Deep Convolutional Neural Network for Precise Skin Cancer Segmentation

Table of Contents

List of Figures

List of Tables

Abstract

Among the most common types of cancer is skin cancer, for which early detection is critical for successful treatment. Unfortunately, current diagnostic methods are sometimes insufficient, especially when it comes to identifying melanoma from other skin lesions. To improve Convolutional Neural Network weaknesses of CNN in automated skin cancer detection.

In this paper, a deep learning system to classify and segment skin cancer lesions from dermoscopic images using CNN structures and Transfer Learning such as InceptionV3, EfficientNet, and VGG19. Data preprocessing methods including image scaling and rotation were used to increase the robustness of the model. Transfer learning was also used to optimise the performance of pre-trained models by fine-tuning them. Some important criteria, such as accuracy, sensitivity, specificity, and F1-score, were used to evaluate the models.

EfficientNetB0 (Freeze Layer) reached 78% accuracy but was at risk of overfitting, while InceptionV3 addressed class imbalances effectively with an accuracy of 75%. The study highlights how crucial it is to improve class balancing and data preprocessing methods to improve model performance.

The study improves the building of skin cancer detection models with more accuracy and indicates the importance of further research including complex methods for data preprocessing such as histogram equalisation and DullRazor algorithms, as well as class balancing techniques that improve model generalisation. Clinical settings will find these models more useful with improved preprocessing and augmentation methods.

Chapter 1

Introduction

1.1 Overview

Since the skin is the largest organ in the human body, visual examination is the primary method used to diagnose skin conditions. The number of cases of skin cancer increases each year, although being comparatively rare in contrast with other forms of cancer. Early skin cancer identification is critical to both treatment outcome and survival. Skin cancer is suitable for deep learning object recognition tasks because it can be identified by signs such as abnormal tissue growth, red or pearly colouration, and dark spots or marks. However, the identification of skin cancer which goes beyond images requires additional support from medical history. Convolution neural networks demonstrate significant analytical in processing abnormal tissue in skin images and general image analysis with the advent of deep learning algorithms. This has aided in the creation of computer-aided diagnosis systems for the analysis of skin cancer.(Dildar et al. 2021)

A 2017 study found that skin cancer accounts for 1.79 per cent of the global disease burden, measured in disability-adjusted life years. Skin cancer causes about 7 per cent of all new cancer cases globally, costing the US Medicare program more than 8 billion dollars in 2011. Clinical data supports the belief that radical disparities exist in the outcomes of skin cancer cases. Patients with darker skin tones have a 20-30 per cent lower risk of developing melanoma than those with lighter skin tones, but they also have a higher or lower mortality risk for specific melanoma types (Rezaoana et al. 2020). In recent years, deep learning has been utilised extensively for supervised and unsupervised learning problems. Among these models, Convolution Neural Networks (CNNs) outperformed all others in tasks involving object detec-

tion and classification. CNNs minimise the need for manual feature construction by developing highly discriminative features through controlled end-to-end practice. Convolution Neural Networks (CNNs) have been applied recently to the classification of skin cancer lesions certain CNNs models have outperformed trained human specialists in the classification of skin cancers. There are other methods, such as transfer learning. The utilisation of large datasets has considerably improved the performance of these simulations. (Shete et al. 2021)

The study topic is briefly summarised, with an emphasis on the growing significance of early detection and correct diagnosis of skin cancer. The part also clarifies the aims and constraints of the research by outlining the precise objectives and scope of the investigation. Through an exploration of the obstacles and possibilities related to the identification of skin cancer, the introduction provides an in-depth overview of the background against which the study is conducted. It also serves as an outline for the rest of the work, identifying the important concepts and themes that will be covered in the sections that follow. In general, the introduction establishes the background and aims that will direct the course of the study and provide a basic framework for the analytical investigation of skin cancer detection. (Gezimati & Singh 2023, Wen et al. 2022)

1.2 Problem Background

The occurrence of melanoma is directly related to sun exposure, and because of their visual similarities, it can be difficult to differentiate between melanoma and nonmelanoma lesions even with advancements in diagnostic techniques. Less than desirable accuracy rates, with visual examination accuracy falling below 80%, are caused by the shortcomings of standard diagnostic techniques, such as dermoscopic image analysis, which can be affected by image noise and the overlapping features of various skin lesions. Growing interest has been shown in the use of automated systems, especially those based on deep learning (DL) and machine learning (ML) algorithms, for the diagnosis and classification of skin cancer as a solution to these issues. The shortcomings of current machine learning techniques, however, are their shallow training performance and requirement for intensive image processing. On the other hand, deep learning techniques, more especially Convolutional Neural Networks (CNNs), have demonstrated potential in many cancer identification tasks because of their capacity to automatically extract characteristics from unprocessed data, which makes them appropriate for improving skin cancer diagnosis accuracy.

(Saleh et al. 2024)

