Comparative Analysis of Vision Transformers and CNNs in Melanoma Classification

Farnaz Haghshenas

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Computer Science) at

Concordia University

Montréal, Québec, Canada

December 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Farnaz Haghshenas**

Entitled: **Comparative Analysis of Vision Transformers and CNNs in Melanoma Classification**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Computer Science)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Ching Y. Suen

_____ Examiner
Dr. Sudhir Mudur

_____ Examiner
Dr. Ching Y. Suen

_____ Supervisor
Dr. Adam Krzyzak

Approved by _____
Charalambos Poullis, Graduate Program Director
Department of Computer Science and Software Engineering

_____ 2024

_____
Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Comparative Analysis of Vision Transformers and CNNs in Melanoma Classification

Farnaz Haghshenas

The increasing number of skin cancers underscores the critical importance of early detection and accurate classification to improve treatment outcomes. Melanoma, a malignant skin cancer, has the highest mortality rate among all skin cancer types. Early detection of melanoma significantly enhances the chances of effective treatment and survival rates. This research presents a comparative analysis of cutting-edge deep learning methodologies in medical imaging, specifically focusing on Vision Transformers (ViT) and Convolutional Neural Networks (CNNs) for melanoma cancer detection. This study further examines the influence of domain-specific transfer learning on improving melanoma detection accuracy by pre-training these deep learning models on various datasets, such as ImageNet, BreakHis, and ISIC 2019. The models are then meticulously fine-tuned using a private annotated dataset of melanoma dermoscopic images. In addition, we employed the k-fold cross-validation technique to evaluate the reliability of our models. Our experimental results highlight the significant performance of advanced deep learning methodologies and transfer learning approaches, with the ViT-B16 model achieving an exceptional diagnostic accuracy of 97.97%, outperforming other models, specifically the pre-trained CNNs models. Moreover, This study highlights the critical role of large, diverse datasets in transfer learning, demonstrating their effectiveness in improving model performance for melanoma detection.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Professor Adam Krzyzak. His unwavering support, insightful guidance, and immense knowledge throughout this research process have been invaluable.

I am also profoundly grateful to the members of my thesis committee, Dr. Ching Y. Suen and Dr. Sudhir Mudur, for their time, and expertise. I am honored to have had the opportunity to have them as my committee members.

Last but not least, I would like to express my heartfelt appreciation to my family and friends for their unconditional love, understanding, and encouragement. To my parents, Reza and Nadia, thank you for always believing in me and supporting my academic pursuits. This success would not have been possible without your support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we discuss our research topic and the key motivations underlying our work (Section 1.1), emphasizing the critical need for advancements in melanoma detection. Following this, Section 1.2 summarizes the primary goals of the study, which aim to enhance diagnostic accuracy through innovative methodologies. Section 1.3 highlights the novel contributions made to this research. Next, Section 1.4 provides an overview of the thesis structure, and Section 1.5 lists the related published articles.

1.1 Motivation

In recent years, the increasing number of cancer cases has highlighted the critical importance of early detection and effective treatment strategies. The year 2020 alone recorded nearly 10 million deaths worldwide due to various forms of cancer, prompting the global health community to escalate its efforts in utilizing advanced technologies to improve diagnostic and treatment protocols Sung et al. (2021). Among the various types of cancer, skin cancer is notably widespread. Melanoma, in particular, is a life-threatening form of skin cancer, yet it is treatable when detected and treated in its early stages Geller, Swetter, Brooks, Demierre, and Yaroch (2007).

The last decade has marked a significant transformation in the landscape of cancer diagnosis, primarily driven by advancements in the integration of Artificial Intelligence (AI) in medical imaging analysis. The introduction of Convolutional Neural Networks (CNNs) has been a cornerstone

of this evolution by automating the detection process and achieving more accurate results. CNNs have been particularly effective in analyzing medical images for the detection of various cancers, including skin, prostate, and breast cancers A. A. Abbasi et al. (2020); Bardou, Zhang, and Ahmad (2018); Brinker et al. (2018).

More recently, Vision Transformers (ViT) have introduced a new paradigm by adapting the transformer model, initially developed for natural language processing, to computer vision tasks Dosovitskiy et al. (2020). ViT employs a self-attention mechanism that allows it to weigh the importance of various parts of an image. This approach has shown promising results in different image classification tasks, including skin cancer detection Azad et al. (2024); Xin et al. (2022).

Despite these technological leaps, adopting Deep Learning (DL) techniques, such as CNNs and ViT, within the medical domain is challenging. One of the primary obstacles is the necessity for extensive datasets to train these models effectively and ensure their reliable performance in real-world applications Aljabri, AlAmir, Al Ghamdi, Abdel-Mottaleb, and Collado-Mesa (2022). Furthermore, the absence of such datasets often leads to overfitting, where models perform well on training data but underperform on new, unseen data Mutasa, Sun, and Ha (2020). The necessity for high levels of expertise for manual annotation and the resource-intensive nature of medical data collection intensifies the difficulty in securing sufficient medical images for specific health conditions. Another significant hurdle is the computational expense and the requirement for extensive memory resources to train CNNs and ViT from scratch. These challenges highlight the complexities involved in integrating CNNs and ViT technologies into practical medical diagnostics and underscore the need for innovative solutions to navigate these obstacles. In response to these challenges, transfer learning (TL) emerges as a valuable strategy. Transfer learning involves utilizing models pre-trained on large, diverse datasets, such as ImageNet, and adapting these models to specific medical diagnostics tasks with fewer data and less computational demand Hosny, Kassem, and Foaud (2018). This approach addresses the challenge of data scarcity and significantly reduces the risk of overfitting, making it a cornerstone of modern medical imaging research. These challenges motivated us to find an effective method to enhance the precision of melanoma detection. Consequently, we performed a comparative analysis of seven advanced deep learning models using three distinct TL approaches.

Our research utilized a private dataset of real-world dermoscopic images of melanoma and non-melanoma cases, created at the Warsaw Maria Skłodowska-Curiebreak National Research Institute of Oncology, Department of Soft Tissue/Bone Sarcoma and Melanoma Gil, Osowski, Swiderski, and Słowińska (2023).

1.2 Main Goals

The main objective of this study is to develop a precise diagnostic framework for melanoma detection. To achieve this, we conducted a comparative analysis of seven advanced deep learning models using three distinct TL approaches. In this study, we rigorously evaluated the performance of various CNNs and two variations of ViT model, in addition to two hybrid models. Our assessment used TL, with each model initially trained on a range of datasets. In this research, we utilized the well-known ImageNet dataset alongside two specialized medical datasets, including BreakHis, which contains histopathological images of breast cancer, and ISIC 2019, which focuses on skin cancer Codella et al. (2017); Hernandez et al. (2024); Spanhol, Oliveira, Petitjean, and Heutte (2016); Tschandl, Rosendahl, and Kittler (2018). We primarily evaluated seven distinct models across three scenarios based on different TL approaches:

- General-purpose dataset (ImageNet): We assessed how transfer learning with ImageNet, a large, general-purpose dataset, influences model performance.
- Different medical domain dataset (BreakHis): We examined the effectiveness of transfer learning from the BreakHis dataset, which belongs to a different medical domain, to see how it impacts the models.
- Domain-specific medical dataset (ISIC 2019): We investigated the impact of specific domain transfer learning using the ISIC 2019 dataset, which is directly related to skin cancer, on the performance of the models.

The deliberate choice of BreakHis and ISIC 2019 aims to investigate the efficacy of applying TL within medical domains instead of relying solely on general-purpose datasets like ImageNet. By systematically evaluating these scenarios, this study aims to provide a robust diagnostic framework

that enhances the accuracy and reliability of melanoma detection using advanced deep learning techniques.

1.3 Novel Contributions

The main novel contributions of this thesis can be summarized as follows:

(1) Comprehensive comparative analysis of seven advanced deep learning models for melanoma detection, including three CNNs, two variations of Vision Transformers, and two hybrid models.

(2) Examination of the potential benefits of hybrid models that use the strengths of both CNNs and ViT architectures for melanoma diagnosis.

(3) Evaluation of the performance of two ViT variations using our private target melanoma dataset, providing insights into their efficacy for melanoma detection.

(4) Systematic evaluation of three distinct transfer learning approaches using datasets from different domains:

- General-purpose dataset (ImageNet)
- Medical domain dataset (BreakHis)
- Domain-specific medical dataset (ISIC 2019)

(5) Investigation of the impact of domain-specific transfer learning on model performance using a subset of the ISIC 2019 dataset.

(6) Assessment of cross-domain transfer learning effectiveness by employing a breast cancer histopathology dataset (BreakHis) for melanoma detection.

These contributions collectively aim to advance the field of automated melanoma diagnosis by providing a comprehensive understanding of how different deep learning architectures and transfer learning strategies can be optimized for improved detection accuracy. We will compare our findings with previous studies that utilized the same private dataset to demonstrate the effectiveness of the proposed method in this research.

1.4 Thesis Structure

This thesis is organized into five chapters, each addressing specific aspects of our research on melanoma detection using advanced deep learning techniques.

- **Chapter 1** introduces the motivation behind our study, outlines the main goal, and highlights the novel contributions of our work. It also provides an overview of the thesis structure and lists relevant publications.

- **Chapter 2** provides a comprehensive background on skin cancer, with a focus on melanoma's characteristics, risk factors, and diagnostic challenges. It reviews computer-aided diagnosis (CAD) systems developed for melanoma detection, highlighting relevant studies utilizing machine learning (ML), deep learning (DL), and ViT models in this area. Additionally, the chapter examines the role of TL in medical imaging, specifically its application to melanoma classification.

- **Chapter 3** offers a detailed description of our proposed method, meticulously explaining each step of the framework developed for the three melanoma detection scenarios. This chapter also includes a comprehensive description of the datasets used, explaining their relevance, structure, and preprocessing steps.

- **Chapter 4** is dedicated to presenting and analyzing the results of all three scenarios. This chapter provides a comprehensive discussion of our findings, offering analytical insights into the outcomes of our research.

- **Chapter 5** concludes the thesis by summarizing our study's main findings and discussing their implications. It also outlines potential directions for future work in this area.

Through this structure, we aim to provide a clear and logical progression of our research, from its conceptual foundations to its practical outcomes and future prospects.

1.5 Publications

- **Haghshenas, F.**, Krzyżak, A., Osowski, S. (2024). Comparative Study of Deep Learning Models in Melanoma Detection. In: Suen, C.Y., Krzyzak, A., Ravanelli, M., Trentin, E., Subakan, C., Nobile, N. (eds) Artificial Neural Networks in Pattern Recognition. ANNPR 2024. Lecture Notes in Computer Science(), vol 15154, pp. 121–131. Springer, Cham. https://doi.org/10.1007/978-3-031-71602-7_11 Haghshenas, Krzyżak, and Osowski (2024).

Chapter 2

Background and Literature Review

2.1 Introduction

This chapter is divided into two main sections. The first section (Section 2.4) provides an overview of cancer and skin cancer. It then defines melanoma, detailing its characteristics and global statistics. Additionally, it discusses the risk factors associated with melanoma and the methods used for its diagnosis. In Section 2.5, we review the latest research on computer-aided diagnosis (CAD) systems for melanoma detection. This includes using Machine Learning (ML), Deep Learning, and Vision Transformers in melanoma detection. We also delve into the utilization of transfer learning in medical image analysis.

2.2 Cancer: An Overview

Cancer is a complex group of diseases characterized by the uncontrolled growth and spread of abnormal cells. It can develop in almost any organ or tissue of the body, resulting from genetic mutations that disrupt normal cell growth and division processes Hanahan and Weinberg (2011). The transformation of normal cells into cancer cells is a multistep process, typically requiring multiple genetic alterations. These changes can be triggered by various factors, including:

- Environmental exposures (e.g., UV radiation, tobacco smoke)
- Lifestyle factors (e.g., diet, physical inactivity)

- Inherited genetic mutations
- Certain infections (e.g., human papillomavirus, hepatitis B virus)

As cancer progresses, it can invade surrounding tissues and metastasize to distant parts of the body, making treatment more challenging Fidler (2003).

2.3 Skin Cancer: A Growing Concern

Skin cancer is the most common form of cancer globally. It primarily develops in sun-exposed areas of the body, although it can occur anywhere on the skin. Skin cancer is broadly categorized into two main types: non-melanoma skin cancer (NMSC) and melanoma skin cancer. In this research, we focus specifically on melanoma detection, given its potential for rapid progression and the critical importance of early diagnosis in improving patient outcomes Gordon (2013).

2.3.1 Non-Melanoma Skin Cancers

Non-melanoma skin cancers primarily include Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC).

Basal Cell Carcinoma

Basal cell carcinoma is the most common type of skin cancer, accounting for approximately 80% of non-melanoma skin cancers Marzuka and Book (2015). BCCs originate in the basal cells of the epidermis and are typically slow-growing. They rarely metastasize but can cause significant local tissue damage if left untreated. Common characteristics of BCCs include:

- Pearly, waxy bumps
- Flat, flesh-colored or brown lesions
- Bleeding or scabbing sores that heal and return

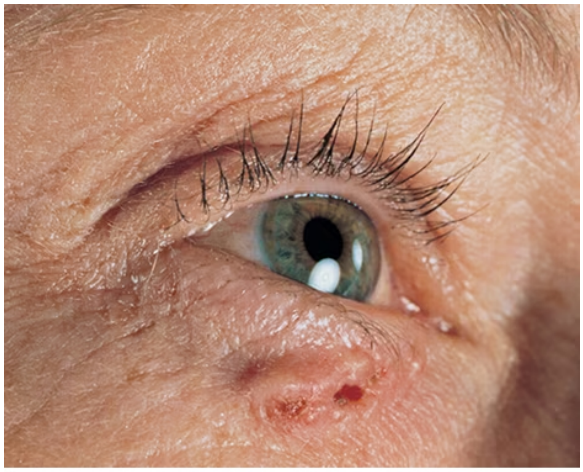

Figure 2.1: Characteristics of basal cell carcinoma. Source: Mayo Clinic (2024a)

Squamous Cell Carcinoma

Squamous cell carcinoma is the second most common type of skin cancer, arising from the squamous cells in the epidermis. SCCs demonstrate a higher tendency to invade deeper layers of the skin and spread to other parts of the body compared to BCCs, although such cases are relatively rare Que, Zwald, and Schmults (2018).

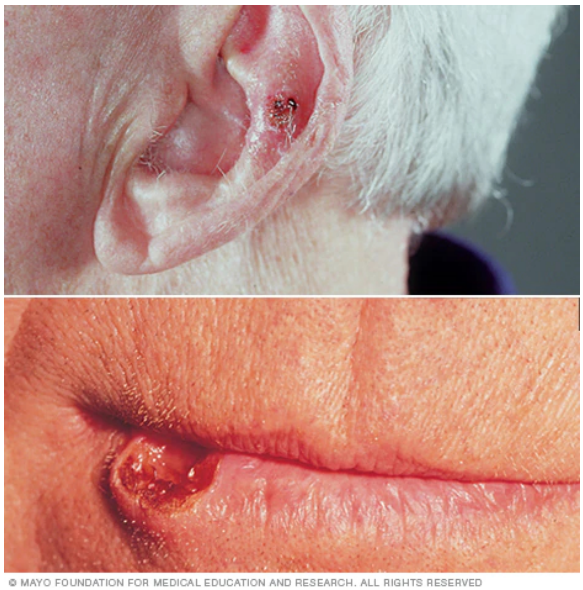

Figure 2.2: Characteristics of squamous cell carcinoma. Source: Mayo Clinic (2024c)

2.3.2 Melanoma Skin Cancer

Melanoma is the most lethal form of skin cancer, developing from melanocytes, which are the cells responsible for producing skin pigment. Melanoma is responsible for the majority of skin cancer deaths due to its aggressive nature and tendency to metastasize Schadendorf et al. (2018).

2.3.3 Prevention and Early Detection

Prevention strategies focus on reducing UV exposure and promoting skin awareness. Early detection remains crucial for improved prognosis, particularly for melanoma. Regular skin self-examinations and professional screenings are recommended, especially for high-risk individuals Wernli et al. (2016).

2.4 Melanoma: A Deeper Dive

2.4.1 Melanoma: Definition and Characteristics

Melanoma originates in melanocytes, the pigment-producing cells in the skin. It can develop anywhere on the body but is most often found on areas exposed to sunlight. What sets melanoma apart is its potential to spread quickly if not caught early Rastrelli, Tropea, Rossi, and Alaibac (2014).

Figure 2.3: Characteristics of melanoma. Source: Mayo Clinic (2024b)

2.4.2 Global Statistics and Epidemiology

Melanoma is a significant health concern worldwide, especially in fair-skinned populations. According to GLOBOCAN 2020, there were an estimated 325,000 new melanoma cases and 57,000 deaths globally in 2020. Incidence rates vary greatly across regions. Australia and New Zealand have the highest rates, followed by Western Europe and North America. In contrast, melanoma is less common in most African and Asian countries Ferlay et al. (2021). In the United States, the American Cancer Society estimates about 100,640 new cases of invasive melanoma will be diagnosed in 2024, with approximately 8,290 expected deaths. The lifetime risk of developing melanoma differs among racial and ethnic groups. For White individuals, it's about 3% (1 in 33), while for Black individuals, it's 0.1% (1 in 1,000), and for Hispanic individuals, 0.5% (1 in 200) American Cancer Society (2024).

2.4.3 Risk Factors for Melanoma

Several factors increase the risk of developing melanoma:

- **UV radiation exposure**: The most significant environmental risk factor, including both natural sunlight and artificial sources like tanning beds Gandini et al. (2011).
- **Genetic factors**: Certain genetic mutations can increase melanoma risk Read, Wadt, and Hayward (2016).
- **Skin type and characteristics**: Fair skin, light hair, and light eyes are associated with higher risk Olsen, Carroll, and Whiteman (2010a).
- **History of sunburns**: Severe sunburns, especially during childhood or adolescence, increase the risk of melanoma later in life Dennis et al. (2008).
- **Number of moles**: Having many moles or atypical moles increases the risk Gandini et al. (2005).
- **Personal or family history**: A personal history of melanoma or a first-degree relative with melanoma increases the risk Olsen, Carroll, and Whiteman (2010b).

