Security Agency. (2022, September). About the uk health security agency [© Crown copyright 2022]. *UK Health Security Agency*. Retrieved May 18, 2024, from https://www.gov.uk/government/organisations/uk-health-security-agency

Verywell Health. (2023). What do malignant and benign mean? Retrieved June 16, 2024, from https://www.verywellhealth.com/what-does-malignant-and-benign-mean-514240