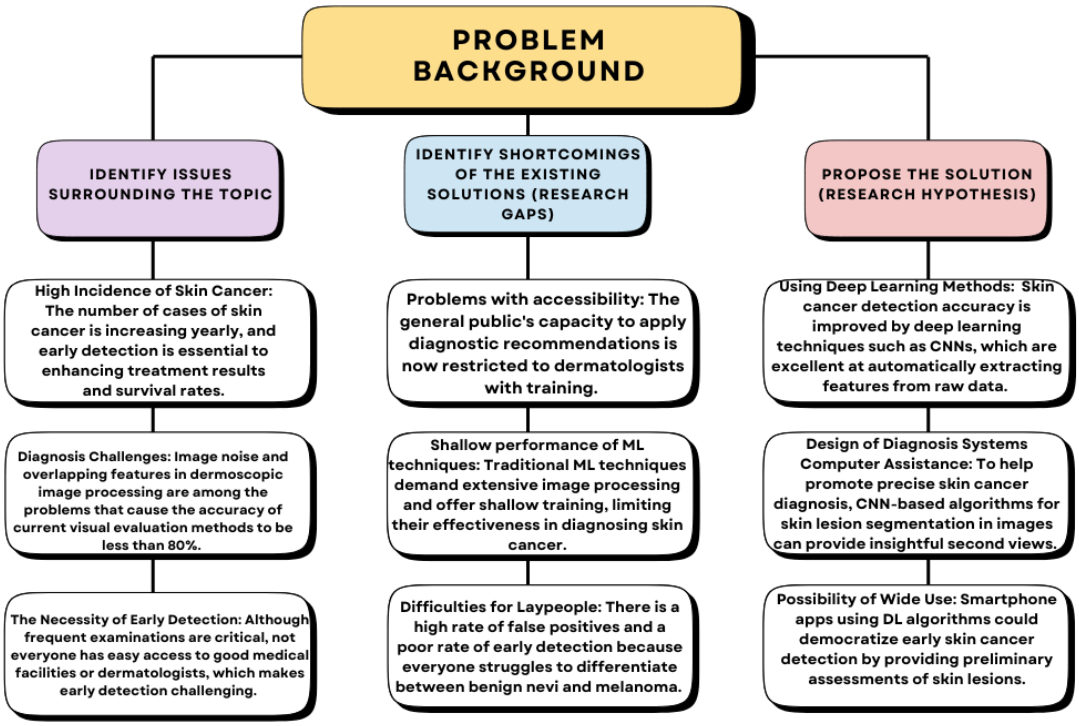

Figure 1.1: Breakdown and Mapping problem and solutions.

Currently, only dermatologists or those with substantial knowledge of skin cancer diagnosis can apply these guidelines. Hardly everyone knows or has access to this kind of knowledge. Physical examinations are the most effective method to detect cancer early on and should be done regularly. Therefore, it is challenging to achieve effective early identification for those who do not have easy accessibility to a hospital. The public can now easily snap images of lesions on their bodies and, if necessary, consult a dermatologist for skin illnesses according to the growing use of smartphones with cameras. Unfortunately, the distinction between benign nevi and melanoma is extremely difficult for laypeople to make, which results in a low rate of early diagnosis and a high false-positive rate. (Janda et al. 2020, MacLellan et al. 2021)

These clinical treatments are based on human visual assessments of dermoscopy images, which are colourful photos of skin lesions from dermatoscopes. For this reason, computer-aided diagnosis of this type of skin cancer becomes essential to give professionals a better second opinion. Because skin lesions frequently have low contrast, uneven lighting, hairy or non-hairy skin, and other abnormalities, analysing skin lesion photos can be a very difficult task. A skilled dermatologist

would find it extremely challenging to distinguish the lesion area from the backdrop in these dermoscopy pictures because of these abnormalities. Therefore, developing a computer-aided system that can autonomously divide the lesion area and lessen the burden on physicians is essential. Suggested to use of a deep Convolutional Neural Networks (CNNs) based system for skin lesion segmentation in dermoscopy pictures rather than creating a method for skin segmentation from scratch. (Baig et al. 2020, Zafar et al. 2020, Mahbod et al. 2020)

1.3 Research Aim

This study aims to create a highly accurate deep learning model, such as a Convolutional Neural Network (CNN), designed for accurately detecting, categorising, and segmenting multiple types of skin cancer from dermatological pictures. The purpose of the concept is to make early diagnosis and treatment easier, improving patient outcomes and decreasing the burden on doctors and nurses. As part of this research, modern deep learning methods will be investigated, CNN structures will be designed and optimised, and comprehensive validation will be conducted to validate the model's clinical accuracy and applicability to a variety of healthcare situations.

1.4 Research Objectives

- To investigate the existing literature, techniques, and approaches in automated skin cancer detection.

- To design and implement into practice a convolutional neural network (CNN) specifically aimed at identifying skin cancer, experimenting with several structures to optimise performance.

- To provide a thorough analysis of current methods for detecting skin cancer, with a focus on deep learning techniques.

- To evaluate the model's overall performance in identifying multiple types of skin cancer along with its accuracy, sensitivity, and specificity.

1.5 Research Scope

The research scope of this study is to create and test Convolutional Neural Networks (CNNs) for the accurate segmentation, classification, and diagnosis of skin cancer using dermoscopic images. The study will focus on examining multiple CNN structures, such as VGG19, Inception, and EfficientNet, to determine the most successful models for recognizing different types of skin lesions. The study will focus on the analysis of publicly available skin cancer image databases, with the addition of patient medical histories where available, to improve diagnostic accuracy. The scope includes designing, implementing, and testing CNN models, as well as using data augmentation techniques such as image flipping, rotation, and scaling to improve model robustness. The performance will be analysed using important metrics such as accuracy, sensitivity, specificity, and F1-score, ensuring a comprehensive evaluation against industry standards.

The study will not, however, include constructing entire new CNN design structures from the ground up or putting these models into actual use in clinical settings. Furthermore, as the focus of the study is on image-based diagnostics, non-dermoscopic techniques for detecting skin cancer, such as biopsies or genetic testing, will not be covered. The study sample is limited to those included in the publicly available datasets, which may mainly include individuals with light skin tones. This may reduce the generality of the results to a wider range of groups. Long-term clinical trials and post-implementation studies are not included in the research timeline; it is restricted to the time frame needed for model development, training, and validation. Deep learning, dermatology, and medical image analysis are the main ideas and theories covered in this study, with a focus on using CNNs to increase diagnostic precision and improve skin cancer early detection.

1.6 Abridged Methodology

A systematic procedure was followed to conduct the analysis and create the model for the diagnosis of skin cancer. First, to ensure reliability and compatibility, we verified the versions of the Python and packages. Collected image datasets for skin cancer from openly accessible resources. To make images suitable for input into the CNN model, preprocessing techniques included, loading data, scaling, and standardising pixel values. Implemented data augmentation methods to improve the diversity and robustness of the training set, including image flipping, rotation, and zoom-in.

We applied Convolutional Neural Networks (CNNs) for model creation because of their performance in image categorisation algorithms. To maximise performance, we experimented with several designs and hyperparameters. In addition, pre-trained models were used in transfer learning to make use of pre-existing information and increase accuracy despite the small dataset. Performance metrics including accuracy, precision, recall, and F1-score were computed to assess the models' effectiveness after they were trained and validated on the processed dataset. This complete strategy ensures a robust framework for image-based skin cancer screening.

1.7 Contribution

This research significantly advances the field of medical diagnostics, especially in the early identification and classification of skin cancer, by developing and evaluating deep-learning models. By using Convolutional Neural Networks (CNNs) on publicly available skin cancer image datasets, the research attempts to improve the efficiency and precision of automated skin cancer identification. It thoroughly examination of current practices, points out areas in need of development and suggests using several CNN structures, such as Vgg19, Inception, and EfficientNet, to maximise performance. The study also emphasises the important data augmentation methods to support model adaptability, such as image flipping, rotation, and scaling.