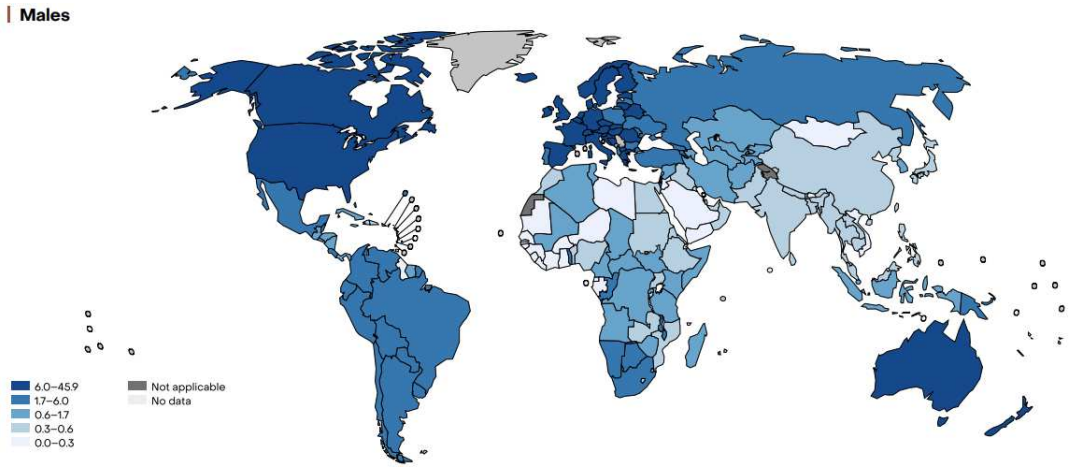

Figure 2.4: Age standardized (World) incidence rates, melanoma of skin, males, all ages. Source: International Agency for Research on Cancer (2022)

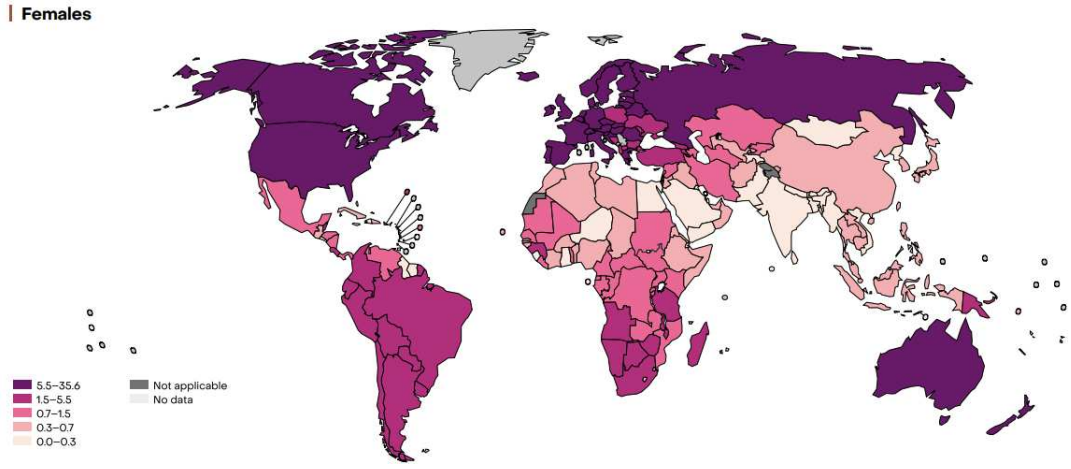

Figure 2.5: Age standardized (World) incidence rates, melanoma of skin, females, all ages. Source: International Agency for Research on Cancer (2022)

2.4.4 Screening and Diagnosis Methods

Early detection is crucial to improve the outcomes of melanoma. Common screening and diagnostic methods include:

- **Self-examination**: The ABCDE criteria (Asymmetry, Border irregularity, Color variation, Diameter $>$ 6 mm, Evolution) is a widely used tool N. R. Abbasi et al. (2004).
- **Full-body skin examinations**: Regular check-ups by healthcare professionals can help detect melanoma early Watts et al. (2015).
- **Dermoscopy**: This non-invasive technique uses a handheld device to visualize structures in the skin not visible to the naked eye, improving diagnostic accuracy Argenziano and Soyer (2012).
- **Biopsy**: The gold standard for diagnosis, where the suspicious lesion is removed and examined under a microscope Swetter et al. (2019).
- **Imaging techniques**: For advanced cases, imaging techniques such as CT scans, MRI, or PET scans may be used to assess the extent of the disease Mohr, Eggermont, Hauschild, and Buzaid (2009).

2.5 Computer-Aided Diagnosis of Skin Cancer

The timely and accurate detection and classification of melanoma are critical for effective treatment and improving patient survival outcomes. In response to this need, the field has seen significant advancements in CAD systems. These systems, powered by AI and ML algorithms, analyze skin images to identify unusual tissue patterns and differentiate between malignant and benign skin lesions. Such AI-empowered technologies have markedly improved the precision of diagnosis, surpassing the sensitivity achieved by experienced dermatologists, who demonstrate 76.9% sensitivity with clinical examinations and 85.7% when using dermoscopy Barros et al. (2020); Chen et al. (2024).

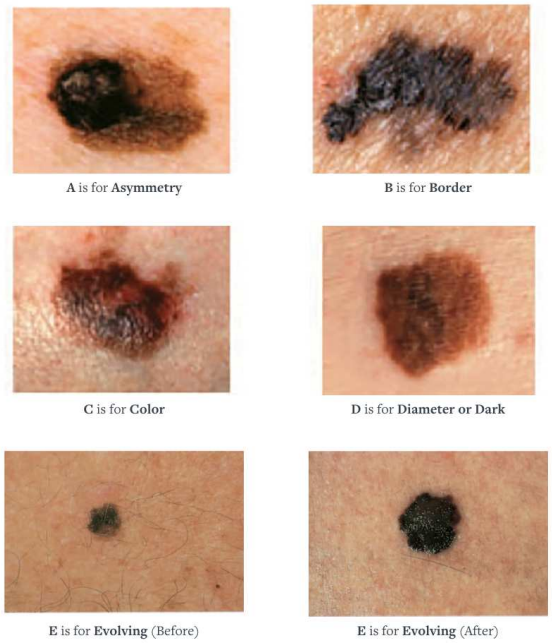

Figure 2.6: The ABCDE criteria for melanoma detection. Source: The Skin Cancer Foundation (2024)

CAD systems utilize various image processing techniques to enhance the visualization of skin lesions that are not visible to the naked eye. By integrating these techniques with AI, along with images obtained from dermoscopy, CAD systems can provide dermatologists with valuable diagnostic insights. One of the key advantages of CAD systems is their ability to handle large volumes of data and learn from vast datasets. This capability enables the development of highly accurate models that can generalize well to new, unseen cases.

Recent research has shown that CAD systems can reach diagnostic accuracy levels similar to those of experienced dermatologists. For example, a study by Esteva et al. (2017) demonstrated that a deep learning algorithm could classify skin cancer with accuracy comparable to that of board-certified dermatologists. These results highlight the significant impact of AI in the field of dermatology. Moreover, Hekler et al. (2019) underscores that CAD systems can assist in reducing inter-observer variability, a common issue in dermatological assessments. By providing consistent and objective evaluations, these systems can improve the reliability of diagnoses and support dermatologists in making informed decisions.

2.5.1 Machine Learning and Deep Learning Application in Skin Cancer Detection

ML and DL techniques have revolutionized skin cancer detection, offering substantial enhancements in diagnostic accuracy, especially when applied to dermoscopic images. Recent studies have demonstrated the potential of these advanced computational methods to augment clinical decision-making and improve patient outcomes. Monika, Arun Vignesh, Usha Kumari, Kumar, and Lydia (2020) developed a machine learning-based system for skin cancer detection and classification using dermoscopic images. The preprocessing stage involved using the Dull Razor method to remove hair and Gaussian and Median filters for image enhancement and noise reduction. The segmentation was performed using color-based k-means clustering. Feature extraction utilized the ABCD method and Gray Level Co-occurrence Matrix (GLCM) to capture statistical and texture features. The classification was conducted using a Multi-class Support Vector Machine (MSVM), achieving an accuracy of 96.25% on the ISIC 2019 Challenge dataset. This study highlights the effectiveness of combining various preprocessing, segmentation, and feature extraction techniques with MSVM for accurate skin cancer detection. Kassem, Hosny, Damaševičius, and Eltoukhy (2021) conducted a comprehensive review of machine learning and deep learning methods for skin lesion classification and diagnosis. The study analyzed 53 articles utilizing traditional machine learning techniques, including those based on the ABCDE rule, and 49 articles employing deep learning approaches, assessing their contributions, methodologies, and outcomes. The findings underscore that deep learning approaches outperform traditional machine learning, particularly when large datasets are available or when data limitations are addressed through augmentation techniques. Pre-trained deep learning models, combined with handcrafted methods based on deep learning, have demonstrated promising results, achieving high-precision accuracy in melanoma detection. The research conducted by Gouda, Sama, Al-Waakid, Humayun, and Jhanjhi (2022) explores the application of deep learning techniques for the detection of skin cancer using skin lesion images. This study utilizes convolutional neural networks (CNNs) to classify skin lesions as either benign or malignant, using the ISIC2018 dataset. The innovative aspect of this study is the use of Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) for image preprocessing, which enhances image quality

before classification. The CNN model achieved an accuracy of 83.2%, while transfer learning models like Resnet50, InceptionV3, and Inception Resnet showed slightly higher accuracies of 83.7%, 85.8%, and 84% respectively. These findings suggest that deep learning models, particularly when combined with advanced preprocessing techniques, can significantly improve the accuracy of skin cancer detection from lesion images. Shimizu, Iyatomi, Celebi, Norton, and Tanaka (2015) proposed a novel computer-aided method for classifying four types of skin lesions: melanoma, nevi, basal cell carcinomas (BCCs), and seborrheic keratoses (SKs). Their approach utilized a layered model with task decomposition strategy, outperforming flat models in classification accuracy. The study employed 964 dermoscopy images and extracted 828 features categorized into color, subregion, and texture. Notably, the layered model with 25 features achieved a 90% detection rate for melanoma, 82.51% for nevi, 82.61% for BCC, and 80.61% for SK. This research is significant for its comprehensive approach to multi-class skin lesion classification, addressing both melanocytic and non-melanocytic lesions, which is crucial for enhancing the capabilities of computer-aided skin cancer detection systems. The research by Ali, Shaikh, Khan, and Laghari (2022) explores the application of EfficientNets in multiclass skin cancer classification, utilizing the HAM10000 dataset. Their study demonstrates the effectiveness of EfficientNet models, particularly EfficientNet B4, which achieved an F1 Score of 87% and a Top-1 Accuracy of 87.91%. The authors highlight the significance of transfer learning and fine-tuning pre-trained ImageNet weights to enhance model performance on this imbalanced dataset. The findings suggest that intermediate complexity models like EfficientNet B4 offer a balanced approach, outperforming both simpler and more complex variants in handling the nuanced task of skin cancer classification. This study underscores the potential of EfficientNets in advancing automated diagnostic systems in dermatology.

2.5.2 Machine Learning and Deep Learning Applications in Melanoma Detection

Machine learning algorithms have significantly advanced the field of skin cancer detection, offering substantial improvements in diagnostic accuracy and efficiency. These algorithms have the capability to analyze complex patterns in medical images, thereby aiding in the early detection and classification of skin lesions. Jafari et al. (2016) developed an efficient system for the automatic detection of melanoma using digital images captured by general-purpose cameras, including

smartphones. The method involves preprocessing to reduce noise and illumination effects, followed by segmentation using k-means clustering in the HSV color space. Morphological operations and guided filtering are applied to enhance the segmentation mask and accurately extract the lesion's border. The system then extracts ten features based on the ABCD rule of dermatology, focusing on asymmetry, border irregularity, and color attributes. The extracted features are classified using a Support Vector Machine (SVM), achieving a diagnostic accuracy of 90%, with an area under the curve (AUC) of 0.9794. This approach demonstrates the potential of using advanced image processing techniques and feature extraction to improve melanoma detection accuracy in a computationally efficient manner suitable for smartphone applications. Kruk et al. (2015) developed a novel system for melanoma recognition using dermoscopic images, employing an extended set of image descriptors and classifiers. The study utilized descriptors such as Haralick, Kolmogorov-Smirnov, and fractal texture analysis to generate diagnostic features. Feature selection methods, including Fisher discrimination, correlation feature selection, and fast correlation-based filter, were applied to enhance classification accuracy. The SVM classifier, combined with Fisher-selected features, achieved the highest accuracy of 93.8%, with a sensitivity of 95.2% and specificity of 92.4%. This approach demonstrates the potential of advanced image descriptors and robust feature selection in improving melanoma detection accuracy. Gil et al. (2023) proposed an ensemble of classifiers for medical image recognition, leveraging deep learning techniques. The study examined various CNN architectures for recognizing melanoma in dermoscopic images and breast cancer in mammograms. Two ensemble approaches were introduced: one combining feature selection methods with classical classifiers like SVM, Random Forest (RF), and softmax, and another using diverse CNN structures directly. The ensemble system integrating multiple classifiers via majority voting significantly improved classification performance. The best results were achieved using the deep ensemble approach, with an accuracy of 98.6% and an AUC of 0.9996 for melanoma recognition. This demonstrates the efficacy of combining different CNN architectures and feature selection methods to enhance diagnostic accuracy in medical imaging.

In the context of skin cancer detection, deep learning algorithms, particularly CNNs, have shown remarkable performance. CNNs are especially effective in image classification tasks, making them well-suited for analyzing dermoscopic images and identifying potential malignancies Esteva et al.

(2017); Yu, Chen, Dou, Qin, and Heng (2017). CNNs, with their deep learning capabilities, have been instrumental in automating the detection and classification of skin cancers. Their architecture, designed to mimic the human visual system, excels in extracting hierarchical features from dermoscopic images, offering substantial improvements over manual diagnostic methods. In the evolving landscape of melanoma diagnosis utilizing deep learning, Belattar, Adjadj, Bakir, and Ait Mehdi (2022) conducted a comprehensive comparison of seven CNN models, including Baseline CNN, InceptionV3, ResNet50, VGG16, Xception, MobileNetV2, and DenseNet201, for melanoma versus nevus classification using a balanced subset of the ISIC 2019 dataset. The study is notable for its thorough evaluation of different architectures, showing that models like Baseline CNN, DenseNet201, and Xception outperformed others. Interestingly, both simpler models like the Baseline CNN and more complex ones such as DenseNet201 achieved high performance, suggesting that model effectiveness is not solely dependent on architectural complexity. Hosseinzadeh Kassani and Hosseinzadeh Kassani (2019) presented a comprehensive analysis comparing several advanced CNNs for melanoma detection using dermoscopic images. By employing diverse preprocessing steps and data augmentation techniques like horizontal and vertical flipping to address class imbalance, this study sought to enhance image quality and model accuracy. Through their experimental evaluation, they assessed the performance of state-of-the-art CNN architectures. The study demonstrated that preprocessing and data augmentation could significantly improve detection accuracy, underscoring the potential of deep learning in early and accurate melanoma diagnosis. Kim, Gaibor, and Haehn (2024) developed a comprehensive melanoma classification framework that integrates 54 combinations of 11 datasets and 24 advanced deep learning models, resulting in 1,296 experimental comparisons. Their study introduces Mela-D, a lightweight model optimized for web deployment, which achieves a 33x increase in processing speed and a 24x reduction in parameters while maintaining an accuracy of 88.8%, comparable to ResNet50. This model is designed for efficient, real-time melanoma detection on consumer-grade hardware. The study by Faghihi, Fatholahi, and Rajabi (2024) presents an innovative approach to melanoma skin cancer detection using customized transfer learning models based on VGG16 and VGG19 architectures. The researchers trained their models on a subset of 2,541 skin lesion images from the International Skin Imaging

Collaboration (ISIC) dataset, employing dropout and early stopping techniques to prevent overfitting. Notably, using k-fold cross-validation, the models achieved average accuracies of 97.51% for VGG16 and 98.18% for VGG19, demonstrating significant improvement over existing approaches without resorting to data augmentation techniques. Mousa, Taha, Kaur, and Afifi (2024) conducted a comprehensive evaluation of five pre-trained deep learning models for melanoma classification using the International Skin Imaging Collaboration (ISIC) dataset. The study compared VGG-16, ResNet50, InceptionV3, DenseNet-121, and Xception models across four experiments with varying hyperparameters and layer configurations. The research is notable for its systematic approach to optimizing model performance through transfer learning and hyperparameter tuning. ResNet50 consistently outperformed other models, achieving exceptional accuracy and F1 scores of around 93% in the third experiment. This study highlights the potential of deep learning in enhancing melanoma diagnosis. The research by Yu et al. (2017) presents a novel approach to automated melanoma recognition using very deep CNNs to address challenges such as low contrast, visual similarities between lesions, and artifacts in dermoscopic images. Their method integrates a fully convolutional residual network (FCRN) for lesion segmentation and a deep residual network (DRN) for classification. The authors demonstrate that substantially deeper networks (more than 50 layers) can acquire richer and more discriminative features, leading to improved recognition accuracy. Their proposed FCRN, which integrates multi-scale contextual information, achieves first rank in classification and second first in classification and second in segmentation tasks performance in both segmentation and classification tasks on the ISBI 2016 Skin Lesion Analysis Towards Melanoma Detection Challenge dataset. This study highlights the potential of very deep CNNs in addressing complex medical image analysis tasks, even with limited training data.

2.5.3 Usage of Vision Transformers in Melanoma and Non-Melanoma Skin Cancer Detection

Despite the excellent performance of CNNs, the advent of ViT has introduced a paradigm shift in the field, particularly in the context of distinguishing between closely resembling skin lesions. Unlike traditional CNNs, which rely on local receptive fields and hierarchical feature extraction, ViT utilizes self-attention mechanisms to capture global contextual information within images. This

ability to process global relationships among visual features is crucial for tasks that require a comprehensive understanding of the entire image, such as differentiating between skin lesions that may appear similar Deininger et al. (2022); Pu, Xi, Yin, Zhao, and Zhao (2024). Yang, Luo, and Greer (2023) introduced a novel ViT-based model for skin cancer classification using clinical skin images. The proposed method involves four key stages: class rebalancing, image preprocessing, transformer encoding, and classification. The model was pretrained on the ImageNet dataset and fine-tuned using the HAM10000 dataset. The ViT model processes images by splitting them into patches, which are then flattened and passed through a transformer encoder. The best configuration of the ViT model achieved a classification accuracy of 94.1%, outperforming the current state-of-the-art model IRv2 with soft attention. This approach demonstrates the efficacy of Vision Transformers in enhancing diagnostic accuracy for skin cancer classification by effectively modeling long-range spatial relationships within images. Several studies have explored the use of ViT for melanoma skin cancer detection. The study by Arshed et al. (2023) underscores the growing importance of ViT models in skin cancer classification. Their research highlights the superiority of pre-trained vision transformers in achieving an accuracy of 92.14%, surpassing CNN-based transfer learning models across several evaluation metrics for skin cancer diagnosis. This comparative analysis suggests that while traditional CNNs remain valuable, ViTs offer a compelling alternative that could redefine the landscape of medical image classification, especially in the context of skin cancer detection. The paper by Flosdorf, Engelker, Keller, and Mohr (2024) explores the application of ViTs in the classification of skin lesion images, highlighting their potential to improve diagnostic accuracy in skin cancer detection. The study compares two configurations of pre-trained ViTs, ViT_L32 and ViT_L16, against traditional models like tree classifiers and k-nearest neighbors, as well as CNNs. The ViT L16 model achieved an accuracy of 92.79%, while ViT L32 reached 91.57%. Despite these high accuracy rates, the recall for melanoma detection was lower, with ViT L16 and ViT L32 achieving 56.10% and 58.54% recall, respectively. These findings highlight the potential of ViTs in enhancing diagnostic accuracy in medical imaging. Cirrincione et al. (2023) introduced a ViT-based model for melanoma detection using dermoscopic images. The study utilized the ISIC 2017 Challenge dataset to train and evaluate the model, focusing on classifying melanoma versus non-cancerous lesions. The proposed ViT architecture effectively models long-range spatial relationships within images,

enhancing classification performance. The best configuration of the ViT model achieved an accuracy of 94.8%, with a sensitivity of 92.8%, specificity of 96.7%. This approach demonstrates the potential of ViTs in improving diagnostic accuracy for melanoma detection. Xie, Wu, Zhu, and Zhu (2021) introduced a novel Swin-SimAM network for melanoma detection using dermoscopic images. This approach integrates the Swin Transformer with the SimAM attention module to enhance feature extraction and focus on critical parts of skin lesions. The model addresses class imbalance using focal loss, which reduces the impact of non-melanoma classes. Experiments conducted on the ISIC-2017 dataset demonstrated that the Swin-SimAM network achieved superior performance, with an accuracy of 90.0% and an AUC of 0.900. This study highlights the effectiveness of combining Swin Transformer and SimAM for improving melanoma detection accuracy.