Furthermore, the study thoroughly evaluates model performance using metrics, such as accuracy, sensitivity, specificity, and F1-score, to ensure accuracy in clinical situations. The main objective is to understand the gap between experimental models and real-world implementation, which will enhance patient outcomes, reduce the workload for dermatologists, and encourage better integration of deep learning in the field. This research contributes significant new insights into how deep learning might improve the precision and efficiency of skin cancer diagnosis.

Chapter 2

Literature Review

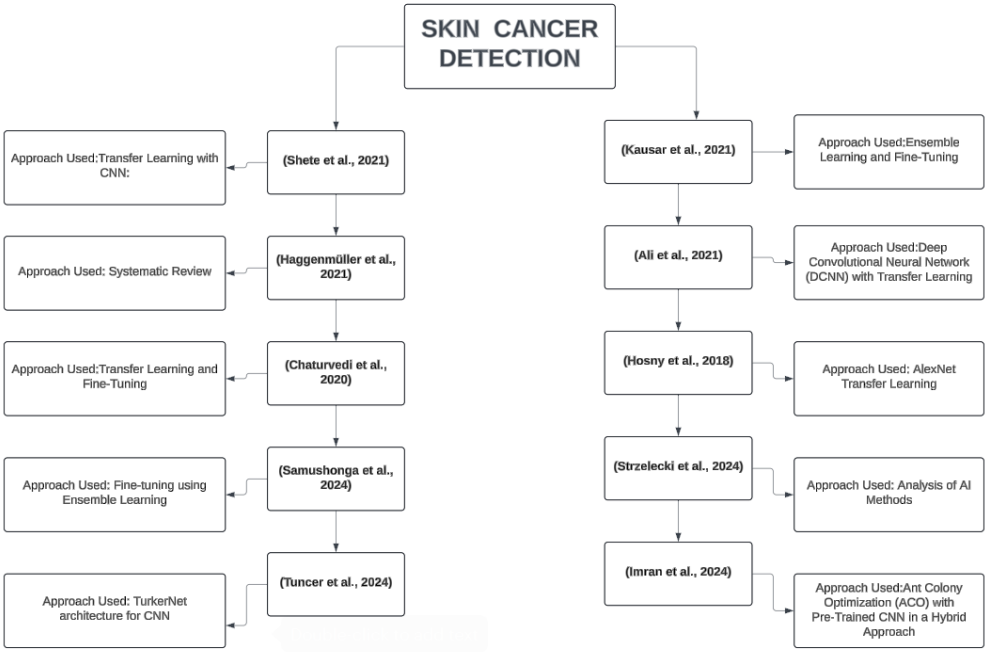

Figure 2.1: Overview of Literature Review

2.1 Transfer Learning with CNN

The research applied transfer learning through the improvement of pre-trained CNN models, specifically VGG16 and ResNet. The HAM10000 dataset was used to retrain these models, which improved the classification accuracy for skin cancer lesions by including new layers and adjusting parameters. Through transfer learning, the models were able to effectively transfer their existing expertise from a huge dataset (ImageNet) to the focused task of skin cancer detection.

2.1.1 Transfer Learning With CNN

Shete et al. (2021) The study proposes a technique for identifying and categorising skin cancer that uses Convolutional Neural Networks (CNNs) combined with Transfer Learning. The main aim is to improve the accuracy of skin cancer diagnostics by achieving high precision and recall in classifying skin lesions into different categories. By demonstrating the effectiveness of CNNs in particular, the VGG16 model when combined with Transfer Learning for the diagnosis of skin cancer, this work contributes to the research of literature. Additionally, it highlights the shortcomings of traditional machine learning algorithms in the classification of skin lesions, underscoring the need for deep learning methodologies. Moreover, it offers empirical support for the idea that Transfer Learning is an essential approach for enhancing model accuracy, particularly in situations with limited data, which makes it particularly relevant to datasets concerning skin cancer.

The research concluded that the CNN model, improved by Transfer Learning using the VGG16 architecture, classified skin cancer lesions with a weighted average precision, recall, and F1-score of 0.88, 0.74, and 0.77, respectively, achieving a 90.51% accuracy rate. Especially on the HAM10000 dataset, it outperformed standard machine learning techniques like Random Forest, XGBoost, and SVM. Transfer Learning's usefulness is shown by the significant performance gain it provides, particularly in situations where there is a lack of data. The scientists do, however, agree that there is still room for development, especially when it comes to handling data imbalance. They propose that future studies concentrate on improved data balancing methods in an effort to increase model accuracy.

2.2 Systematic Review

The research implemented the systematic review approach, carrying out comprehensive research and examination of existing research to determine CNNs' performance in skin cancer classification compared to human professionals.

2.2.1 Convolutional Neural Network

Haggenmüller et al. (2021) Although taking up only 4% of skin cancers, malignant melanoma is responsible for over 75% of associated deaths. Even with advances like dermoscopy, diagnosis accuracy is still around 80%, despite the need for early

detection. In order to evaluate Convolutional Neural Networks (CNNs) clinical significance and potential for everyday practice, a systematic evaluation compares their performance in melanoma categorisation to that of human experts. The review notes that while CNNs show potential, their current application in clinical practice is hampered by the artificial nature of most available studies. It emphasises the need for additional clinically relevant evaluations in various, realistic scenarios.

For peer-reviewed articles from 2017 to 2021, the authors did a comprehensive search across PubMed, Medline, and ScienceDirect. Their focus was on the use of CNNs for skin cancer classification and their comparison with human specialists, particularly in the diagnosis of melanoma. 19 research were found to meet the inclusion criteria; these studies demonstrated that CNNs performed as well as or better than human specialists in the categorisation of skin cancer, mainly based on dermoscopic images. However, the majority of research lacked clinical diversity because it was carried out in experimental settings using holdout test sets. A single study that evaluated CNNs in an actual clinical scenario found that dermatologists performed better than AI, underscoring CNNs' shortcomings in the absence of complete patient data. In order to more accurately evaluate CNNs' clinical significance, the review highlights its importance for future research to make use of different datasets and more realistic test situations.

2.3 Transfer Learning and Fine-Tuning

Transfer learning was used in the study to specifically fine-tune pre-trained CNN models for multi-class skin cancer categorization. Using minimum preprocessing, this method utilises the knowledge from huge image datasets and applies it to the particular task at issue.

2.3.1 Transfer Learning with CNN

Chaturvedi et al. (2020) Given the clear similarities between different skin lesions, diagnosing skin cancer is still a considerable issue. The goal of this research is to accurately differentiate between various forms of skin cancer by creating an automated computer-aided diagnostic (CAD) system that makes use of deep convolutional neural networks (CNNs). To determine the best method for multi-class skin cancer classification, the study evaluates the performance of five pre-trained CNN models and four ensemble models. The paper highlights the benefits of transfer

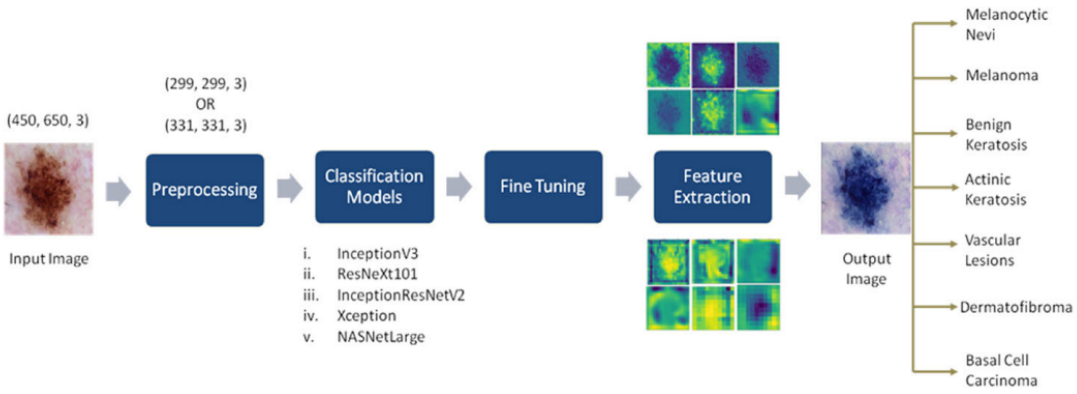

Figure 2.2: CNN Architecture

learning with fine-tuning in obtaining high accuracy using a comprehensive review of several pre-trained CNN models. It also shows that performance may be improved by ensemble approaches even with higher computing needs. The work addresses the problem of multi-class classification, which is frequently disregarded in support of binary classification in previous research. Figure 2.2 shows the CNN architecture used in the current study.

To maintain model generality, the study used the HAM10000 dataset, which included 10,015 dermoscopic images from seven different skin cancer classes. No significant data augmentation or preprocessing was done. Transfer learning was used to improve five pre-trained CNN models: Xception, ResNeXt101, InceptionV3, InceptionResNetV2, and NASNetLarge. Four ensemble models were evaluated in addition to the models being optimised using SGD and Adam optimizers. The highest individual model accuracy was 93.20% for ResNeXt101 and InceptionResNetV2, while the ensemble of InceptionResNetV2 + ResNeXt101 gained 92.83%. These models performed better than both the accuracy of regular dermatologists (62% to 80%) and current deep learning techniques. The study made clear that whereas ensemble models can increase accuracy, they also increase training time and complexity. Clinical visuals may be used in future studies to enhance the diagnostic accuracy even more.

2.4 Ensemble Learning and Fine-Tuning

By optimising five pre-trained CNN models and combining their outputs using majority voting and weighted majority voting approaches, the study performed ensemble learning. Through integrating the advantages of multiple models, classification

performance is improved.

2.4.1 Fine-Tuning

Kausar et al. (2021) The objective of this work is to improve multiclass skin cancer classification through the development of ensemble models built from optimised deep learning architectures. The study uses ensemble approaches to improve classification performance, addressing the issues of significant visual similarity between skin cancer types and the limitations of individual models. By addressing inter-class visual similarity, showcasing the advantages of ensemble techniques in medical image analysis through a thorough performance comparison, and validating the effective use of ensemble models in enhancing accuracy.

The ISIC 2019 dataset, which included 25,331 dermoscopic images classified into eight classifications, such as basal cell carcinoma and melanoma, was utilised in the study. To preserve model generality, images were scaled to 224 x 224 pixels without undergoing significant preprocessing or lesion segmentation. In order to improve classification accuracy, two ensemble techniques—majority voting and weighted majority voting—were created to aggregate judgments from separate CNN models. The best-performing individual models were InceptionV3, DenseNet, and ResNet, each with an accuracy of roughly 91%, and InceptionV3. Weighted majority voting achieved 98.6% accuracy, and majority voting achieved 98% accuracy, indicating a significant superiority of the ensemble models over the individual models. These results show the efficacy of ensemble approaches in multiclass skin cancer classification, surpassing not just the accuracy of individual CNNs but also that of ordinary dermatologists.

2.5 Deep Convolutional Neural Network (DCNN) with Transfer Learning

The study uses a transfer learning-enhanced DCNN model and compares it with many pre-trained models, including ResNet, AlexNet, VGG-16, DenseNet, and MobileNet. The proposed approach has been improved and enhanced to achieve higher classification accuracy and lower execution times. To further improve feature extraction and the classification process, this method applies augmentation, preprocessing, and multiple transfer learning models.

2.5.1 DCNN

Ali et al. (2021) To improve accuracy and reduce computation times, the study will apply deep convolutional neural networks (DCNNs) combined with transfer learning to create an advanced deep learning model for the classification of skin cancer lesions from dermoscopic images. The more accurate DCNN model it presents improves classification accuracy while reducing calculation time, satisfying the need for faster and accurate diagnostic instruments in clinical settings. The study closes a significant gap in the field by showing how transfer learning may be used to achieve high accuracy, especially when there is a lack of data. The paper also provides a thorough comparison between the suggested model and a number of existing transfer learning models, demonstrating the model's advantages in terms of accuracy and efficiency.

The HAM10000 dataset, which included 10,015 dermoscopic images of skin lesions divided into benign and malignant groups, was used in the study. In order to evaluate the models, the dataset was divided into testing, validation, and training sets. Based on factors including accuracy, precision, recall, F1-score, and computational efficiency (execution time per epoch), the proposed DCNN model was evaluated compared to other transfer learning models. With a testing accuracy of 91.43%, the model outperformed the others and needed a lot less computing time every epoch, which made it more useful in real-world applications. The HAM10000 dataset is smaller, but the use of transfer learning, which makes use of pre-existing knowledge from large-scale datasets, significantly improved the model's performance and allowed it to achieve high accuracy.