2.6 Transfer Learning in the Medical Domain

A pivotal aspect of the evolution of AI in medical imaging is the exploration of TL techniques, which use pre-existing models trained on large datasets to improve the accuracy and efficiency of models in medical diagnostics. The integration of TL techniques with CNNs and ViT, employing pre-trained models, has further enhanced their diagnostic accuracy, mitigating the challenges posed by the limited number of annotated medical datasets and the computational intensity of training models from scratch. Menegola et al. (2017) explored transfer learning with deep learning for melanoma screening, evaluating the use of models pre-trained on ImageNet and Retinopathy. They focused on the benefits of fine-tuning and using deeper models for enhanced diagnostic performance. Their findings highlighted a preference for deeper, fine-tuned models trained on ImageNet, demonstrating significant improvements in accuracy with AUC scores of up to 84.5% on skin lesion datasets. This study underscores the potential of utilizing transfer learning to improve melanoma detection through deep learning. Shamshiri, Krzyzak, Kowal, and Korbicz (2023) have extended the scope of transfer learning research by applying a compatible-domain transfer learning approach for the classification of breast cancer. They proposed the use of a pre-trained model on a histopathological image dataset for the classification of breast cancer cytological biopsy samples. By utilizing a dataset that is more compatible with the target medical domain, they demonstrated that features

learned during the pre-training phase are more relevant and thus significantly improve the model's accuracy. Their findings indicate an impressive enhancement in classification accuracy—by 6% to 17% over traditional machine learning techniques and about 7% over deep learning methods that did not employ compatible-domain transfer learning, achieving up to 98.73% validation accuracy and 94.55% test accuracy. Inspired by the work of Shamshiri et al. (2023), our research seeks to explore the application of compatible-domain transfer learning to the task of melanoma cancer detection. We plan to employ a range of computational models, including CNNs, ViTs, and hybrid models, to investigate whether the pre-training on domain-relevant datasets can similarly improve the accuracy of melanoma classification. To the best of the authors' knowledge, this study represents the first attempt to apply compatible-domain transfer learning to melanoma detection, potentially pioneering a new direction in the field and setting a precedent for future research.

2.7 Summary

In summary, this chapter is organized into two primary sections. The first section of this chapter provides a comprehensive overview of cancer, with a focus on skin cancer, including a detailed examination of melanoma. It covers melanoma's characteristics, global statistics, risk factors, and diagnostic methods. The second section reviews recent advancements in CAD systems for melanoma detection. It highlights the application of ML, DL, and ViT in enhancing diagnostic accuracy. Additionally, this section highlights the necessity of using TL techniques in medical image analysis to improve the accuracy and efficiency of diagnostic models. It also reviews different TL approaches, such as domain-specific transfer learning in medical analysis. All the reviewed research confirms that CAD systems utilizing CNNs, ViTs, and TL approaches can significantly assist dermatologists in detecting melanoma accurately and quickly.

Chapter 3

Methodology

3.1 Introduction

In this chapter, we thoroughly explain the methodology used in our research on melanoma cancer diagnosis. To thoroughly value our proposed approach and the most accurate model for melanoma detection, we took a series of essential steps. Section 3.2 provides an overview of our proposed method, outlining the strategic framework we developed for the comparison between different approaches in this study, we will explain each step separately. In section 3.3, we delve into the datasets selected for this study, explaining the criteria and reasoning behind their choice. This section focuses on the target melanoma dataset used for fine-tuning and also elaborates on the three distinct datasets used during the pre-training phase, highlighting their characteristics. In section 3.4, we describe the preprocessing steps undertaken, including how the data was divided into training, validation, and test subsets. Lastly, section 3.5 presents the deep learning architectures utilized, detailing their implementation and relevance to our study.

3.2 Overview

This research thoroughly evaluates and compares seven distinct deep learning models for the binary classification of melanoma cancer using a private annotated dataset of melanoma dermoscopic

images. The models under investigation include VGG16, Inception-V3, MobileNet, two variations of the ViT model (ViT-B16 and ViT-B32), and two innovative hybrid models—one combining ViT-B16 with VGG16, and the other integrating ViT-B16 with Inception-V3. Our methodology is implemented through three carefully crafted scenarios, each designed to assess the models' performance under different transfer learning conditions. Transfer learning is crucial for melanoma cancer detection due to the limited size of our target melanoma dataset. The following sections provide a detailed explanation of each scenario.

- **First scenario**: The first scenario investigates the models' performance utilizing transfer learning from the ImageNet dataset, a widely used resource. This scenario explores how transfer learning from a large, general-purpose dataset can impact our models' accuracy. All seven models are initially pre-trained on ImageNet and then meticulously fine-tuned on our targeted melanoma dataset. This phase is crucial for evaluating the models' capacity to adapt and apply transfer learning from a large and general dataset to a specific medical imaging task.

- **Second scenario**: In the second scenario, we shift to a more specialized domain for transfer learning, utilizing the BreakHis dataset, which is focused on breast cancer histopathology. The models are pre-trained on the BreakHis dataset before fine-tuning on the target melanoma dataset. The goal of this scenario is to evaluate the effectiveness of transfer learning between distinct medical domains, investigating whether the knowledge acquired from breast cancer histopathology can be used to improve melanoma diagnosis.

- **Third scenario**: For the third and final scenario, we apply domain-specific transfer learning for Melanoma detection. In this scenario, we pre-train our models on a subset of the ISIC 2019 dataset, which shares the dermatological focus of the target dataset, before fine-tuning with the target Melanoma dataset. This scenario is designed to assess the impact of transfer learning within the same medical domain on the precision and efficiency of melanoma detection.

Fig. 3.1 illustrates the difference between the typical TL approach for classifying melanoma

cancer by using ImageNet dataset and our proposed TL approaches. The typical TL approach highlights the effectiveness of using well-known datasets like ImageNet to enhance binary classification results. To build on this, we propose two methods: medical-domain and domain-specific transfer learning approaches. For the medical-domain approach, we selected the BreakHis dataset, which focuses on breast cancer histopathology. This choice explores whether knowledge from a distinct medical domain can improve melanoma diagnosis accuracy. For the domain-specific approach, we utilized a custom-made subset of the ISIC 2019 dataset. This strategy investigates whether learning from the same domain can enhance melanoma diagnosis. To evaluate these proposed TL strategies, we will assess the performance of the seven DL models across all three TL approaches. Our ultimate goal is to identify the most effective model and transfer learning approach for melanoma binary classification. This comprehensive methodology allows us to evaluate each model's potential and determine the best strategy for improving melanoma diagnosis. All models were developed in Python using the Keras and TensorFlow libraries.

The method pipeline for addressing the binary classification of melanoma cancer is illustrated in Fig. 3.2. Our proposed TL approaches are integrated into this pipeline. We utilized three datasets—ImageNet, BreakHis, and ISIC 2019—for pre-training our models. In the first step, we prepared a subset of the ISIC 2019 dataset for binary classification and selected images with 400X magnification from the BreakHis dataset. Next, we applied data preprocessing steps, including normalization, augmentation, and resizing, to the subsets from BreakHis and ISIC2019, as well as our target melanoma dataset. In the third step, we pre-trained the models using ImageNet, BreakHis, and ISIC 2019. Finally, we fine-tuned the models with our target melanoma dataset to perform the binary classification task. In this study, we applied complete fine-tuning, updating the weights of all layers throughout the process.

3.3 Datasets

This section provides an in-depth review of each dataset used in this research and its characteristics. Furthermore, we explain the reasoning behind each dataset selection. This study is structured around two key phases: pre-training and fine-tuning. Accordingly, the datasets are categorized into

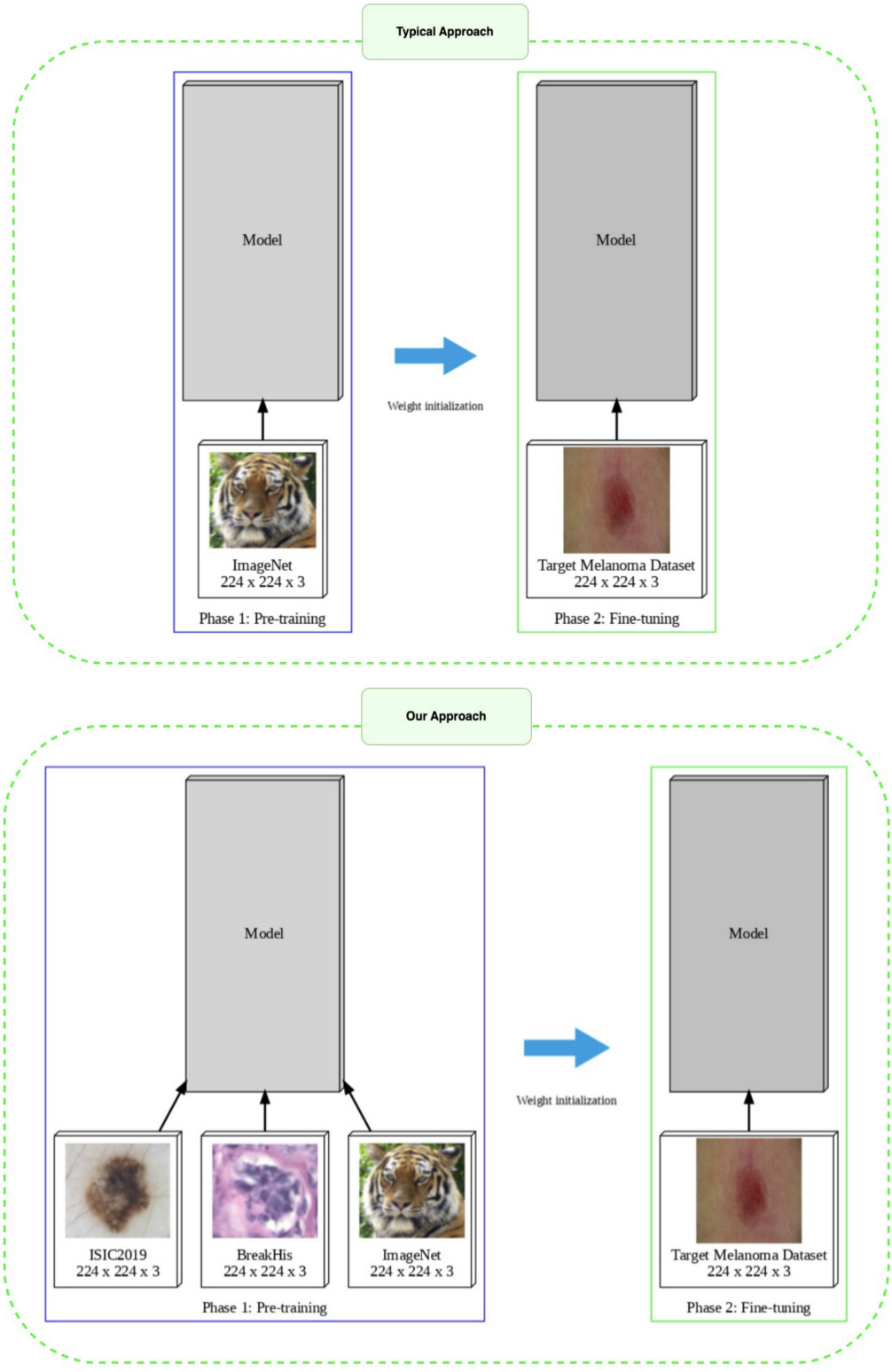

Figure 3.1: Illustrates both the typical transfer learning approach and our transfer learning approach.

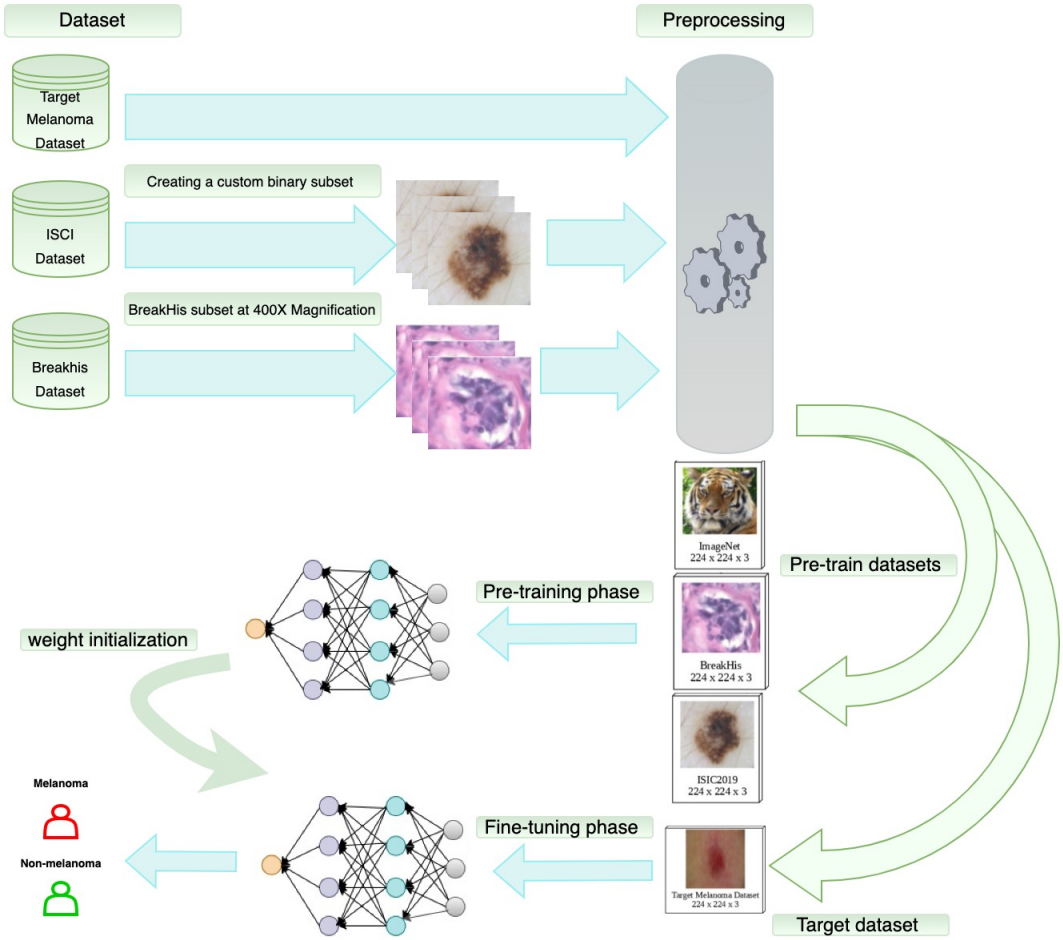

Figure 3.2: Method pipeline for binary classification of melanoma cancer. The process involves pre-training models using ImageNet, BreakHis (400X magnification), and a custom-made subset of ISIC 2019 datasets, followed by fine-tuning with the target melanoma dataset.

pre-training datasets, used during the initial phase, and the target dataset, employed in the fine-tuning stage. The pre-training phase utilizes three diverse datasets—ImageNet, BreakHis, and ISIC 2019— each originating from a different domain, which lays the groundwork for our model development. The target melanoma dataset, on the other hand, is specifically used in the fine-tuning phase to refine the model's accuracy.

3.3.1 Target Dataset

The melanoma dataset was developed at the Warsaw Maria Skłodowska-Curiebreak National Research Institute of Oncology, Department of Soft Tissue/Bone Sarcoma and Melanoma Gil et al. (2023). It features 112 RGB images of verruca seborrhoica as non-melanoma examples and 134 images of basal cell carcinoma categorized as melanoma. These images were captured at a 20× magnification using dermoscopy from various body areas. Expert dermatologists annotated the target dataset using the ABCDE criteria, with confirmation by exact pathomorphological inspection. The images, stored in JPEG format, were taken at different times and with various resolutions, leading to a range in image sizes from 767×576 to 4273×2848 pixels. Fig. 3.10 shows some examples of melanoma and non-melanoma cases in the target dataset.

Figure 3.3: Sample images from the melanoma dataset. From left to right, the first three samples (a,b, and c) belong to non-melanoma cases, and the next three samples belong to melanoma cases.

3.3.2 Pre-train Datasets

ImageNet

ImageNet is renowned for its extensive collection of approximately 14 million high-resolution images representing over 21,000 categories. This dataset contains a diverse range of general images, from common items and animals to more abstract concepts, offering a rich visual foundation Deng

et al. (2009). The versatility of ImageNet makes it an ideal option for TL, allowing us to fine-tune our models on a more specialized melanoma dataset, which contains a limited amount of annotated data. By using the wide variety and large number of images in ImageNet, our models become better at generalizing, making it a valuable resource for the initial training phase. In Fig. 3.4, you can find samples of ImageNet images.

Figure 3.4: Sample images from the ImageNet dataset.

BreakHis

BreakHis dataset is a specialized collection of histopathological images for breast cancer research. This dataset is divided into two primary categories: benign and malignant. The benign category consists of 2,480 samples, while the malignant category comprises 5,429 samples. Each sample is available at four magnification levels: 40X, 100X, 200X, and 400X. Moreover, this dataset includes eight subcategories, with the benign and malignant categories each further classified into four distinct types. The benign category contains adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA), and the malignant category subgroups are ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC). In this dataset, each image has three RGB channels and a dimension of 700×460 pixels Spanhol et al. (2016). For the purpose of this study, we selected a subset of 1,820 images for analysis, including

588 benign and 1,232 malignant samples, all captured at a magnification of 400X. The BreakHis dataset, like our target melanoma dataset, originates from the medical domain but represents a different form of cancer. By using the BreakHis dataset for the initial training of our models, we are investigating the efficacy of transfer learning across related yet distinct medical domains; we aspire to use domain-specific insights derived from breast cancer histopathology for improving melanoma detection. Fig. 3.5 shows a sample from each sub-category in this dataset.