2.6 AlexNet Transfer Learning

Using the AlexNet model that had already been trained, the study performed a transfer learning strategy. To categorise three distinct types of skin lesions, a softmax layer was added in place of the original final layer, which was intended to classify 1,000 classes in the ImageNet dataset. Backpropagation was used to fine-tune the model, using a low learning rate to prevent big changes in the convolutional layers while allowing considerable updates in the fully connected layers. The PH2 dataset performed data augmentation, which included rotation, to enhance the model's ability to generalise based on the restricted amount of accessible data.

2.6.1 AlexNet

Hosny et al. (2018) This study's main objective is to create an automated system for classifying skin lesions that can reliably differentiate between melanoma, atypical nevus, and common nevus by utilising deep learning and transfer learning approaches. The research addresses the weakness of earlier techniques that depended on major preprocessing and large datasets. It shows that even with a short dataset, transfer learning using a pre-trained model like AlexNet can achieve excellent accuracy in skin lesion categorization. The approach becomes more effective and less complicated to implement in clinical settings by streamlining the classification procedure and lowering the requirement for intensive preprocessing. The paper also offers a thorough comparison with current approaches, emphasising the advantages of the proposed approach.

The 200 RGB colour images contained in the PH2 dataset—40 of which were of melanoma, 80 of which were of common nevus, and 80 of which were of atypical nevus—were used in the study. Due to the small size of the dataset, data augmentation was used to increase the quantity of training images. Accuracy, sensitivity, specificity, and precision were used to measure the model's performance and provide a comprehensive evaluation of its categorization abilities. The predicted model's classification accuracy was a high 98.61%, with corresponding values for sensitivity, specificity, and precision of 98.33%, 98.93%, and 97.73%. These outcomes outperformed those of previous approaches, demonstrating the strength of merging data augmentation and transfer learning. The AlexNet model was especially well-suited to the particular task of skin lesion categorization when it was fine-tuned. In contrast with common machine learning methods and previous deep learning models, the suggested method showed enhanced classification accuracy and robustness across various types of lesions.

2.7 Fine-tuning using Ensemble Learning

The study used ensemble learning. By combining the advantages of various models, this method improves classification performance. The HAM10000 dataset was used for improving each CNN in the ensemble, adjusting the pre-trained models to the particular purpose of classifying skin cancer. The ensemble model used the various capacities for learning the various CNN architectures to enhance generalisation and decrease the possibility of overfitting.

2.7.1 Ensemble Learning

Samushonga et al. (2024) The purpose is to enhance the accuracy and reliability of skin cancer diagnosis by combining the advantages of several CNN architectures using ensemble techniques. When compared to single CNN models, the study shows that ensemble learning using optimised CNN models may greatly improve classification accuracy. Additionally, it fills in an understanding space regarding the integration of different CNN models by demonstrating how this strategy could reduce overfitting and boost prediction adaptability in challenging medical image processing tasks. Furthermore, the paper offers a thorough examination of several ensemble techniques, including majority voting and weighted voting, and their effects on the overall model.

The present research made use of the HAM10000 dataset, which includes more than 10,000 different dermoscopic images of skin lesions and is well-known for covering a wide range of skin lesion types. The performance of the ensemble model in classifying skin lesions was examined using measures including accuracy, precision, recall, and F1-score. Additionally, computational efficiency was assessed by comparing the training and inference durations of the model across various ensemble designs. With an accuracy of 93.7% overall, the recommended ensemble CNN model outperformed the individual models that were a part of the ensemble. Its reliability for skin cancer prediction was highlighted by its great accuracy and recall. When compared to single CNN designs, the error rate was greatly decreased by the application of ensemble learning. The pre-trained CNNs needed to be fine-tuned in order to better capture the complex patterns of skin lesions. This allowed the models to be adjusted to the unique features of the dermoscopic images. The study highlighted that this trade-off between computational cost and accuracy is appropriate in essential uses like skin cancer diagnosis, even if the ensemble model needed more computer resources according to the combined learning of several models.

2.8 TurkerNet architecture for CNN

The research introduces TurkerNet, a unique CNN architecture designed to be a lightweight model with less than 10 million trainable parameters. The input block, residual bottleneck block, efficient block, and output block make up TurkerNet's four main structural blocks. The architecture includes novel characteristics including squeeze-and-excitation blocks and residual connections, taking influence from MobileNet. These developments aim to reduce computing demands while increasing

the efficiency of feature extraction.

2.8.1 TurkerNet

Tuncer et al. (2024) In this work, a novel lightweight Convolutional Neural Network (CNN) called TurkerNet is presented. Its purpose is to provide excellent classification accuracy with a small amount of trainable parameters. The goal of the research is to create a model that effectively strikes a balance between computational cost and performance, making it especially appropriate for usage in restricted resource situations. TurkerNet solves the urgent demand for accurate yet efficient CNN models, especially in applications such as skin cancer classification, by significantly enhancing the balance between model performance and complexity. Squeeze-and-excitation and residual blocks are combined by TurkerNet to extract complex features without increasing computing requirements. In addition, the research introduces a new standard in the field by thoroughly comparing the novel model with the most advanced lightweight CNNs.

The study used a dataset of skin cancer images from Kaggle that was made available to the public and included images divided into malignant and benign. TurkerNet's performance was evaluated by dividing the dataset into training and testing sets. The study included in-depth mathematical expressions that outlined the functions of every TurkerNet design section.

$InputBlock : I1 = Conv33, 32, Stride = 2(I)$

$I2 = Conv33, 16, Stride = 2(I1)$

$I3 = Conv11, 64(I2)$

$ResidualBlock :$

$Rn = Conv1x1, C(Conv3x3, C/4(Conv1x1, C(Rn1))) + Rn1$

$OutputBlock :$

$Outn = Conv11, C(Conv11, 4C(GAP(Rn)))Rn$

Evaluation metrics including accuracy, recall, precision, and F1-score were used to evaluate the model's performance. By calculating the amount of trainable parameters and comparing TurkerNet's performance with that of other lightweight CNNs, such as MobileNetV2, SqueezeNet, and EfficientNetB0, computational efficiency was additionally evaluated. TurkerNet outperformed other lightweight CNNs with a test

classification accuracy of 92.12%. It performed well in terms of recall, accuracy, and F1-score; it was especially good at recognizing skin lesions that were cancerous. With just 2.2 million trainable parameters, TurkerNet keeps a small computing footprint despite its lightweight design. TurkerNet's scaling to handle larger data sets was noted in the study, indicating that it might find utility in areas other than skin cancer detection. A comparative study showed that TurkerNet produced better results than other top lightweight models and outperformed MobileNetV2 by over 5% in accuracy.