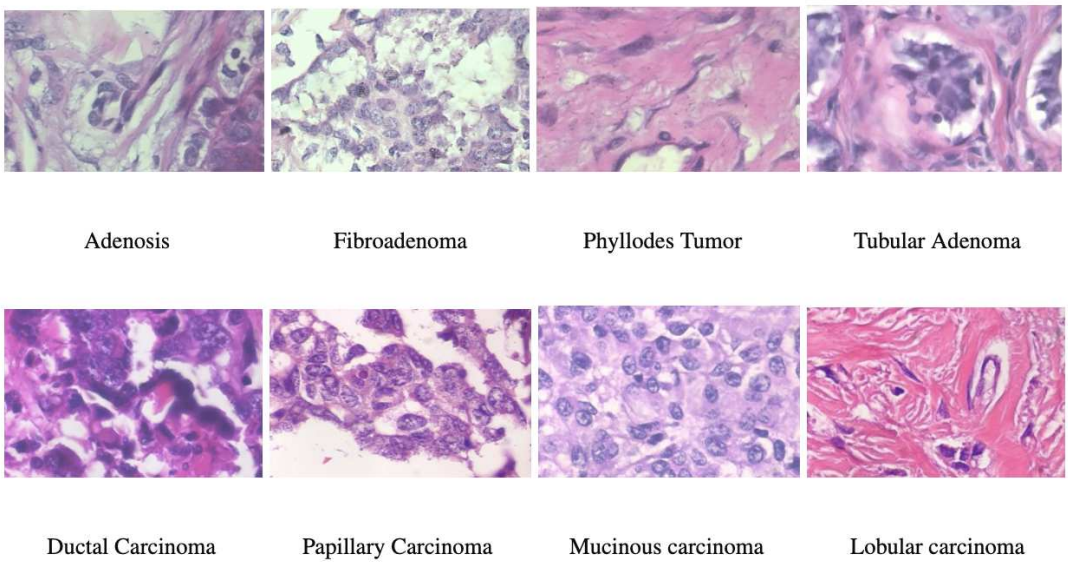

Figure 3.5: Example images from the BreakHis dataset. Adenosis, Fibroadenoma, Phyllodes Tumor, and Tubular Adenoma are benign while Ductal Carcinoma, Papillary Carcinoma, Mucinous carcinoma, and are malignant.

ISIC 2019

The ISIC 2019 challenge dataset is a widely used dataset in dermatological diagnosis, including a vast collection of 25,331 dermoscopic images. This dataset consists of nine diagnostic categories: Melanoma (MEL), Melanocytic nevus (NV), Basal cell carcinoma (BCC), Actinic keratosis (AK), Benign keratosis (BKL), Dermatofibroma (DF), Vascular lesion (VASC), Squamous cell carcinoma (SCC), and an unspecified category (UNK) Codella et al. (2017); Hernandez et al. (2024); Tschandl et al. (2018). Given the computational demands of training on the entire dataset, we selected six categories for our study. The Melanoma (MEL) category was chosen due to its direct relevance to

our research focus in melanoma detection. We grouped together BKL, AK, DF, VASC, and SCC into a collective non-melanoma category. BKL was included as it aligns with the non-melanoma lesions present in our target dataset. Additionally, to maintain a balance between melanoma and non-melanoma samples and to introduce a variety of skin lesion types into our pre-training dataset, we included AK, DF, VASC, and SCC. Consequently, our dataset comprises 4,522 samples categorized as melanoma and 4,611 samples categorized as non-melanoma. The reason behind utilizing the ISIC dataset for pre-training lies in assessing the potential benefits of domain-specific transfer learning. Specifically, we investigate how TL from a dataset within the same domain can affect the model's performance when subsequently fine-tuned on our designated target dataset for melanoma recognition. Fig. 3.6 presents images from the ISIC dataset, providing an example for each category.

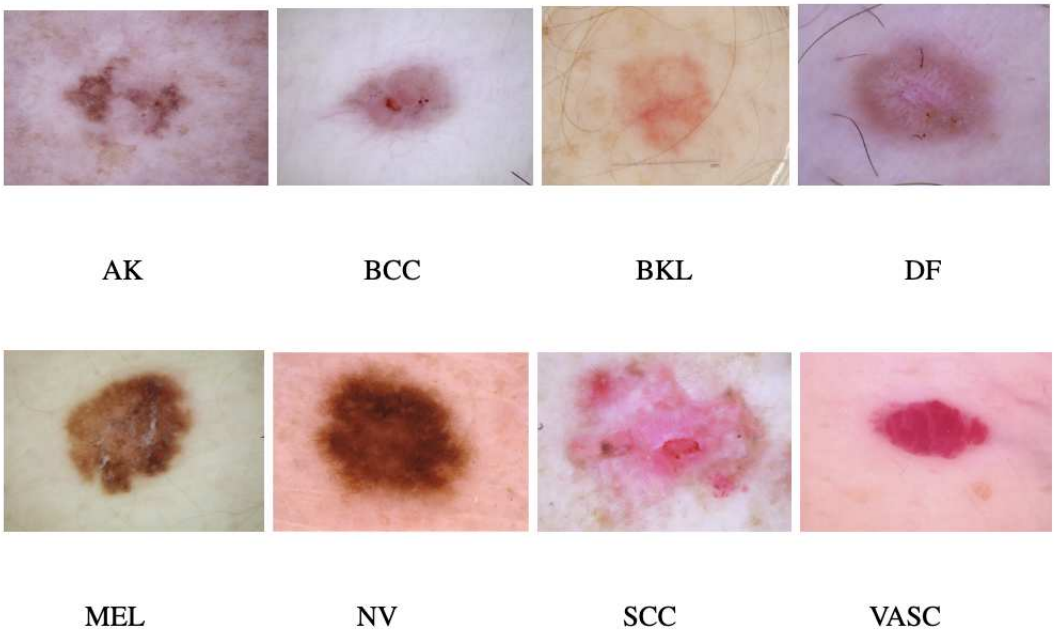

Figure 3.6: Representative images from the ISIC dataset illustrating the nine categories: Actinic keratosis (AK), Basal cell carcinoma (BCC), Benign keratosis (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic nevus (NV), Squamous cell carcinoma (SCC), Vascular lesion (VASC), and an unspecified category (UNK).

3.4 Preprocessing

For this study, three preprocessing steps were implemented to enhance the models' performance and generalization ability. The following sections provide a detailed explanation of each step.

3.4.1 Normalization

In our approach, we used normalization for our dataset. We applied normalization to adjust the pixel values of images to a range of [0, 1], which is crucial for facilitating model convergence during the training phase. This process can be mathematically represented as follows

$$x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{1}$$

where x' represents the normalized value, x represents the original pixel value, x_{min} represents the minimum pixel value in the dataset, and x_{max} represents the maximum pixel value in the dataset.

3.4.2 Augmentation

Recognizing the importance of model robustness and generalization, we utilized a data augmentation step exclusively for the training data in our datasets. We first split the data using a stratified sampling approach to ensure that each subset maintained the same class distribution as the original dataset. Specifically, we divided the dataset into training, validation, and test sets. Initially, we allocated 20% of the data to the test set. The remaining 80% of the data was then split again, with 20% of this subset assigned to the validation set and the remaining 80% used for training. After splitting the data, we applied data augmentation techniques exclusively to the training set to enhance the model's generalization capabilities. For each training example, we generated three augmented examples using a variety of transformations. These transformations included rotation, width and height shifts, shear, zoom, and horizontal flips. These augmentations artificially expand the variety of training data, simulating real-world variations in image orientation, scale, and perspective. This approach significantly enhances the model's ability to generalize from the training data to unseen images.

3.4.3 Resizing

One of the common challenges in medical image analysis is diversity in image sizes. This is particularly evident in our target melanoma dataset, which contains dermatoscopic images of varying dimensions. To address this issue, a critical preprocessing step involves resizing the images to ensure they meet our models' input requirements. For this purpose, all images are standardized to a resolution of 224×224 pixels. This step is vital for facilitating batch processing and ensuring that the network is trained on a consistent input format, thereby reducing the likelihood of bias towards specific image sizes.

3.5 Deep Learning Models

Deep learning has significantly impacted the field of skin cancer detection, particularly melanoma, by enabling more accurate and early diagnosis. These models replicate the neural networks of the human brain, allowing them to learn complex patterns from extensive datasets. This capability is crucial for identifying subtle differences in skin lesions that may indicate melanoma. In our study, we evaluated the performance of seven DL models: VGG16, Inception-V3, MobileNet, two variations of the ViT model (ViT-B16 and ViT-B32), and two innovative hybrid models—one combining ViT-B16 with VGG16, and the other integrating ViT-B16 with Inception-V3. Each model offers unique advantages in terms of architecture and feature extraction capabilities, which are essential for handling the variability in skin lesion images. Our primary objective is to identify the most precise model for melanoma detection. The following sections provide detailed explanations of each model's architecture and performance, highlighting their contributions to improving melanoma detection. By comparing these models, we aim to advance the development of automated systems that can assist dermatologists in diagnosing melanoma more accurately and efficiently.

3.5.1 VGG16

Introduced by Simonyan and Zisserman (2015), the VGG16 model represented a major leap forward in the field of computer vision during the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). It set itself apart from earlier models like AlexNet by implementing several

key innovations. One of the primary advancements of the VGG16 model was its use of multiple 3 x 3 convolutional filters, replacing the larger 11 x 11 and 5 x 5 filters used in previous architectures. This change allowed for a reduction in computational complexity by decreasing the number of parameters, while simultaneously enhancing the model's ability to capture intricate features from input images. By adopting these smaller filters, VGG16 was able to construct a deeper network, extending its architecture to between 16 and 19 layers compared to the 8 layers present in AlexNet. This increased depth allowed the model to learn more complex patterns and hierarchical representations, which significantly improved its performance in image classification tasks. The uniform structure of VGG16, characterized by consistent use of convolutional layers followed by max-pooling layers, contributed to its effectiveness and ease of implementation. Moreover, VGG16 utilized Rectified Linear Unit (ReLU) activation functions, which played a crucial role in addressing the vanishing gradient problem, thereby facilitating faster and more efficient training of the deep network. This combination of architectural choices made VGG16 a foundational model in the realm of deep learning, influencing subsequent research and development in the field. Its design principles have been widely adopted and adapted in various applications, underscoring its lasting impact on the advancement of convolutional neural networks.

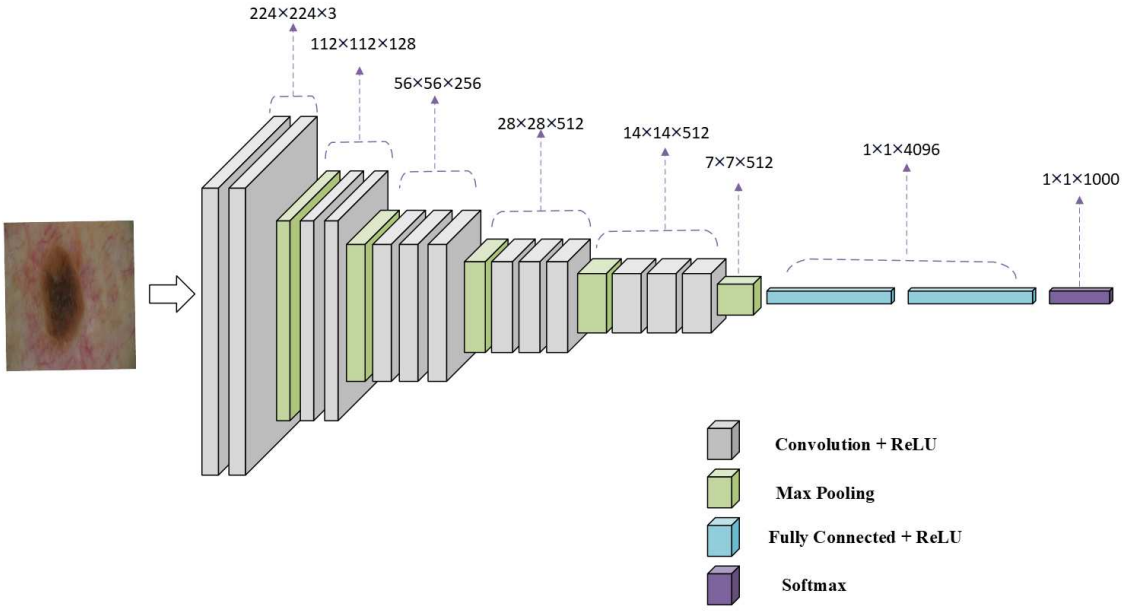

Figure 3.7: VGG16 Architecture.

3.5.2 Inception-V3

Inception-V3, introduced by Szegedy et al. (2016), marks a significant advancement in the development of Inception networks. This architecture is renowned for its innovative design, which improves computational efficiency while maintaining high accuracy in image classification tasks. The model is distinguished by the integration of Inception modules, which consist of parallel convolutional layers with filters of varying sizes, such as 1×1 and 3×3, along with max-pooling layers. Notably, Inception-V3 replaces the 5×5 convolutions used in earlier versions with two consecutive 3×3 convolutions to reduce computational complexity. These modules enable the network to perform diverse convolution operations efficiently, optimizing parameter usage and minimizing memory consumption.

Inception-V3 incorporates several improvements over its predecessors. One such enhancement is the use of factorized convolutions, which break down larger convolutions into smaller, more manageable operations. This approach reduces computational cost and improves the model's ability to generalize from the training data. Additionally, Inception-V3 employs batch normalization, which helps stabilize the learning process and accelerates convergence. Another significant feature of Inception-V3 is the use of label smoothing, a regularization technique that reduces overfitting by preventing the model from becoming too confident about its predictions. This technique contributes to improved generalization performance on unseen data. The architecture also includes an auxiliary classifier, which provides additional gradient signals during training, further aiding the learning process. Overall, Inception-V3's design reflects a careful balance between model complexity and computational efficiency, making it a powerful tool for image classification tasks. Its architectural innovations have set a benchmark in the field of deep learning, influencing subsequent research and development in convolutional neural networks.

3.5.3 MobileNet

MobileNet, introduced by Howard et al. Howard et al. (2017) in their 2017 paper titled "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" represents a

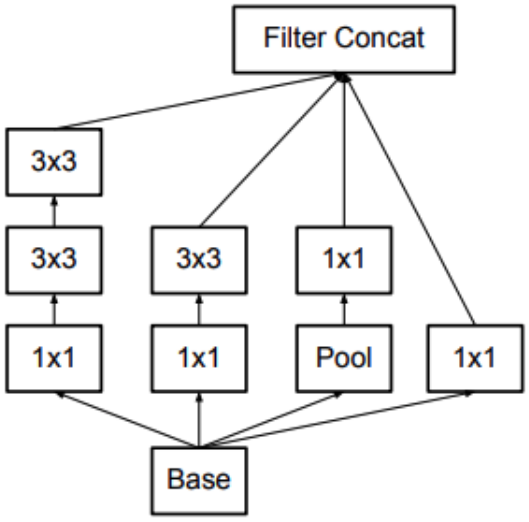

Figure 3.8: Inception modules where each 5×5 convolution is substituted with a pair of 3×3 convolutions. Source: Szegedy et al. (2016)

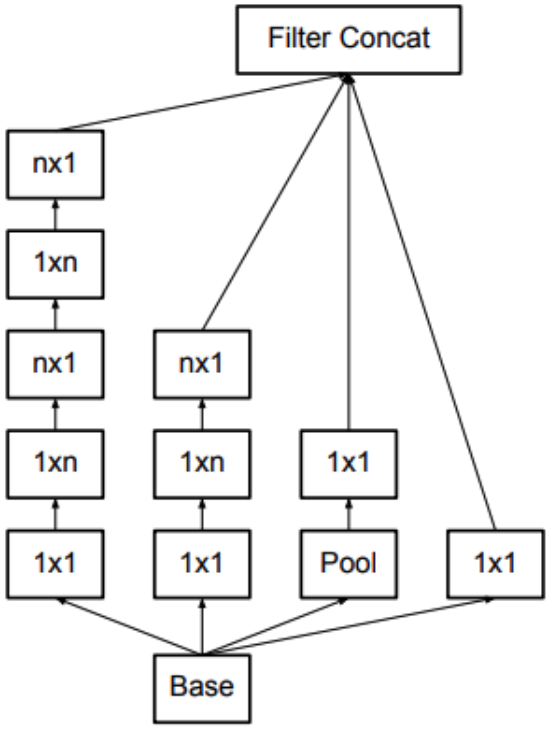

Figure 3.9: Inception modules following the factorization of the n×n convolutions. Source: Szegedy et al. (2016)

notable advancement in neural network design specifically tailored for mobile and embedded devices. This architecture is outstanding for its use of depthwise separable convolutions, which drastically reduce the model's size and computational demands while maintaining robust performance. Depthwise separable convolutions decompose the standard convolution process into two distinct stages: a depthwise convolution, which applies a single filter to each input channel independently, followed by a pointwise convolution that combines these outputs through a 1×1 convolution.

This method significantly lowers the computational burden and the number of parameters, making MobileNet highly efficient for applications where computational resources are limited. Moreover, MobileNet introduces a width multiplier, a hyperparameter that allows the model to trade-off between latency and accuracy, providing flexibility for deployment on devices with varying capabilities. The architecture also includes a resolution multiplier, which adjusts the input image resolution, further enabling control over the trade-off between speed and accuracy. These innovations make MobileNet particularly well-suited for real-time applications on mobile and embedded platforms, where efficiency is paramount. Its design principles have influenced a range of subsequent models, underscoring its impact on the development of lightweight neural networks.

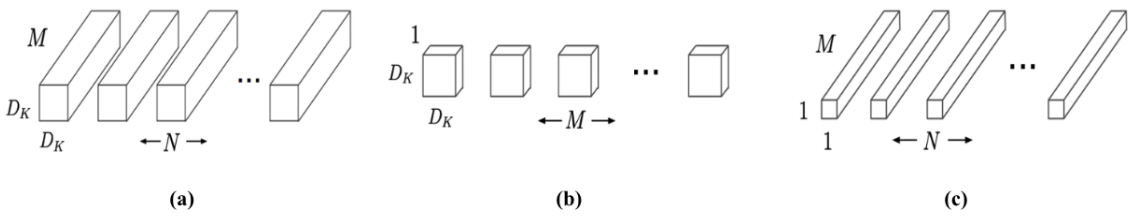

Figure 3.10: The standard convolutional filters shown in (a) are substituted with two layers: (b) depthwise convolution and (c) pointwise convolution, which together create a depthwise separable filter. Source: Howard et al. (2017); Wang et al. (2020)

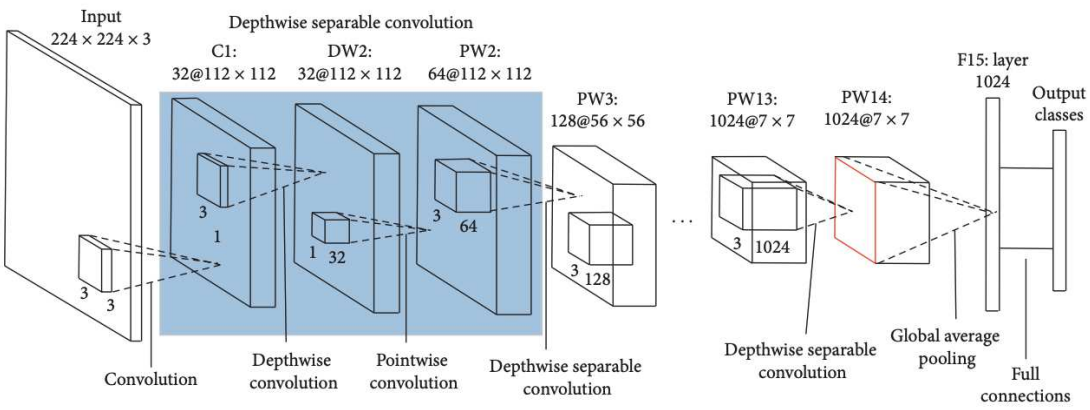

Figure 3.11: MobileNet Architecture. Source: Wang et al. (2020)

3.5.4 Vision Transformers

In our study, we employed the ViT architecture, which represents a substantial shift from traditional CNNs. The ViT architecture introduces a groundbreaking approach by leveraging self-attention mechanisms, a technique originally designed for natural language processing, and adapting it for image classification tasks. This innovative method involves dividing images into smaller segments, referred to as patches, which allows the ViT to evaluate and prioritize the significance of various regions within an image. This approach enables a more detailed and context-aware examination of visual information, allowing the model to capture complex and intricate relationships within the data Dosovitskiy et al. (2020). The Vision Transformer operates by first converting each image into a sequence of flattened patches, similar to the way words are processed in a sentence. Each patch is then linearly embedded and combined with positional embeddings to retain spatial information. These embeddings are processed through a series of transformer layers, which apply self-attention to model the dependencies between patches. This mechanism allows the ViT to focus on different parts of the image dynamically, enhancing its ability to recognize patterns and features across the entire image.

In our research methodology, we implemented two distinct configurations of the Vision Transformer: ViT-B16 and ViT-B32. The ViT-B16 configuration divides images into 16×16 pixel patches, while the ViT-B32 configuration processes images using larger 32×32 pixel patches. These configurations allow for flexibility in balancing computational efficiency with the level of detail captured

from the images. By employing these configurations, our study aims to explore the effectiveness of the Vision Transformer in capturing complex patterns and relationships in visual data, contributing to advancements in melanoma detection.

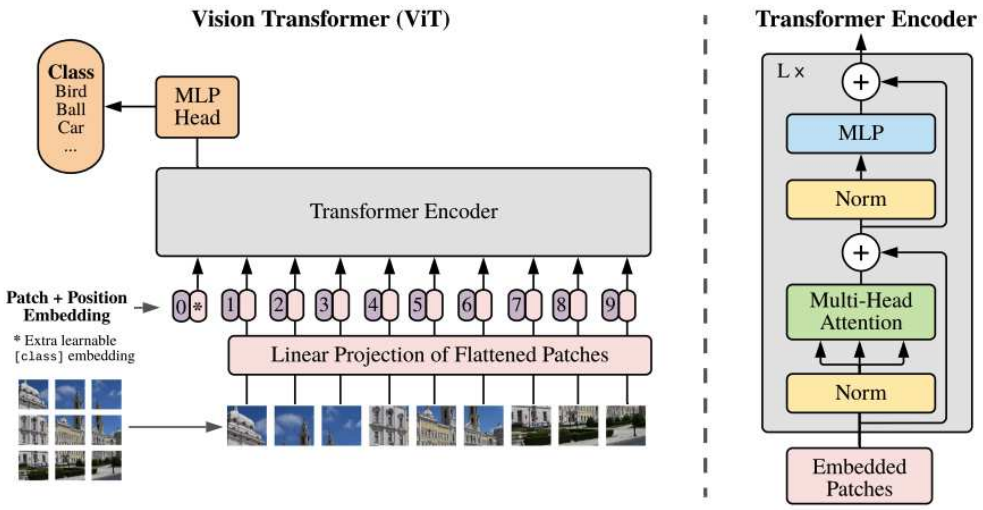

Figure 3.12: Vision transformers model overview. Source: Dosovitskiy et al. (2020)

3.5.5 Hybrid models

In our research, we developed two hybrid models to explore their effectiveness in detecting melanoma. The goal of creating these models is to combine the strengths of CNNs with ViT architectures, using the advantages of both approaches. By integrating the feature extraction capabilities of CNNs with the self-attention mechanisms of ViTs, these hybrid models aim to enhance diagnostic accuracy and robustness. These models are part of a broader set of architectures in our comparative analysis. We will assess their performance against other well-established models to validate their diagnostic capabilities and determine their potential benefits in melanoma detection.

ViT-B16-VGG16

The first hybrid model in our study, ViT-B16-VGG16, represents an innovative integration of the ViT-B16 and VGG16 architectures. This model is designed to harness the strengths of both architectures, combining the sophisticated self-attention mechanism of ViT-B16, known for its effectiveness in image classification tasks, with the robust feature extraction capabilities of VGG16. The ViT model's self-attention mechanism allows it to focus on different parts of the image, capturing complex relationships and dependencies effectively.

To adapt these architectures for our binary classification task, we made specific modifications. We removed the top layers of both models to customize them for our needs. The outputs from each model were then flattened to create a unified feature set. These features were subsequently merged and passed through a dense layer with ReLU activation, which enhances the model's ability to learn complex patterns by introducing non-linearity. To mitigate the risk of overfitting, a dropout layer was incorporated, which randomly deactivates a portion of the neurons during training. The final layer of the model employs a sigmoid activation function, which is ideal for binary classification tasks.

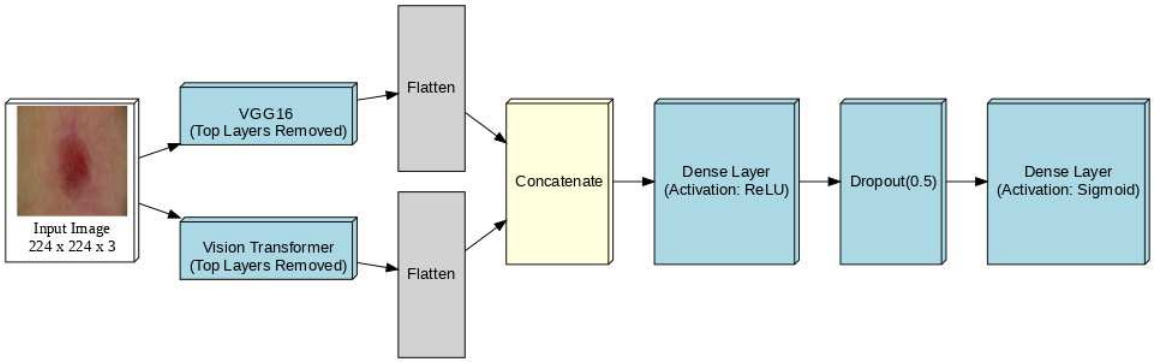

Figure 3.13: ViT-B16-VGG16 Model architecture.

ViT-B16-Inception-V3

In this study, we present ViT-B16-Inception-V3 as our second hybrid model, designed to merge the capabilities of Inception-V3 with the ViT to improve image analysis. Inception-V3 is celebrated for its ability to efficiently and accurately process visual information, employing a network

of convolutional layers and inception modules that effectively extract key features from images. On the other hand, the ViT is adept at identifying intricate patterns using its self-attention mechanism, which enables it to concentrate on various segments of an image and discern complex relationships within the data.

To develop this hybrid model, we started by eliminating the top layers of both Inception-V3 and ViT, customizing them for our binary classification objective. The outputs from these architectures were then flattened and integrated to create a unified feature representation. This combined output was fed into a dense layer with ReLU activation, which adds non-linearity and boosts the model's capacity to learn sophisticated features. To reduce the likelihood of overfitting, a dropout layer was incorporated, which randomly disables a portion of neurons during training, enhancing the model's robustness and ability to generalize. The last layer of the ViT-B16-Inception-V3 model uses a sigmoid activation function, which is particularly suited for binary classification tasks as it outputs probabilities that can be interpreted as class predictions.

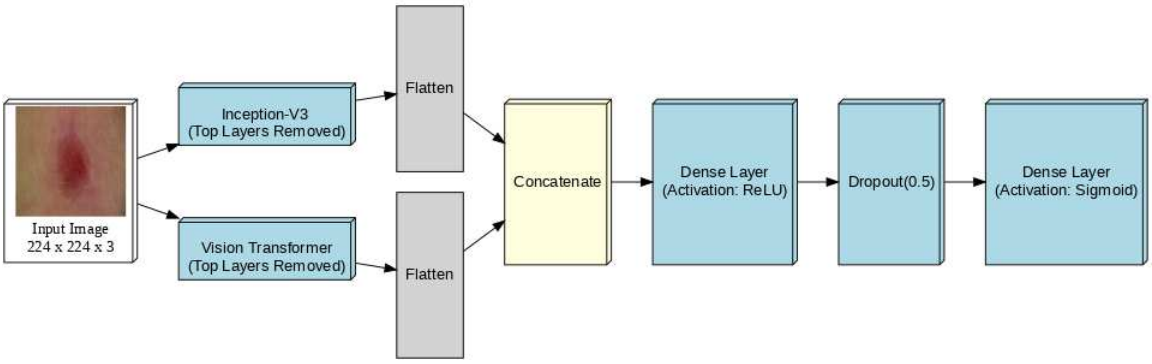

Figure 3.14: ViT-B16-Inception-V3 model architecture.

3.6 Summary

In this chapter, we presented a detailed explanation of the methodology utilized in this study. Our research focused on comparing three distinct TL techniques for melanoma binary classification, each implemented in a separate scenario. To evaluate these TL techniques, we utilized seven DL models, carefully explaining each scenario and its corresponding TL approach. We used real-world

dermoscopic melanoma images to ensure an effective and realistic evaluation of these scenarios and models. Section 3.2 presents a detailed pipeline of our melanoma detection process, outlining all the steps involved in our study. This section also includes a comparison of the different TL approaches we investigated. In section 3.3, we discuss the datasets used for both pre-training and fine-tuning stages, providing insight into the data that formed the foundation of our research. Then in section 3.4, we have meticulously described all processing steps undertaken in this study. Furthermore, Section 3.5 offers an in-depth explanation of each deep learning model employed in our research. This includes a discussion of their architectures and key features.

Chapter 4

Experimental Results

4.1 Introduction

This chapter presents the results from our comprehensive study on melanoma detection using various DL models and TL techniques. Our research aimed to evaluate the effectiveness of different approaches in accurately classifying melanoma from dermoscopic images. As mentioned earlier, we explored three distinct TL scenarios, each implemented with seven DL models, to identify the most effective method for this crucial medical imaging task. In the following sections, we will present our experimental results and provide a thorough analysis of each model's performance across the different scenarios. We will examine key metrics like accuracy, loss, precision, and recall to evaluate how well the models detect melanoma. Additionally, we will compare our results with recent studies that used the same dataset. By sharing these findings, we hope to contribute valuable insights to the field of automated melanoma detection and lay the groundwork for future research in this area.

4.2 Results

In this section, we analyze the outcomes of melanoma detection using seven widely recognized deep learning architectures: Inception-V3, VGG16, MobileNet, and two versions of the Vision Transformer, ViT-B16 and ViT-B32, and two hybrid models, ViT-B16-VGG16 and ViT-B16-Inception-V3. This study, along with certain parts of its findings, has been published in Haghshenas

et al. (2024).

In our research, we employed a 5-fold cross-validation technique to ensure a comprehensive and reliable assessment of our models. This method allowed us to maximize the use of our dataset for training, validation, and testing purposes. In our implementation of the 5-fold cross-validation, we divided our dataset into five segments. Our training process involved five iterations, where we used four segments for training and the remaining segment for testing. We added an additional step by randomly allocating 20% of the training data for validation in each iteration. This approach ensures that every data sample serves as a test sample once and contributes to the training process four times. The 5-fold cross-validation method proved particularly valuable in evaluating the consistency of our models' performance and their ability to generalize to novel data. By implementing this technique, we aimed to minimize overfitting risks and achieve a more accurate estimation of our models' capabilities. Furthermore, it enabled us to identify potential variations in our results, which is essential for assessing the robustness of our methodology.

Table 4.1 presents an overview of the advanced CNNs utilized across our three experimental scenarios. These models are arranged in ascending order based on their parameter count, providing insight into their relative complexity. We selected a diverse range of CNNs, varying in complexity and depth, to identify the most effective model for melanoma detection. Additionally, Table 4.2 provides a comprehensive overview of the ViT models examined in this study. To further explore the potential of combining different architectures, we also investigated two hybrid models: ViT-B16-VGG16 and ViT-B16-Inception-V3. Each of these architectures brings unique strengths to the task of melanoma detection. By comparing their performance across various metrics, we aimed to identify the most effective approach for this critical medical imaging task. Detailed descriptions of all these architectures are provided in the methodology chapter.

Table 4.1: The details of the cutting-edge CNNs investigated in our research.

Model	Parameters	Depth	Size
MobileNet	4.2 M	28	16 MB
Inception-V3	23.9 M	48	92 MB
VGG16	138.4 M	16	528 MB

Table 4.2: The details of the Vision Transformer models we used in this study.

Model	ViT Variant	Patch Size	Parameters	Layers	Heads	Hidden Size	MLP Size
ViT-B16	ViT-Base	16×16	86M	12	12	768	3072
ViT-B32	ViT-Base	32×32	86M	12	12	768	3072

Table 4.3 presents a detailed overview of the experimental configuration and training parameters utilized in this study. This comprehensive table includes essential information such as batch sizes, learning rates, and optimization techniques implemented across our various model architectures. To ensure consistency and facilitate meaningful comparisons, we maintained a uniform experimental setup for all models throughout our research. We selected the Adam optimizer for its adaptive learning rate capabilities, with an initial learning rate of 10^{-5}. This choice allowed for efficient optimization across different network structures. For all training iterations, we employed binary cross-entropy as the loss function, which is particularly well-suited for our binary classification task of distinguishing between melanoma and non-melanoma cases. To mitigate the risk of overfitting, we implemented an early stopping mechanism with a patience of 10 epochs. Additionally, we incorporated a learning rate scheduler that reduced the rate by 10% every 10 epochs. This gradual refinement of the learning rate allowed for more precise optimization as training progressed. The models underwent training for up to 20 epochs, with the training process executed on GPU-accelerated environments to enhance computational time.

Table 4.3: The experimental settings employed.

Parameters	Values
Optimizer	Adam
Learning Rate	10^{-5}
Loss Function	Binary Cross-Entropy
Batch Size	32
Epochs	20
Early Stopping Patience	10
Execution Environment	GPU

4.2.1 Performance Metrics

To evaluate our models' performance, we employed a diverse set of performance metrics that are standard in medical image analysis research. These carefully chosen metrics offer a comprehensive view of each model's capabilities, allowing for a thorough examination of their effectiveness in melanoma detection tasks. By utilizing these well-established evaluation metrics, we ensure a balanced analysis of our models' performance, providing insight into various aspects of their performance. The performance metrics we used are accuracy, area under the curve(AUC), precision, and recall. In the following sections, we offer a detailed explanation of each metric

Accuracy:

This metric reflects the ratio of correctly identified cases to the total cases evaluated, offering a general overview of the model's effectiveness.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Area Under the Curve (AUC):

This metric assesses the model's ability to discriminate between classes, with an AUC of 1.0 indicating perfect classification and 0.5 suggesting no discriminative power.

Precision:

Precision calculates the proportion of true positive findings out of all positives predicted by the model, indicating the reliability of positive classifications. The precision score ranges from 0 to 1, where a score of 1 indicates that all positive predictions were correct, and a score of 0 indicates that all samples labeled as positive are negative.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

Recall:

Recall evaluates the percentage of true positives, highlighting its ability to detect all true instances. The range of recall is from 0 to 1. A recall score of 1 means the model correctly identified all positive cases. In contrast, a score of 0 means the model did not correctly identify any positive cases.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

F1-score:

The F1-score represents the harmonic mean of precision and recall, providing a balanced measure of model performance. This metric is particularly useful for imbalanced datasets, as it considers both false positives and false negatives. The F1-score ranges from 0 to 1, with 1 indicating optimal performance.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

4.2.2 Attention Visualization in Vision Transformer Architecture

To gain a deeper understanding of the decision-making process in the ViT model, we generated attention map plots for a selection of melanoma and non-melanoma images from our target dataset. Using the ViT-Keras library in Python, we visualized the regions within each input image that the model identified as most influential for classification. The process of creating these attention maps involves several key steps, utilizing the self-attention mechanism fundamental to transformer architectures. Initially, the input image is divided into a grid of fixed-size patches, which are then linearly embedded and combined with position embeddings. These embedded patches are processed through a series of transformer encoder layers, each containing a multi-head self-attention mechanism. To generate the attention map, we focus on the attention weights from the final layer of the transformer encoder, as these weights represent the model's assessment of each patch's relevance to the overall classification decision. The attention map generation process involves the following steps:

(1) Extraction of attention weights from each TransformerBlock in the model.

(2) Reshaping of the weights to separate layers, heads, and attention matrices.

(3) Averaging of weights across all attention heads to obtain a single set of attention scores.

(4) Addition of an identity matrix to the attention matrix to account for residual connections, followed by re-normalization of the weights.

(5) Recursive multiplication of weight matrices, starting from the last layer, to combine attention information from all layers.

(6) Extraction and resizing of the final attention map to match the original image dimensions.

The resulting attention map is then applied to the input image, with brighter areas corresponding to regions of higher importance and darker areas indicating less relevant regions. This visualization technique enables us to identify the areas of the image that the model considers most important for classification. Our analysis revealed that the ViT model successfully identified key features crucial for distinguishing between melanoma and non-melanoma cases. Fig. 4.1 presents three sample skin lesion images from our target dataset, with the top two samples representing non-melanoma cases and the third sample illustrating a melanoma case. These attention maps demonstrate the model's remarkable precision in locating the lesions and its ability to distinguish these lesions from non-relevant background elements, such as hair. This illustrates the model's proficiency in highlighting crucial diagnostic indicators while effectively filtering out irrelevant background elements. Beyond showcasing the model's precise decision-making, these visualizations offer valuable insights into its internal processes, providing essential interpretability crucial for medical applications. By examining these attention maps, dermatologists and researchers can not only validate the model's decisions against established diagnostic criteria for melanoma but also gain new insights into subtle patterns and features that may inform future diagnostic practices or research.