2.9 Analysis of AI Methods

Convolutional neural networks (CNNs), ontology-based expert systems, and machine learning (ML) algorithms are only a few of the AI approaches that are thoroughly reviewed in this paper. It compares and contrasts more modern techniques, such as comprehensive imaging systems and smartphone applications, with standard methods of dermatoscopic analysis of images. In order to improve the accuracy of skin lesion identification and classification, a major emphasis is placed on the integration of data from multiple sources, deep learning architectures, and complex image-processing techniques.

2.9.1 AI Applications

Strzelecki et al. (2024) In the case of skin cancer in particular, early identification is essential for efficient treatment and improved patient outcomes. The purpose of this study is to present a thorough overview of the state of artificial intelligence (AI) applications in skin cancer diagnosis and detection. It focuses on several AI methods and evaluates their advantages, disadvantages, and potential to change dermatological procedures. The work fills up a critical knowledge empty space about the current capabilities and limitations of artificial intelligence in dermatology by providing a comprehensive review. Future research aimed at improving the usability and dependability of AI in healthcare settings might benefit significantly from the consideration of difficulties like the need for vast data sets and the interpretability of AI models. The study addresses the drawbacks of single methods by promoting the integration of several AI techniques and data sources to improve diagnosis accuracy.

The study, which focuses on AI techniques applied to skin lesion analysis, makes use of a wide range of previously conducted research and clinical data. It gives

special attention to important datasets like ISIC (International Skin Imaging Collaboration) and HAM10000, which are widely used in AI model training for skin cancer diagnosis. The study highlights notable developments in the use of ensemble learning techniques and convolutional neural networks (CNNs), both of which have significantly improved the precision and dependability of skin cancer diagnosis. The identification of malignant lesions with high sensitivity and specificity has been shown to be achievable with techniques like transfer learning and a combination of deep learning architectures.

Despite these developments, the paper notes ongoing difficulties, such as certain AI models' "black-box" personality, which can make them difficult to understand and damage medical professionals' confidence. Furthermore, one major obstacle to the broad use of AI in clinical practice is its substantial dependence on huge, well-annotated datasets. The study highlights the need for more validation and clinical integration, even though AI systems have the potential to increase diagnostic accuracy, decrease the workload for dermatologists, and promote early diagnosis of skin cancer especially in places with limited resources. These technologies cannot be completely trusted and widely used until they have undergone extensive testing and been accepted into standard practice.

2.10 Ant Colony Optimization (ACO) with Pre-Trained CNN in a Hybrid Approach

The study employs a hybrid methodology that integrates a pre-trained Convolutional Neural Network (CNN) model, EfficientNetB0, with Ant Colony Optimization (ACO) for feature selection. Initially, deep features are extracted from both the raw and preprocessed images using EfficientNetB0. These extracted features are then combined through serial feature fusion to create a comprehensive feature vector. ACO is subsequently applied to this fused vector to refine and select the most pertinent features by mimicking the pheromone trail-following behaviour of ants. The final step involves classifying these optimised features using different Support Vector Machine (SVM) kernels, including Linear, Quadratic, and Cubic SVMs.

2.10.1 Ant Colony Optimization (ACO)

Imran et al. (2024) Using a feature optimisation technique inspired by nature and a pre-trained Convolutional Neural Network (CNN), this study seeks to increase the accuracy of skin cancer categorization. In particular, the study aims to improve the accuracy of skin lesion detection by using several Support Vector Machine (SVM) classifiers in conjunction with an improved feature selection procedure. A revolutionary method for detecting skin cancer is provided by the creative combination of EfficientNetB0 and Ant Colony Optimization (ACO), which increases computing efficiency and accuracy. The work reduces the high computing costs that are usually associated with deep learning models by optimising feature selection using ACO, making the suggested approach possible for real-time applications. The study not only shows higher accuracy but also provides a thorough investigation of the effects of optimization, feature fusion, and preprocessing methods on model performance.

The ISIC (International Skin Imaging Collaboration) dataset, which includes dermoscopic images of skin lesions categorised as benign or malignant, provided a unique dataset for the study. Several preparation approaches are used to improve this dataset. The pheromone update formula for Ant Colony Optimization (ACO) is explained in entirely in the paper:The ISIC (International Skin Imaging Collaboration) dataset, which includes dermoscopic images of skin lesions categorised as benign or malignant, provided a unique dataset for the study. Several preparation approaches are used to improve this dataset. The pheromone update formula for Ant Colony Optimization (ACO) is explained in entirely in the paper

After feature fusion, the suggested algorithm used the Cubic SVM classifier to obtain an outstanding accuracy of 99.01%, outperforming other classifiers and earlier techniques. Furthermore, the paper emphasises how the use of ACO for feature optimization led to important improvements in training time and prediction performance. Combining serial feature fusion with ACO reduced dimensionality while maintaining essential data from both the original and treated datasets, improving classification performance.

2.11 Literature Review Summary

The literature review highlights the different approaches and methodologies that have been used in the past for the classification of skin cancer and related fields. It specifically highlights the use of ensemble learning techniques and Convolutional

Neural Networks (CNNs). This is summarised in Table 2.1 below. It can be seen from the studied papers, which cover the years 2018 to 2024, that significant advancements have been made in the connected fields of transfer learning, fine-tuning, and the use of light architectures like TurkerNet. Such experiments by Shete et al. (2021) and Chaturvedi et al. (2020), use CNNs like VGG16, ResNet, and InceptionV3 with extremely high accuracy rates of over 90 percent. Ensemble learning frequently works with these CNNs to improve performance.To improve generalisation and usefulness, it must also overcome issues with size, computing difficulty, and larger datasets with more variance.