4.2.3 Results for the First Scenario

In our first experimental scenario, we evaluated the performance of models pre-trained on ImageNet and subsequently fine-tuned on our private annotated melanoma dataset. This approach

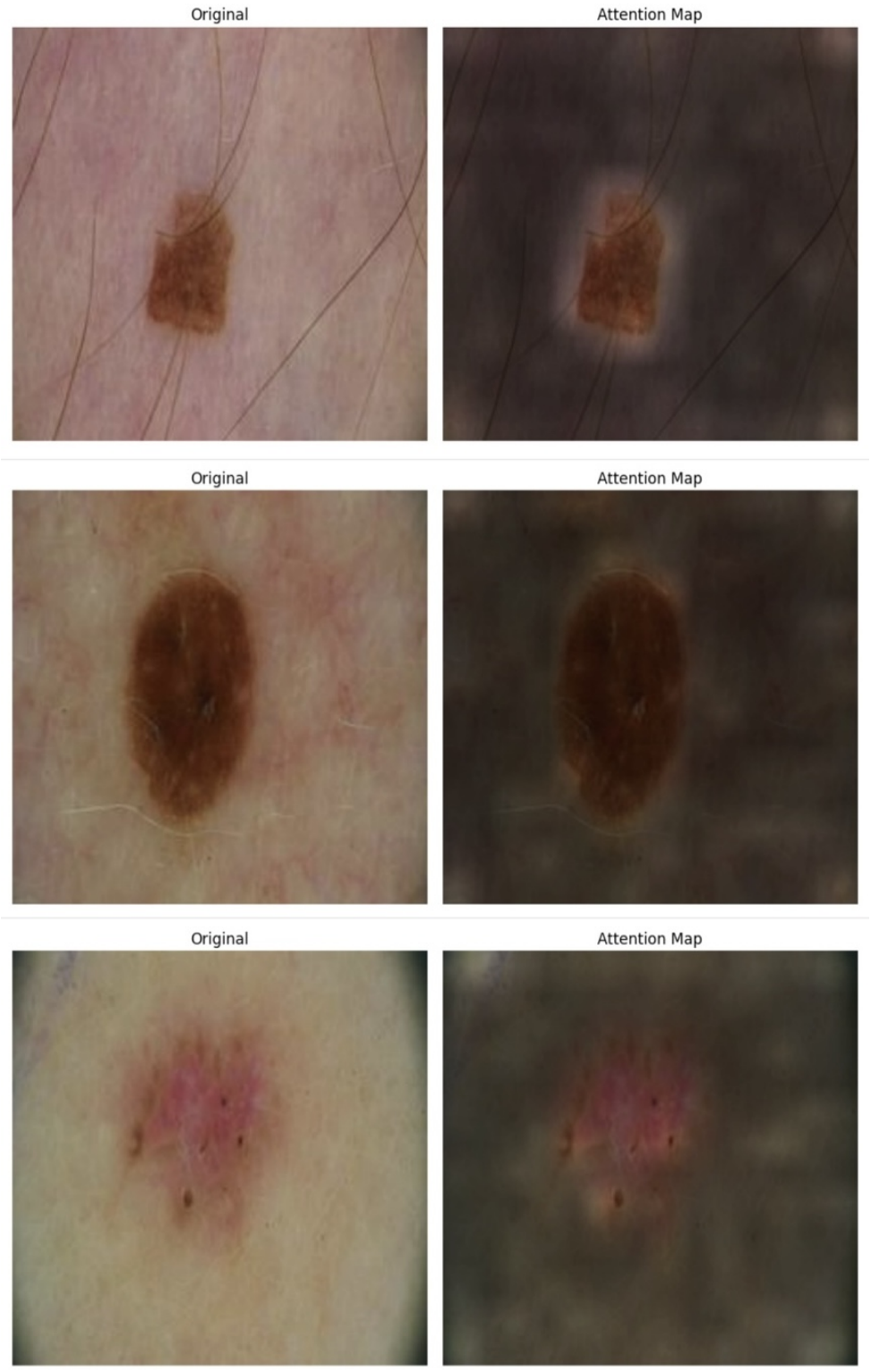

Figure 4.1: Attention Visualization in Vision Transformer Architectures: The top two samples represent non-melanoma cases, while the third sample illustrates a melanoma case.

allowed us to use the general feature extraction capabilities developed on a large-scale dataset and adapt them to our specific melanoma detection task.In this study, we employed a complete fine-tuning strategy. In this approach, we adjusted all layers of the networks, allowing the models to fully adapt to the characteristics of melanoma images. This comprehensive fine-tuning process was applied consistently across all architectures to ensure a fair comparison.To ensure the robustness of our results, we utilized k-fold cross-validation. This technique provides a more reliable estimate of model performance by reducing the impact of data partitioning bias. The results presented in Table 4.4 reflect the average performance across all folds, offering a comprehensive view of each model's capabilities. Among the various architectures tested, the Vision Transformer, ViT-B16, emerged as the top-performing model in this scenario. It achieved an impressive average accuracy of 97.97% across the k-fold cross-validation, surpassing the performance of other models, including traditional CNNs. The performance metrics for all models are detailed in Table 4.4, providing a comprehensive overview of the outcomes in this first scenario. Additionally, Fig. 4.3 illustrates the progression of accuracy and loss metrics over the course of training epochs.

Table 4.4: The classification results of models pre-trained on the ImageNet and fully fine-tuned (all layers) on the private annotated melanoma dataset.

Model	Ave. Accuracy	AUC	Precision	Recall	F1-Score	Loss
ViT-B16	**97.97%**	**0.9925**	**0.9857**	**0.9772**	**0.9814**	**0.0976**
ViT-B16-Inception-V3	96.74%	0.9911	0.9712	0.9698	0.9705	0.1193
ViT-B16-VGG16	96.33%	0.9907	0.9519	0.9852	0.9683	0.1243
ViT-B32	95.76%	0.9974	0.9575	0.9653	0.9614	0.1623
VGG16	90.24%	0.9567	0.8993	0.9256	0.9123	0.2896
Inception-V3	87.39%	0.9469	0.8689	0.9100	0.8890	0.2796
MobileNet	85.49%	0.9492	0.9058	0.8212	0.8614	0.3132

4.2.4 Results for the Second Scenario

In our second experimental scenario, we explored the impact of medical domain pre-training on model performance. Table 4.5 presents the classification results for models initially pre-trained on histopathological images from the BreakHis dataset before being fine-tuned on our private annotated melanoma dataset. This approach allowed us to assess the potential benefits of using a more closely related pre-training dataset for melanoma detection. Maintaining consistency with our first scenario,

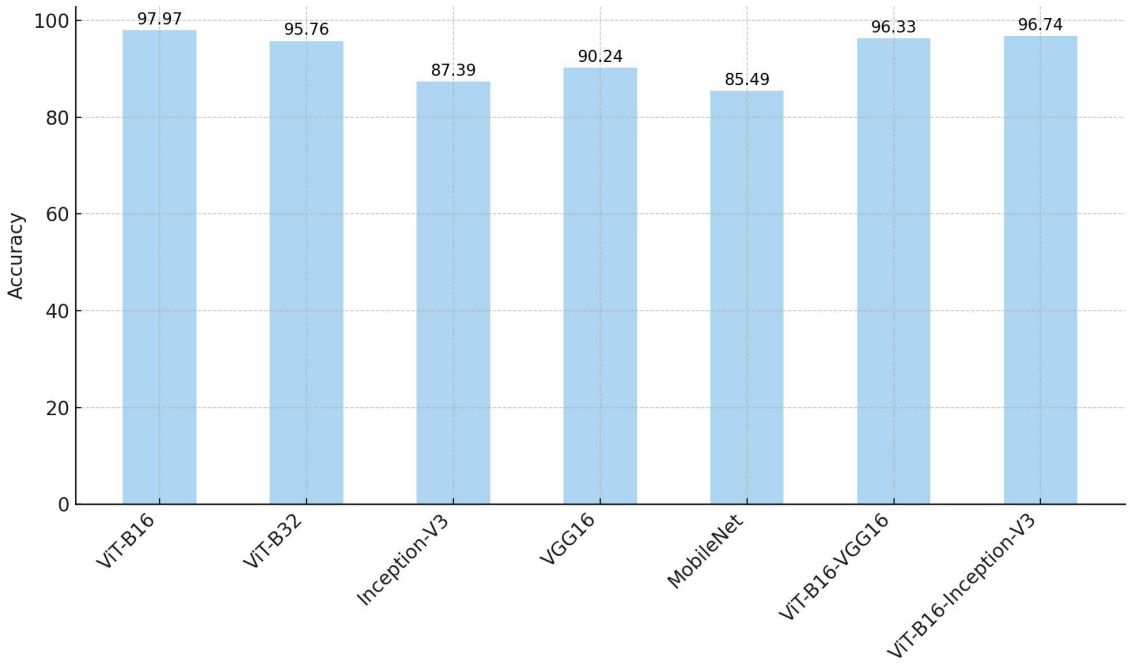

Figure 4.2: Accuracy of models (first scenario).

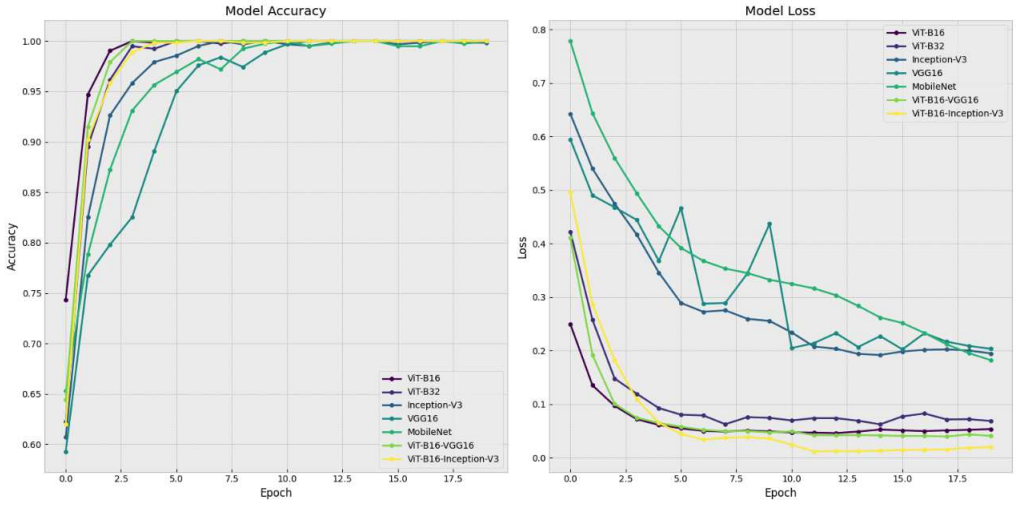

Figure 4.3: Loss and accuracy curves plotted against the number of epochs, corresponding to the training fold with the highest performance in the first scenario.

we applied complete fine-tuning to all model layers and employed 5-fold cross-validation to ensure result reliability. All the results are based on the unseen test set data, providing a robust evaluation of the models' generalization capabilities. This methodology ensures that our findings are both reliable and applicable to real-world scenarios. The results reveal interesting patterns in model performance. Once again, the Vision Transformer architecture, specifically the ViT-B16 model, demonstrated superior performance. It achieved an average accuracy of 86.59% across the 5-fold cross-validation, outperforming other tested models. The performance metrics for all evaluated models are detailed in Table 4.5, offering a comprehensive view of the results in this second scenario. Fig. 4.5 presents the evolution of accuracy and loss metrics across the training epochs for the second scenario. These graphs offer a visual representation of the models' learning processes, revealing important trends in their performance as training advances.

Table 4.5: The classification results of models pre-trained on BreakHis and fully fine-tuned (all layers) on the private annotated melanoma dataset.

Model	Ave. Accuracy	AUC	Precision	Recall	F1-Score	Loss
ViT-B16	**86.59%**	**0.9336**	**0.8857**	**0.8735**	**0.8796**	**0.3959**
ViT-B32	85.07%	0.9541	0.9145	0.8148	0.8618	0.5051
Inception-V3	84.59%	0.9176	0.9104	0.7991	0.8511	0.5271
ViT-B16-Inception-V3	84.58%	0.9028	0.9083	0.7974	0.8492	0.5138
ViT-B16-VGG16	83.33%	0.9308	0.8662	0.8444	0.8552	0.4130
MobileNet	80.18%	0.8661	0.8357	0.795	0.8148	0.4598
VGG16	77.62%	0.8565	0.8188	0.7530	0.7845	0.4973

4.2.5 Results for the Third Scenario

In our final experimental scenario, we investigated the impact of pre-training on a domain-specific dataset closely related to our target task. Table 4.6 presents the classification results for models initially pre-trained on the ISIC 2019 dataset before fine-tuning on our private annotated melanoma dataset. Consistent with our previous scenarios, we employed complete fine-tuning across all model layers and utilized 5-fold cross-validation to ensure robust evaluation. All reported results are based on unseen test set data, providing a reliable measure of the models' generalization capabilities. In this scenario, the Vision Transformer, ViT-B32, out performed other models, achieving an impressive average accuracy of 90.95% across the 5-fold cross-validation, surpassing

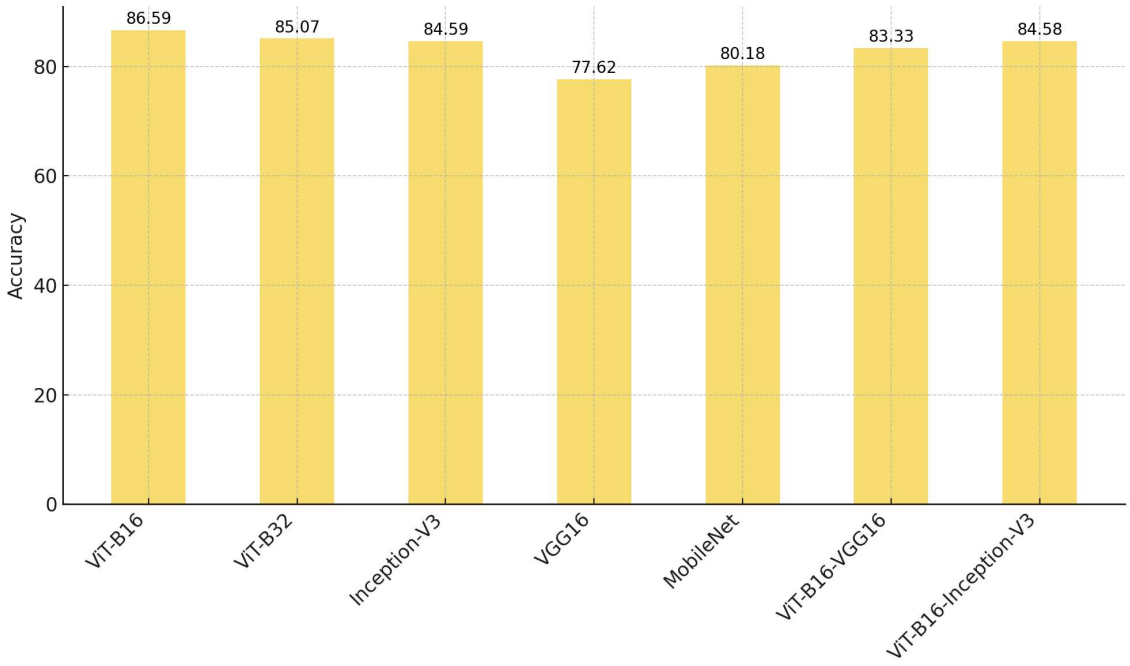

Figure 4.4: Accuracy of models (second scenario).

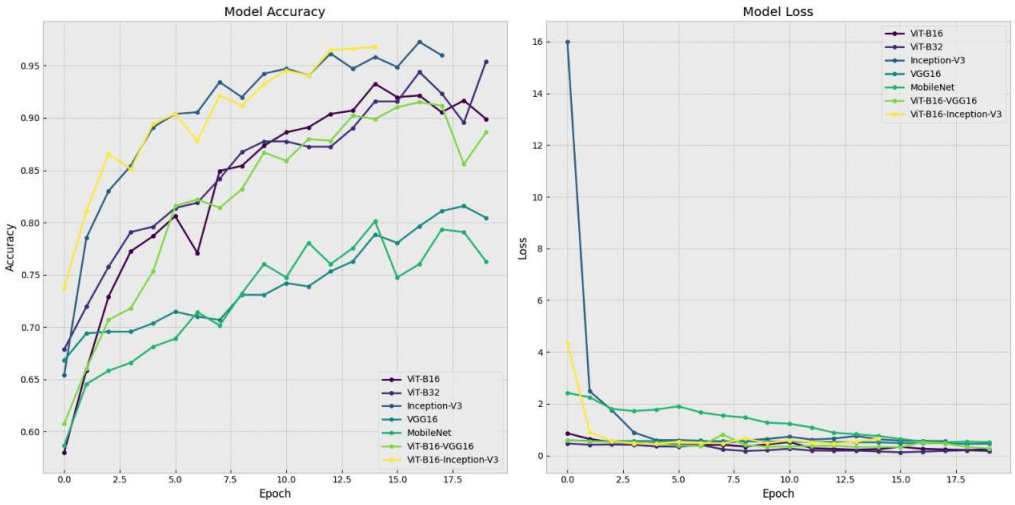

Figure 4.5: Loss and accuracy curves plotted against the number of epochs, corresponding to the training fold with the highest performance in the second scenario.

the performance of other tested architectures. Figure 4.7 illustrates the progression of accuracy and loss metrics over the training epochs for this scenario.

Table 4.6: The classification results of models pre-trained on the ISIC 2019 dataset and fully fine-tuned (all layers) on the private annotated melanoma dataset.

Model	Ave. Accuracy	AUC	Precision	Recall	F1-Score	Loss
ViT-B32	**90.95%**	**0.9548**	**0.9220**	**0.9169**	**0.9194**	**0.3835**
ViT-B16-VGG16	87.83%	0.9440	0.9257	0.8510	0.8868	0.3686
ViT-B16	87.81%	0.9345	0.9606	0.8140	0.8812	0.4782
Inception-V3	86.99%	0.9456	0.9266	0.8282	0.8746	0.3392
ViT-B16-Inception-V3	85.42%	0.9256	0.8761	0.8507	0.8632	0.4282
VGG16	81.27%	0.9082	0.9141	0.7308	0.8122	0.4485
MobileNet	73.58%	0.7983	0.8466	0.6366	0.7267	0.5860

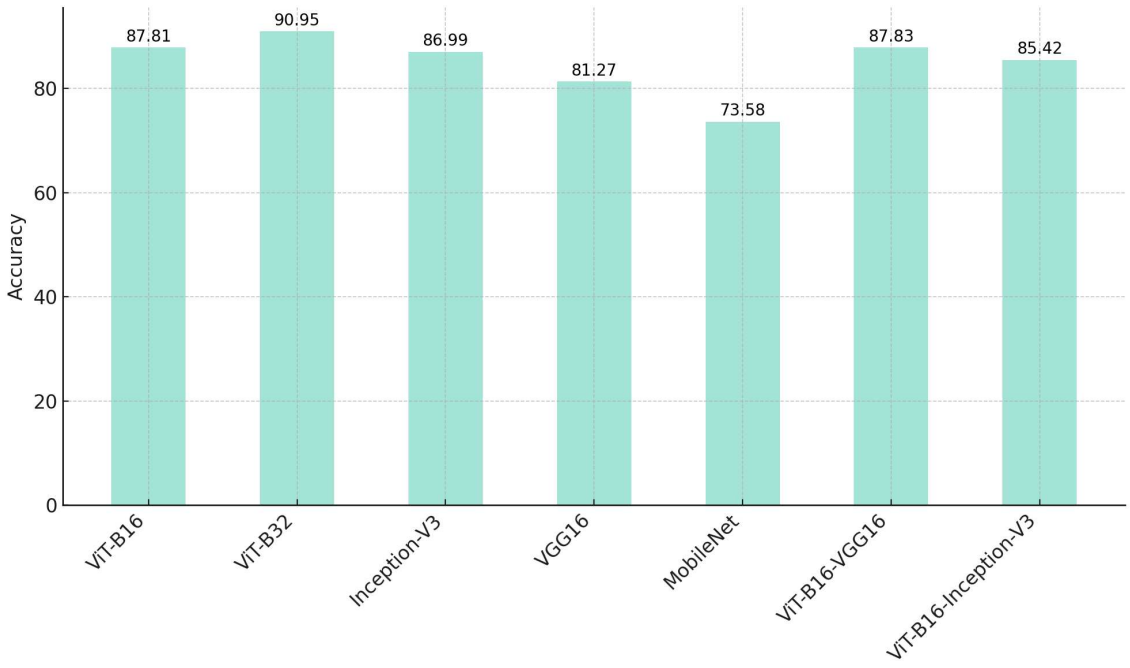

Figure 4.6: Accuracy of models (third scenario).

4.3 Discussion

In the field of medical image classification, particularly for melanoma detection, the limited availability of large-scale annotated datasets cause a significant challenge for training deep learning

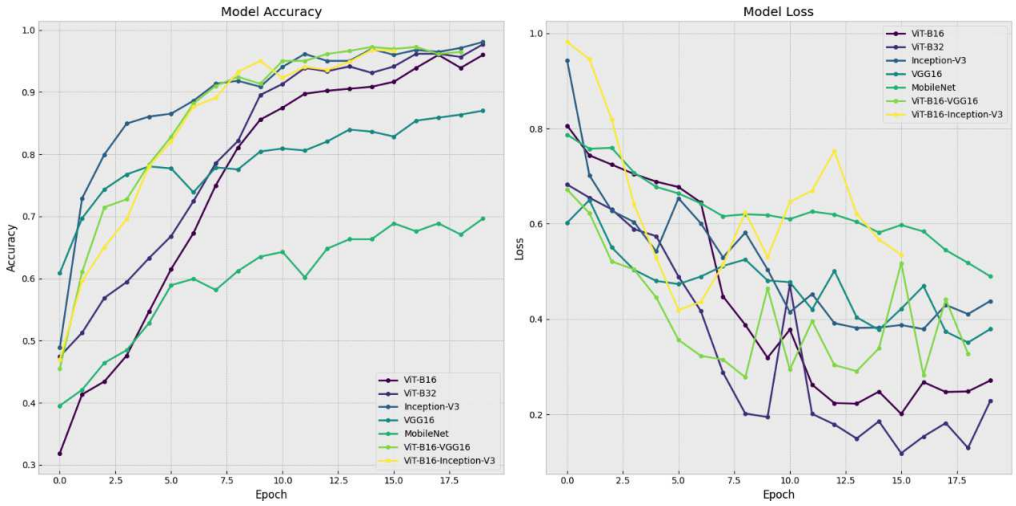

Figure 4.7: Loss and accuracy curves plotted against the number of epochs, corresponding to the training fold with the highest performance in the third scenario.

models from scratch. To address this limitation, transfer learning has emerged as a popular and effective strategy. Traditionally, models pre-trained on the ImageNet dataset have been widely adopted for various computer vision tasks, including medical image analysis. However, our study takes a novel approach by exploring the potential of domain-specific transfer learning in melanoma detection. We propose and evaluate two innovative transfer learning strategies using specialized medical datasets: the BreakHis dataset, and the ISIC 2019 dataset.

We conducted a comprehensive evaluation of seven deep learning models across three different TL approaches. This multifaceted methodology allows us to compare the effectiveness of different TL strategies and also facilitates the identification of the most effective model for melanoma diagnosis. The performance metrics presented in Table 4.8 offer significant insights into the efficacy of the three transfer learning approaches applied to melanoma detection. These approaches utilize models pre-trained on ImageNet, BreakHis, and ISIC 2019, providing a comprehensive comparison across different domains. The differences in pre-train dataset size, as outlined in Table 4.7, play a crucial role in the models' ability to generalize and perform well in the target melanoma classification task.

ImageNet, with its vast dataset size of 14 million samples, offers a highly diverse range of images, enabling models pre-trained on it to develop a broad understanding of visual patterns. This

Table 4.7: Number of samples in pre-training datasets.

Dataset	Number of samples
ImageNet	14 Million
BreakHis	2259
ISCI 2019(Custom-made subset)	9133

advantage is clearly reflected in the performance of models like ViT-B16, which achieved an average accuracy of 97.97%, an AUC of 0.9925, and precision of 0.9857. The extensive size of the ImageNet dataset likely enhances the generalization abilities seen in this scenario, especially for transformer-based models like ViT-B16. The hybrid models, such as ViT-B16-Inception-V3 and ViT-B16-VGG16, also performed well, indicating that combining Vision Transformers with traditional CNN architectures can be beneficial when using such a large-scale dataset. These findings confirm the effectiveness of using ImageNet pre-training for medical image analysis, particularly melanoma detection.

In contrast, the BreakHis dataset, with only 2,259 samples, represents a much smaller and more domain-specific dataset focused on breast cancer histopathological images. Despite its smaller size, pre-training on BreakHis allowed the ViT-B16 model to maintain competitive performance, achieving an accuracy of 86.59% and an AUC of 0.9336. The relatively smaller size of BreakHis likely limited the ability of certain models, such as VGG16, which only achieved 77.62% accuracy and an AUC of 0.8565. This suggests that while medical domain pre-training has its merits, it can also be constrained by the limited data available for certain diseases. Nonetheless, transformer-based models appear to benefit even from limited-sized medical domain datasets like BreakHis, further highlighting their robustness across different dataset sizes and domains.

The custom-made subset of ISIC 2019 dataset, with 9,133 samples, falls between ImageNet and BreakHis in terms of size, yet it is closely aligned with the target task of melanoma detection. Pre-training on ISIC 2019 led to strong results, particularly for ViT-B32, which achieved the highest accuracy of 90.95% and an AUC of 0.9548. The recall of 0.9169 for ViT-B32 suggests that pre-training on a domain-specific dataset greatly enhances the model's sensitivity to melanoma cases. While the dataset is significantly smaller than ImageNet, the domain alignment between ISIC 2019

and melanoma detection likely contributed to the model's ability to generalize well on the target task. The performance of hybrid models, particularly ViT-B16-VGG16, further underscores the value of specialized pre-training, as this model achieved an AUC of 0.9440 and precision of 0.9257, demonstrating strong performance across key metrics.

Overall, this analysis underscores the critical importance of dataset size and domain relevance in enhancing transfer learning for melanoma detection. ImageNet, as a large-scale and diverse dataset, provides models with the ability to capture complex and varied visual patterns, facilitating strong generalization across a wide range of tasks. This adaptability makes it an excellent choice for pre-training in transfer learning applications. In contrast, BreakHis, a dataset focused on histopathological images of breast cancer, is the smallest among the three datasets utilized across three distinct transfer learning scenarios in this study. As a result, it achieved the lowest performance across all scenarios. Moreover, while BreakHis is a medical dataset, its limited domain relevance to melanoma detection—due to substantial differences between histopathological and dermoscopic images—significantly diminishes its effectiveness. This lack of domain alignment, combined with its small size, further constrains its utility as a pre-training dataset for melanoma diagnosis. On the other hand, ISIC 2019, a domain-specific dataset focused on dermoscopic images of skin lesions, is closely aligned with the target task of melanoma detection. Its stronger domain relevance and larger sample size compared to BreakHis enable models to extract domain-specific features more effectively, resulting in superior performance for melanoma detection compared to BreakHis. Additionally, the results suggest that vision transformers demonstrate exceptional performance across all scenarios, underscoring their potential as a powerful tool for melanoma classification tasks.

Table 4.10 compares the findings of this study with previous research using the same dataset. A prior study, Gil et al. (2023), achieved an accuracy of 98.6% on the same private target dataset using a deep ensemble method comprising nine CNN models: GoogleNet, Inceptionv4, DenseNet201, ResNet50, InceptionResNetv2, NasNetlarge, EfficientNetb0, AlexNet, and ShuffleNet. Although our study's best accuracy of 97.97% is slightly lower than the 98.6% achieved by the ensemble approach in Gil et al. (2023), it is noteworthy for utilizing a single model rather than an ensemble of nine complex CNN models. This highlights the efficiency of our approach, which requires significantly fewer computational resources, making it highly practical for real-world applications.

Table 4.8: Summary of results from three transfer learning scenarios: 1) Models pre-trained on ImageNet, 2) Models pre-trained on BreakHis, and 3) Models pre-trained on ISIC 2019. All models were fully fine-tuned on the target melanoma dataset.

TL Approach	Model	Ave. Accuracy	AUC	Precision	Recall	F1-Score	Loss
ImageNet	ViT-B16	**97.97%**	**0.9925**	**0.9857**	**0.9772**	**0.9814**	**0.0976**
	ViT-B16-Inception-V3	96.74%	0.9911	0.9712	0.9698	0.9705	0.1193
	ViT-B16-VGG16	96.33%	0.9907	0.9519	0.9852	0.9683	0.1243
	ViT-B32	95.76%	0.9974	0.9575	0.9653	0.9614	0.1623
	VGG16	90.24%	0.9567	0.8993	0.9256	0.9123	0.2896
	Inception-V3	87.39%	0.9469	0.8689	0.9100	0.8890	0.2796
	MobileNet	85.49%	0.9492	0.9058	0.8212	0.8614	0.3132
BreakHis	ViT-B16	**86.59%**	**0.9336**	**0.8857**	**0.8735**	**0.8796**	**0.3959**
	ViT-B32	85.07%	0.9541	0.9145	0.8148	0.8618	0.5051
	Inception-V3	84.59%	0.9176	0.9104	0.7991	0.8511	0.5271
	ViT-B16-Inception-V3	84.58%	0.9028	0.9083	0.7974	0.8492	0.5138
	ViT-B16-VGG16	83.33%	0.9308	0.8662	0.8444	0.8552	0.4130
	MobileNet	80.18%	0.8661	0.8357	0.795	0.8148	0.4598
	VGG16	77.62%	0.8565	0.8188	0.7530	0.7845	0.4973
ISIC2019	ViT-B32	**90.95%**	**0.9548**	**0.9220**	**0.9169**	**0.9194**	**0.3835**
	ViT-B16-VGG16	87.83%	0.9440	0.9257	0.8510	0.8868	0.3686
	ViT-B16	87.81%	0.9345	0.9606	0.8140	0.8812	0.4782
	Inception-V3	86.99%	0.9456	0.9266	0.8282	0.8746	0.3392
	ViT-B16-Inception-V3	85.42%	0.9256	0.8761	0.8507	0.8632	0.4282
	VGG16	81.27%	0.9082	0.9141	0.7308	0.8122	0.4485
	MobileNet	73.58%	0.7983	0.8466	0.6366	0.7267	0.5860

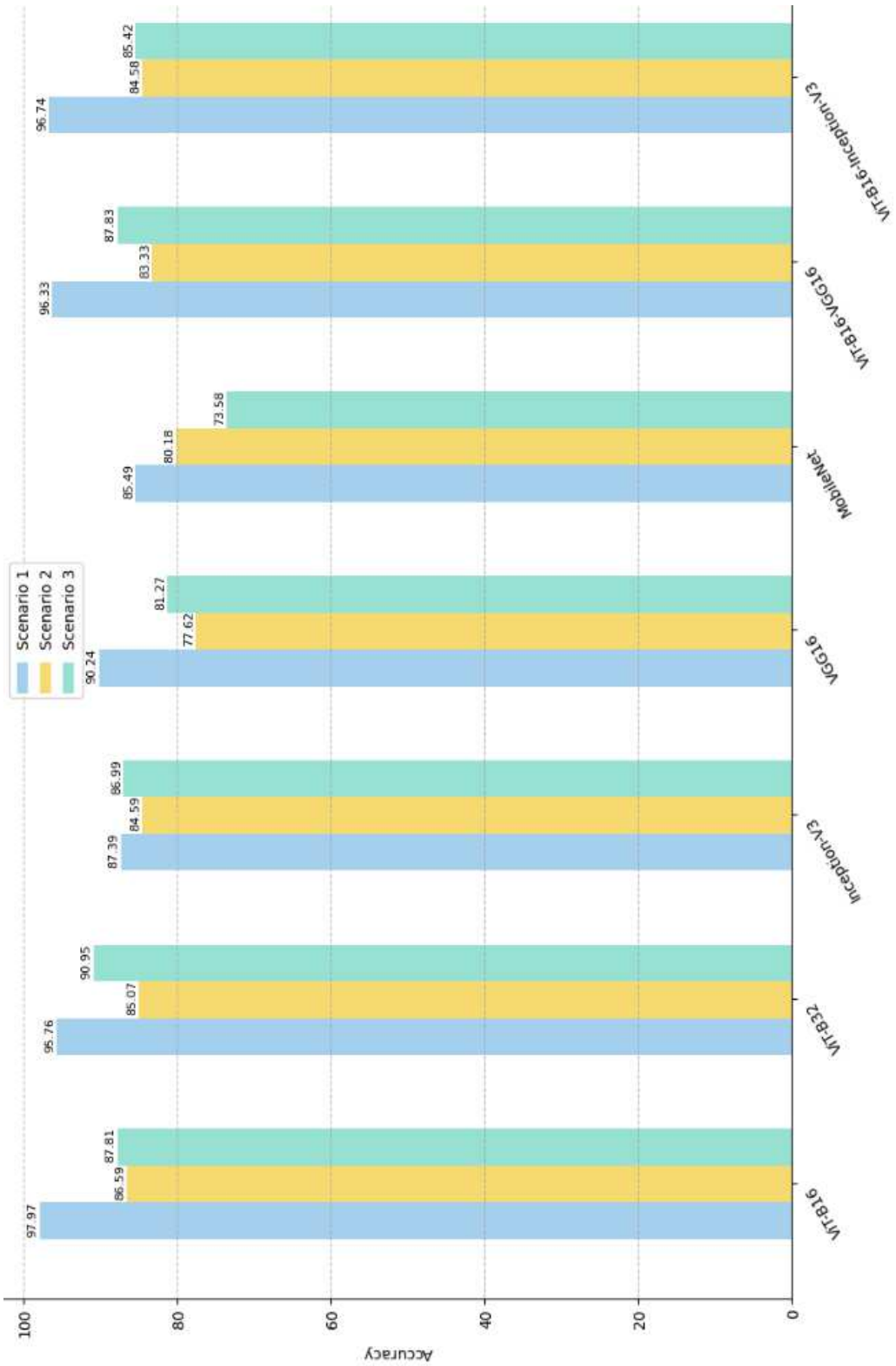

Figure 4.8: Comparison of models across scenarios.

Table 4.9: Detailed Overview of State-of-the-Art Studies.

Author	Method	Best Performed Model	Dataset	Classification	Best Results (ACC.)
Belattar et al. (2022)	Baseline CNN, InceptionV3,ResNet50, VGG16, Xception, MobileNetV2, DenseNet201	Baseline CNN	ISIC 2019	Binary (Melanoma/Nevus)	98.9%
Hosseinzadeh Kassani and Hosseinzadeh Kassani (2019)	AlexNet,VGGNet16, VGGNet19,ResNet50, Xception	ResNet50	ISIC 2018	Multi-class	92.08%
Faghihi et al. (2024)	VGG16, VGG19	VGG19	ISIC	Binary (Melanoma/Benign)	98.1%
Arshed et al. (2023)	ViT,ResNet50, ResNet101,ResNet152, ResNet50V2,VGG16, ResNet101V2,VGG19, ResNet152V2, DenseNet121, DenseNet169	ViT	HAM10000 & ISIC	Multi-class	92.14%
Kruk et al. (2015)	SVM and RF	SVM	Private Melanoma dataset	Binary (Melanoma/Benign)	93.8 %
Gil et al. (2023)	Shallow Ensemble (SVM, RF, Softmax) and deep ensemble including nine deep CNN models	Deep ensemble	Private Melanoma dataset	Binary (Melanoma/Benign)	98.6%
Proposed	ViT-B16,ViT-B32, Inception-V3,VGG16, MobileNet, and two hybrid models, ViT-B16-VGG16 and ViT-B16-Inception-V3	ViT-B16	Private Melanoma dataset	Binary (Melanoma/Benign)	97.97%

Table 4.10: Comparison of this study with previous research using the same dataset.

Author	Classification Method	Evaluation Method	Best Results (ACC.)
Kruk et al. (2015)	SVM and RF	K-fold CV	93.8 %
Gil et al. (2023)	Shallow Ensemble (SVM, RF, Softmax) and deep ensemble including nine deep CNN models	Mean of 10 experiments	98.6%
Proposed	ViT-B16, ViT-B32, Inception-V3, VGG16, MobileNet, and two hybrid models, ViT-B16-VGG16 and ViT-B16-Inception-V3	K-fold CV	97.97%

4.4 Summary

This chapter presents a comprehensive analysis of our research on melanoma detection using advanced deep learning models and transfer learning techniques. Our study explored three distinct transfer learning scenarios, each implemented across seven different deep learning architectures, to identify the most effective approach for accurate melanoma classification from dermoscopic images. We begin by introducing the performance metrics used to evaluate our models, including accuracy, precision, recall, and AUC. These metrics provide a detailed view of each model's capabilities in distinguishing between melanoma and non-melanoma cases. A unique aspect of our study is the attention maps visualization in Vision Transformer architectures, offering insights into the model's decision-making process and highlighting the areas of focus during image analysis. The results are presented for each of the three scenarios: models pre-trained on ImageNet and fine-tuned on our private melanoma dataset; models pre-trained on histopathological images (BreakHis) before fine-tuning; and models pre-trained on the ISIC 2019 dataset prior to fine-tuning. Each scenario's results are thoroughly analyzed, comparing the performance of different architectures and highlighting the best-performing models. We observe that the Vision Transformer models consistently demonstrate superior performance across scenarios, with the ViT-B16 model achieving particularly impressive

results. The chapter concludes with a discussion section, where we interpret our findings in the context of existing literature and explore the implications for future research in automated skin cancer detection. We also address the strengths and limitations of our approach, providing a balanced view of our contribution to the field.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

The early and precise detection of melanoma is crucial for improving patient survival rates. In this research, we explored three distinct transfer learning scenarios across seven different deep learning architectures to evaluate the effectiveness of various TL approaches for melanoma diagnosis using a private dataset of dermoscopic melanoma images. We applied meticulous pre-processing techniques, including normalization, resizing, and augmentation, to prepare the dataset. To ensure robustness, we employed k-fold cross-validation, which provided reliable measures of model performance. Our results reveal that the Vision Transformer model, ViT-B16, demonstrated exceptional performance in the first scenario when pre-trained on ImageNet, achieving an accuracy of 97.97% and an AUC of 0.9925. Another Vision Transformer variant, the ViT-B32 model, also showed strong results in the first scenario with an accuracy of 95.76%, outperforming traditional CNN models such as Inception-V3, VGG16, and MobileNet in both accuracy and AUC. These findings underscore the potential of Vision Transformers to enhance melanoma detection, surpassing traditional CNNs, which generally exhibited lower accuracy and higher loss rates. Furthermore, our findings highlight the critical importance of pre-training dataset selection in determining model effectiveness. Models pre-trained on the large-scale, diverse ImageNet dataset demonstrated strong generalization capabilities, with ViT-B16 achieving the highest accuracy. In contrast, models pre-trained on the domain-specific ISIC 2019 dataset excelled in recall and precision metrics compared

to those pre-trained on the BreakHis dataset. The ISIC 2019 dataset's closer alignment with the task of melanoma detection, combined with its larger sample size relative to BreakHis, enabled models to extract more relevant features. These results highlight the significance of dataset size and domain relevance in pre-training datasets, particularly for improving performance in specialized tasks such as melanoma classification.