Research such as Haggenmüller et al. (2021) bring focus to the exciting chance for CNNs to perform better than human experts, but they also highlight some of their problems with regard to the diversity and variability of clinical data within datasets. Similar to this, Hosny et al. (2018) and Ali et al. (2021) achieve excellent results on transfer learning; however, this is because they use smaller, more specific datasets, such as PH2 or HAM10000. The same review covers a number of novel approaches, such as those created by Imran et al. (2024), who use pre-trained CNNs or use Ant Colony Optimization in hybrid methods that achieve high accuracy but with increasing computing needs. Although there are still issues with data imbalance, computational resources, and the requirement for more extensive testing on a variety of datasets, the literature normally shows a pattern of high accuracy in settings with less data.

Table 2.1: Literature review summary

#	Study	Method Description	Pros	Cons
1	**Shete et al., 2021**	Transfer Learning with CNNs (VGG16, ResNet)	High accuracy (90.51%) on HAM10000 dataset, effectiveness in limited data	Challenges with data imbalance require further improvement
2	**Haggenmüller et al., 2021**	Systematic Review of CNNs vs. Human Experts in Skin Cancer Classification	CNNs show potential, comparable or superior to human experts in some cases	Lack of clinical diversity, experimental settings may limit real-world applicability
3	**Chaturvedi et al., 2020**	Transfer Learning and Fine-Tuning with CNNs (Xception, ResNeXt101, InceptionV3, etc.)	High accuracy (93.20%), benefits of ensemble models	Increased computational complexity due to ensemble models
4	**Kausar et al., 2021**	Ensemble Learning and Fine-Tuning (InceptionV3, DenseNet, ResNet)	High accuracy (98.6%) with weighted majority voting, improved classification performance	Increased computational cost and complexity
5	**Ali et al., 2021**	Deep Convolutional Neural Network (DCNN) with Transfer Learning	High accuracy (91.43%), reduced computation time	Limited by smaller dataset (HAM10000)
6	**Hosny et al., 2018**	AlexNet Transfer Learning with data augmentation	High accuracy (98.61%), effective on small datasets	Limited to a specific dataset (PH2)
7	**Samushonga et al., 2024**	Fine-tuning using Ensemble Learning	High accuracy (93.7%), reduced overfitting, improved generalization	Increased computational resources required
8	**Tuncer et al., 2024**	TurkerNet architecture for CNN, lightweight with fewer parameters	High accuracy (92.12%), low computational cost	Limited testing on larger datasets
9	**Strzelecki et al., 2024**	Analysis of AI Methods (CNNs, ontology-based expert systems, ML algorithms)	Comprehensive review, identifies AI strengths and weaknesses	Dependency on large, well-annotated datasets
10	**Imran et al., 2024**	Ant Colony Optimization (ACO) with Pre-Trained CNN (EfficientNetB0) in a Hybrid Approach	Very high accuracy (99.01%), efficient feature selection	Increased computational complexity due to ACO

Chapter 3

Methodology

3.1 Overview

This research presents a very accurate deep learning model using CNN for improved skin cancer diagnosis, classification, and segmentation based on dermoscopic images. Vgg19, Inception, and EfficientNet are parts of the Transfer Learning that improve skin cancer detection accuracy, allowing for earlier diagnosis and better patient outcomes. To increase the models' robustness, more validation and testing will be done using data augmentation techniques, such as scaling, rotating, and flipping images on publicly available skin cancer datasets. Any such method of available data augmentation can only make the diagnostic process more difficult.

This research aims to build accurate CNN models for the identification and classification of skin cancer lesions, which could offer professionals an honest second opinion when they need it. The metrics will be taken into consideration while testing the model for accuracy, sensitivity, specificity, and F1 score. This will assist in closing the gaps in the literature currently providing deep learning in dermatology and make the model more clinically helpful. In conclusion, the present work indicates to improve early skin cancer diagnosis by increased integration of deep learning, thus closing the gap between experimental models and real-world medical applications.

3.2 Research Framework

Figure 3.1 presents a comprehensive research framework to improve the diagnosis of skin cancer through the use of convolutional neural networks and transfer learn-

ing. First HAM10000 dataset public-accessible skin cancer datasets are collected and preprocessed, this includes normalisation, data augmentation (scaling, rotating, flipping), and image resizing to ensure consistency. CNN architectures are created with basic elements including convolution, activation, pooling, and fully connected layers throughout the model-building stage. Transfer learning models, such as InceptionV3, EfficientNet, and VGG19 are combined, and new layers are fine-tuned to adjust to the current work, while basic layers remain blocked. To maximise its performance, the model is then trained by applying methods such as data splitting, batch processing, early stopping, and learning rate scheduling. To make sure the model fits the medical requirements, its accuracy, sensitivity, specificity, and F1 score are evaluated after training. After validation, the model is used to predict and diagnose skin cancer, offering a useful tool for medical diagnosis.

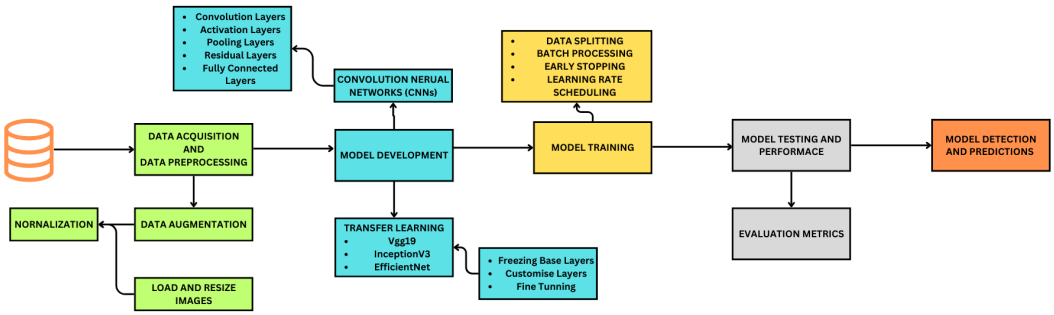

Figure 3.1: Overall research framework

3.2.1 Datasets

The first step in any Deep learning and machine learning project is data collection. The quality and diversity of the dataset are important factors in the diagnosis of skin cancer. The most popular dataset for detecting skin cancer that is accessible to the public is the HAM10000 dataset (Tschandl et al. 2018). The comprehensive collection of high-resolution dermatoscopic images is designed to aid research in building automated algorithms for recognising and classifying skin diseases. In general, one of the most well-known sources of skin imaging data is the ISIC Archive.