5.2 Future Work

Future research should focus on optimizing ViT models by fine-tuning hyperparameters to improve precision and recall in melanoma detection. Although ViT models have already demonstrated outstanding performance, further investigation into hyperparameter adjustments could lead to even greater improvements in accuracy. Additionally, a detailed study of the impact of preprocessing techniques, such as image resizing, would offer valuable insights for enhancing model performance. Future studies could also explore using Generative Adversarial Networks (GANs) to create artificial melanoma images, addressing the challenge of limited annotated data and potentially enhancing model generalizability. Furthermore, incorporating ensemble learning techniques, such as majority voting, to aggregate predictions from multiple deep learning models could provide a more robust and accurate melanoma classification system. Incorporating these advanced models into clinical applications could revolutionize melanoma diagnostics, ultimately leading to better patient outcomes and care.

References

Abbasi, A. A., Hussain, L., Awan, I. A., Abbasi, I., Majid, A., Nadeem, M. S. A., & Chaudhary, Q.-A. (2020). Detecting prostate cancer using deep learning convolution neural network with transfer learning approach. *Cognitive Neurodynamics*, *14*, 523–533. doi: https://doi.org/10.1007/s11571-020-09587-5

Abbasi, N. R., Shaw, H. M., Rigel, D. S., Friedman, R. J., McCarthy, W. H., Osman, I., ... Polsky, D. (2004). Early diagnosis of cutaneous melanoma: revisiting the abcd criteria. *JAMA*, *292*(22), 2771–2776. doi: https://doi.org/10.1001/jama.292.22.2771

Ali, K., Shaikh, Z. A., Khan, A. A., & Laghari, A. A. (2022). Multiclass skin cancer classification using efficientnets – a first step towards preventing skin cancer. *Neuroscience Informatics*, *2*(4), 100034. doi: https://doi.org/10.1016/j.neuri.2021.100034

Aljabri, M., AlAmir, M., Al Ghamdi, M., Abdel-Mottaleb, M., & Collado-Mesa, F. (2022). Towards a better understanding of annotation tools for medical imaging: a survey. *Multimedia Tools and Applications*, *81*, 25877—25911. doi: https://doi.org/10.1007/s11042-022-12100-1

American Cancer Society. (2024). *Key statistics for melanoma skin cancer.* Retrieved from https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/key-statistics.html (Accessed: 2024-07-28)

Argenziano, G., & Soyer, H. P. (2012). Twenty years of dermoscopy. *Journal of the American Academy of Dermatology*, *67*(1), 125–126. doi: https://doi.org/10.1016/j.jaad.2011.06.048

Arshed, M. A., Mumtaz, S., Ibrahim, M., Ahmed, S., Tahir, M., & Shafi, M. (2023). Multi-class skin cancer classification using vision transformer networks and convolutional neural network-based pre-trained models. *Information*, *14*(7), 415. doi: https://doi.org/10.3390/

info14070415

Azad, R., Kazerouni, A., Heidari, M., Aghdam, E. K., Molaei, A., Jia, Y., ... Merhof, D. (2024). Advances in medical image analysis with vision transformers: A comprehensive review. *Medical Image Analysis*, *91*, 103000. doi: https://doi.org/10.1016/j.media.2023.103000

Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Classification of breast cancer based on histology images using convolutional neural networks. *IEEE Access*, *6*, 24680–24693. doi: https://doi.org/10.1109/ACCESS.2018.2831280

Barros, W., Morais, D., Fernandes Lopes, F., Torquato, M., De Melo Barbosa, R., & Fernandes, M. (2020). Proposal of the cad system for melanoma detection using reconfigurable computing. *Sensors*, *20*, 3168. doi: https://doi.org/10.3390/s20113168

Belattar, K., Adjadj, M., Bakir, M., & Ait Mehdi, M. (2022). A comparative study of cnn architectures for melanoma skin cancer classification. In *ICT Innovations* (pp. 74–89).

Brinker, T. J., Hekler, A., Utikal, J. S., Grabe, N., Schadendorf, D., Klode, J., ... Von Kalle, C. (2018). Skin cancer classification using convolutional neural networks: systematic review. *Journal of medical Internet research*, *20*(10), 365–372. doi: https://doi.org/10.2196/11936

Chen, J. Y., Fernandez, K., Fadadu, R. P., Reddy, R., Kim, M.-O., Tan, J., & Wei, M. L. (2024). Skin cancer diagnosis by lesion, physician, and examination type: A systematic review and meta-analysis. *JAMA Dermatology*. doi: https://doi.org/10.1001/jamadermatol.2024.4382

Cirrincione, G., Cannata, S., Cicceri, G., Prinzi, F., Currieri, T., Lovino, M., ... Vitabile, S. (2023). Transformer-based approach to melanoma detection. *Sensors*, *23*(12), 9335–9351. doi: https://doi.org/10.3390/s23125677

Codella, N. C. F., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S. W., Gutman, D. A., ... Halpern, A. (2017). Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). *arXiv preprint arXiv:1902.03368*. doi: https://doi.org/10.48550/arXiv.1902.03368

Deininger, L., Stimpel, B., Yuce, A., Abbasi-Sureshjani, S., Schönenberger, S., Ocampo, P., ... Gaire, F. (2022). A comparative study between vision transformers and cnns in digital pathology. *arXiv preprint arXiv:2206.00389*. doi: https://doi.org/10.48550/arXiv.2206.00389

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Li, F.-F. (2009). Imagenet: a large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). doi: https://doi.org/10.1109/CVPR.2009.5206848

Dennis, L. K., Vanbeek, M. J., Beane Freeman, L. E., Smith, B. J., Dawson, D. V., & Coughlin, J. A. (2008). Sunburns and risk of cutaneous melanoma: does age matter? a comprehensive meta-analysis. *Annals of Epidemiology*, *18*(8), 614–627. doi: https://doi.org/10.1016/j.annepidem.2008.03.006

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... others (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. doi: https://doi.org/10.48550/arXiv.2010.11929

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. doi: https://doi.org/10.1038/nature21056

Faghihi, A., Fathollahi, M., & Rajabi, R. (2024). Diagnosis of skin cancer using vgg16 and vgg19 based transfer learning models. *arXiv preprint arXiv:2404.01160*. doi: https://doi.org/10.48550/arXiv.2404.01160

Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bray, F., ... Bray, F. (2021). Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA Dermatology*, *157*(5), 549–555. doi: https://doi.org/10.1001/jamadermatol.2020.5391

Fidler, I. J. (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature reviews cancer*, *3*(6), 453–458. doi: https://doi.org/10.1038/nrc1098

Flosdorf, C., Engelker, J., Keller, I., & Mohr, N. (2024). Skin cancer detection utilizing deep learning: Classification of skin lesion images using a vision transformer. *arXiv preprint arXiv:2407.18554*. doi: https://doi.org/10.48550/arXiv.2407.18554

Gandini, S., Sera, F., Cattaruzza, M. S., Pasquini, P., Abeni, D., Boyle, P., & Melchi, C. F. (2005). Meta-analysis of risk factors for cutaneous melanoma: I. common and atypical naevi. *European Journal of Cancer*, *41*(1), 28–44. doi: https://doi.org/10.1016/j.ejca.2004.10.015

Gandini, S., Sera, F., Cattaruzza, M. S., Pasquini, P., Picconi, O., Boyle, P., & Melchi, C. F. (2011). Meta-analysis of risk factors for cutaneous melanoma: Ii. sun exposure. *European Journal of*

Cancer, *47*(5), 720–731. doi: https://doi.org/10.1016/j.ejca.2010.11.008

Geller, A. C., Swetter, S. M., Brooks, K., Demierre, M.-F., & Yaroch, A. L. (2007). Screening, early detection, and trends for melanoma: current status (2000-2006) and future directions. *Journal of the American Academy of Dermatology*, *57*(4), 555–572. doi: https://doi.org/10.1016/j.jaad.2007.06.032

Gil, F., Osowski, S., Swiderski, B., & Słowińska, M. (2023). Ensemble of classifiers based on deep learning for medical image recognition. *Metrology and Measurement Systems*, *30*(1), 139–156. doi: https://doi.org/10.24425/mms.2023.144400

Gordon, R. (2013). Skin cancer: An overview of epidemiology and risk factors. *Seminars in Oncology Nursing*, *29*(3), 160-169. doi: https://doi.org/10.1016/j.soncn.2013.06.002

Gouda, W., Sama, N. U., Al-Waakid, G., Humayun, M., & Jhanjhi, N. Z. (2022). Detection of skin cancer based on skin lesion images using deep learning. *Healthcare*, *10*(7). doi: https://doi.org/10.3390/healthcare10071183

Haghshenas, F., Krzyżak, A., & Osowski, S. (2024). Comparative study of deep learning models in melanoma detection. In *Artificial Neural Networks in Pattern Recognition (ANNPR)* (pp. 121–131). Springer, Cham. doi: https://doi.org/10.1007/978-3-031-71602-7_11

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646-674. doi: https://doi.org/10.1016/j.cell.2011.02.013

Hekler, A., Utikal, J. S., Enk, A. H., Solass, W., Schmitt, M., Klode, J., ... Brinker, T. J. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer*, *118*, 91-96. doi: https://doi.org/10.1016/j.ejca.2019.06.012

Hernandez, C., Combalia, M., Podlipnik, S., Codella, N., Rotemberg, V., Halpern, A., ... Malvehy, J. (2024). Bcn20000: Dermoscopic lesions in the wild. *Scientific Data*, *11*(1), 641. doi: https://doi.org/10.1038/s41597-024-03387-w

Hosny, K. M., Kassem, M. A., & Foaud, M. M. (2018). Skin cancer classification using deep learning and transfer learning. In *9th Cairo International Biomedical Engineering Conference (CIBEC)* (pp. 90–93). doi: https://doi.org/10.1109/CIBEC.2018.8641762

Hosseinzadeh Kassani, S., & Hosseinzadeh Kassani, P. (2019). A comparative study of deep

learning architectures on melanoma detection. *Tissue and Cell*, *58*, 76–83. doi: https://doi.org/10.1016/j.tice.2019.04.009

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. doi: https://doi.org/10.48550/arXiv.1704.04861

International Agency for Research on Cancer. (2022). *Globocan 2022: Estimated cancer incidence, mortality and prevalence worldwide in 2022 - melanoma of skin fact sheet.* Retrieved from https://gco.iarc.who.int/media/globocan/factsheets/cancers/16-melanoma-of-skin-fact-sheet.pdf (Accessed: 2024-08-04)

Jafari, M. H., Samavi, S., Karimi, N., Soroushmehr, S. M. R., Ward, K., & Najarian, K. (2016). Automatic detection of melanoma using broad extraction of features from digital images. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (p. 1357-1360). doi: https://doi.org/10.1109/EMBC.2016.7590959

Kassem, M. A., Hosny, K. M., Damaševičius, R., & Eltoukhy, M. M. (2021). Machine learning and deep learning methods for skin lesion classification and diagnosis: A systematic review. *Diagnostics*, *11*(8). doi: https://doi.org/10.3390/diagnostics11081390

Kim, S., Gaibor, E., & Haehn, D. (2024). Web-based melanoma detection. *arXiv preprint arXiv:2403.14898*. doi: https://doi.org/10.48550/arXiv.2403.14898

Kruk, M., Świderski, B., Osowski, S., Kurek, J., Słowińska, M., & Walecka, I. (2015). Melanoma recognition using extended set of descriptors and classifiers. *EURASIP Journal on Image and Video Processing*, *2015*(43), 1–10. doi: https://doi.org/10.1186/s13640-015-0099-9

Marzuka, A. G., & Book, S. E. (2015). Basal cell carcinoma: pathogenesis, epidemiology, clinical features, diagnosis, histopathology, and management. *The Yale Journal of Biology and Medicine*, *88*(2), 167—179.

Mayo Clinic. (2024a). *Basal cell carcinoma - symptoms and causes.* Retrieved from https://www.mayoclinic.org/diseases-conditions/basal-cell-carcinoma/symptoms-causes/syc-20354187 (Accessed: 2024-07-28)

Mayo Clinic. (2024b). *Melanoma - symptoms and causes.* Retrieved from https://www.mayoclinic.org/diseases-conditions/melanoma/

symptoms-causes/syc-20374884 (Accessed: 2024-07-28)

Mayo Clinic. (2024c). *Squamous cell carcinoma - symptoms and causes.* Retrieved from https://www.mayoclinic.org/diseases-conditions/squamous-cell-carcinoma/symptoms-causes/syc-20352480 (Accessed: 2024-07-28)

Menegola, A., Fornaciali, M., Pires, R., Bittencourt, F. V., Avila, S., & Valle, E. (2017). Knowledge transfer for melanoma screening with deep learning. In *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (p. 297-300). doi: https://doi.org/10.1109/ISBI.2017.7950523

Mohr, P., Eggermont, A. M., Hauschild, A., & Buzaid, A. (2009). Staging of cutaneous melanoma. *Annals of Oncology*, *20*(suppl_6), vi14–vi21. doi: https://doi.org/10.1093/annonc/mdp252

Monika, M. K., Arun Vignesh, N., Usha Kumari, C., Kumar, M., & Lydia, E. L. (2020). Skin cancer detection and classification using machine learning. *Materials Today: Proceedings*, *33*, 4266-4270. doi: https://doi.org/10.1016/j.matpr.2020.07.366

Mousa, Y., Taha, R., Kaur, R., & Afifi, S. (2024). Melanoma classification using deep learning. In *Image and Video Technology* (p. 259–272). doi: https://doi.org/10.1007/978-981-97-0376-0_20

Mutasa, S., Sun, S., & Ha, R. (2020). Understanding artificial intelligence based radiology studies: What is overfitting? *Clinical Imaging*, *65*, 96–99. doi: https://doi.org/10.1016/j.clinimag.2020.04.025

Olsen, C. M., Carroll, H. J., & Whiteman, D. C. (2010a). Estimating the attributable fraction for melanoma: a meta-analysis of pigmentary characteristics and freckling. *International Journal of Cancer*, *127*(10), 2430–2445. doi: https://doi.org/10.1002/ijc.25548

Olsen, C. M., Carroll, H. J., & Whiteman, D. C. (2010b). Familial melanoma: a meta-analysis and estimates of attributable fraction. *Cancer Epidemiology and Prevention Biomarkers*, *19*(1), 65–73. doi: https://doi.org/10.1158/1055-9965.EPI-09-0929

Pu, Q., Xi, Z., Yin, S., Zhao, Z., & Zhao, L. (2024). Advantages of transformer and its application for medical image segmentation: a survey. *BioMedical Engineering OnLine*, *23*, 14. doi: https://doi.org/10.1186/s12938-024-01212-4

Que, S. K. T., Zwald, F. O., & Schmults, C. D. (2018). Cutaneous squamous cell carcinoma: Incidence, risk factors, diagnosis, and staging. *Journal of the American Academy of Dermatology*, *78*(2), 237-247. doi: https://doi.org/10.1016/j.jaad.2017.08.059

Rastrelli, M., Tropea, S., Rossi, C. R., & Alaibac, M. (2014). Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo*, *28*(6), 1005–1011.

Read, J., Wadt, K. A., & Hayward, N. K. (2016). Melanoma genetics. *Journal of Medical Genetics*, *53*(1), 1–14. doi: https://doi.org/10.1136/jmedgenet-2015-103150

Schadendorf, D., van Akkooi, A. C., Berking, C., Griewank, K. G., Gutzmer, R., Hauschild, A., ... Ugurel, S. (2018). Melanoma. *The Lancet*, *392*(10151), 971–984. doi: https://doi.org/10.1016/S0140-6736(18)31559-9

Shamshiri, M. A., Krzyzak, A., Kowal, M., & Korbicz, J. (2023). Compatible-domain transfer learning for breast cancer classification with limited annotated data. *Computers in Biology and Medicine*, *154*, 106575. doi: https://doi.org/10.1016/j.compbiomed.2023.106575

Shimizu, K., Iyatomi, H., Celebi, M. E., Norton, K.-A., & Tanaka, M. (2015). Four-class classification of skin lesions with task decomposition strategy. *IEEE Transactions on Biomedical Engineering*, *62*(1), 274-283. doi: https://doi.org/10.1109/TBME.2014.2348323

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556v6*. doi: https://doi.org/10.48550/arXiv.1409.1556

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, *63*(7), 1455-1462. doi: https://doi.org/10.1109/TBME.2015.2496264

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. doi: https://doi.org/10.3322/caac.21660

Swetter, S. M., Tsao, H., Bichakjian, C. K., Curiel-Lewandrowski, C., Elder, D. E., Gershenwald, J. E., ... others (2019). Guidelines of care for the management of primary cutaneous melanoma. *Journal of the American Academy of Dermatology*, *80*(1), 208–250. doi:

https://doi.org/10.1016/j.jaad.2018.08.055

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2818–2826). doi: https://doi.org/10.1109/CVPR.2016.308

The Skin Cancer Foundation. (2024). *Melanoma warning signs and images.* Retrieved from https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-warning-signs-and-images/ (Accessed: 2024-07-28)

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

Wang, W., Li, Y., Zou, T., Wang, X., You, J., & Luo, Y. (2020). A novel image classification approach via dense-mobilenet models. *Mobile Information Systems*, *2020*(1), 1–8. doi: https://doi.org/10.1155/2020/7602384

Watts, C. G., Dieng, M., Morton, R. L., Mann, G. J., Menzies, S. W., & Cust, A. E. (2015). Clinical practice guidelines for identification, screening and follow-up of individuals at high risk of primary cutaneous melanoma: a systematic review. *British Journal of Dermatology*, *172*(1), 33–47. doi: https://doi.org/10.1111/bjd.13403

Wernli, K. J., Henrikson, N. B., Morrison, C. C., Nguyen, M., Pocobelli, G., & Blasi, P. R. (2016). Screening for Skin Cancer in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*, *316*(4), 436-447. doi: https://doi.org/10.1001/jama.2016.5415

Xie, J., Wu, Z., Zhu, R., & Zhu, H. (2021). Melanoma detection based on swin transformer and simam. In *IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (Vol. 5, p. 1517-1521). doi: https://doi.org/10.1109/ITNEC52019.2021.9587071

Xin, C., Liu, Z., Zhao, K., Miao, L., Ma, Y., Zhu, X., ... Chen, H. (2022). An improved transformer network for skin cancer classification. *Computers in Biology and Medicine*, *149*, 105939. doi: https://doi.org/10.1016/j.compbiomed.2022.105939

Yang, G., Luo, S., & Greer, P. (2023). A novel vision transformer model for skin cancer

classification. *Neural Processing Letters*, *55*, 9335–9351. doi: https://doi.org/10.1007/s11063-023-11204-5

Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P.-A. (2017). Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging*, *36*(4), 994-1004. doi: https://doi.org/10.1109/TMI.2016.2642839