Ten different types of skin diseases are represented by the images included in this dataset, and each presents a different set of difficulties for machine learning, deep learning models and human dermatologists. These ten classes are as follows and shown in Figure 3.2:

Melanoma: A skin cancer that starts in melanocytes, the cells that give skin its colour, and has the potential to be deadly. Early disease detection is essential for effective treatment.

Actinic Keratosis: A skin area that is rough and scaly due to years of sun exposure. It is regarded as a precancerous condition with a chance to progress to squamous cell carcinoma.

Basal Cell Carcinoma: The most common form of skin cancer is marked by abnormal, rapidly growing growths or lesions in the skin's basal cells, usually as a result of frequent sun exposure.

Dermatofibroma: A hard, raised mass that is usually a benign fibrous skin lesion. Although mostly benign, it may be associated with more severe illnesses.

Nevus: A benign skin lesion made up of melanocyte clusters. Although most nevi are benign, some varieties, such as dysplastic nevi, are precancerous and have a chance to progress to malignant melanoma.

Monkey Pox: A rare viral illness that typically appears as rashes on the skin with flu-like symptoms. Even though they were often limited to isolated outbreaks in sub-Saharan Africa, they have recently gained worldwide concern.

Pigmented Benign Keratosis: A benign skin lesion that is frequently incorrectly identified as a dark, warty patch of skin. It is common for older persons and hardly requires medical attention.

Seborrheic Keratosis: A benign, noncancerous skin growth that looks like a wart and can be light tan, black, or brown in appearance. Although frequently misidentified as other, more serious skin conditions, these lesions are generally benign.

Squamous Cell Carcinoma: The most common kind of skin cancer starts in the skin's squamous cells. It is more aggressive than basal cell carcinoma and, if left untreated, can spread to other regions of the body.

Vascular Lesion: This is a general referral for skin lesions that involve blood vessels, hemangiomas and spider veins are two examples of these types of lesions. Although these skin lesions are mostly benign, cosmetic treatment may be necessary.

Figure 3.2: 10 Classes of Data

3.3 Methodology Pseudo-code

Data Collection Load images from dataset directories and preprocess them by resizing and normalizing.

```
load_images(data_dir):
    for file in data_dir:
        img = load_image(file)
        img_resized = resize(img, (224, 224))
        img_normalized = img_resized / 255.0
    return images
```

Data Augmentation Apply transformations like rotation, flipping, and zooming for better generalization.

```
augment_data(image):
    image_rotated = random_rotate(image)
    image_flipped = flip(image)
    image_zoomed = zoom(image)
    return [image_rotated, image_flipped, image_zoomed]
```

CNN Model Development Load a pre-trained model (e.g., EfficientNet, VGG19), freeze base layers, and add custom layers for fine-tuning.

```
build_model(pretrained_model):
    base_model = load_pretrained_model(pretrained_model,
                                       include_top=False)
    freeze_layers(base_model)
    custom_layers = add_custom_layers(base_model)
    model = compile_model(custom_layers, optimizer='adam',
                          loss='categorical_crossentropy')
    return model
```

Add Custom Layers Add dense and dropout layers for classification.

```
add_custom_layers(base_model):
    x = base_model.output
    x = GlobalAveragePooling2D()(x)
    x = Dense(512, activation='relu')(x)
```

```
    x = Dropout(0.5)(x)
    predictions = Dense(num_classes, activation='softmax')(x)
    return Model(inputs=base_model.input, outputs=predictions)
```

Model Training Train the model using the training dataset and validate it.

```
train_model(model, train_data, val_data):
    early_stopping = EarlyStopping(monitor='val_loss', patience=5)
    model.fit(train_data, validation_data=val_data, epochs=100,
              batch_size=32, callbacks=[early_stopping])
```

Model Evaluation Evaluate the model using accuracy, sensitivity, specificity, and F1-score metrics.

```
evaluate_model(model, test_data):
    accuracy = model.evaluate(test_data, metric='accuracy')
    sensitivity = calculate_sensitivity(test_data)
    specificity = calculate_specificity(test_data)
    f1_score = calculate_f1_score(test_data)
    return accuracy, sensitivity, specificity, f1_score
```

3.4 Model Implementation

3.4.1 Data Acquisition and Preprocessing

The initial step in the process was compatibility with the necessary libraries to set the computational environment for the methodology. In this case, runtime errors were avoided, and results were consistent across multiple executions. The system was checked in this line by verifying that the Python version was at least 3.7. This is crucial since many modern deep learning libraries and frameworks, like TensorFlow and Keras, depend on Python versions 3.7 and above.

Data acquisition is from ISIC Archive a publicly accessible source that involves uploading skin cancer image data to a Google Drive folder. The images included different classes based on the type of skin cancer and whether it was benign or not. First, the data set was arranged into some subdirectories, each of which corresponded to a different class label.

Key steps in data preprocessing:

Load and Resize Images: Using the functions of the OS library to browse the file system and load images from the dataset directory. The images were downsized to a fixed resolution of 224 by 224 pixels. The decision was made since pre-trained CNN, like the CNN for the ImageNet dataset, typically uses this size for fine-tuning for particular tasks, including the classification of skin cancer.

Normalisation: This learning process is made available by normalising pixel values. Images are usually standardised to have zero mean and unit variance, which speeds up their meeting in training, or they are normalised by rescaling pixel values in the range of 0 to 1 using 225, or maximum pixel value.

Data Augmentation: It is used to prevent overfitting, particularly given the frequently restricted availability of classified medical images. By introducing random image changes, like the following, these methods affect the training dataset and intentionally expand its size:

- Rotation: To simulate various viewing angles, random rotations, such as those between 20 and 30 degrees, are used
- Flipping: Both horizontal and vertical flips are performed to make sure focused features are not learned by the model.
- Zooming and Scaling: To approximate changes in distance from the camera, use different zooms.

3.4.2 Model Development:

Convolutional Neural Network: A convolutional neural network provides the foundation for the model building used in the classification of skin cancer. CNN performance in recognising images can be related to its power to automatically and adaptively learn spatial feature hierarchies with backpropagation. A CNN normal design consists of the following important elements as shown in Figure 3.3:

Convolution Layers: To automatically generate spatial hierarchies of features from input images, it is expected that CNN architecture was designed using multiple convolutional layers followed by pooling layers. In a CNN, the convolutional layer recognises patterns that include simple ones like edges and textures to complex types seen in higher layers.

Activation Function: To include non-linearity into the model, an activation function typically a Rectified Linear Unit (ReLU) is applied to each convolution process.

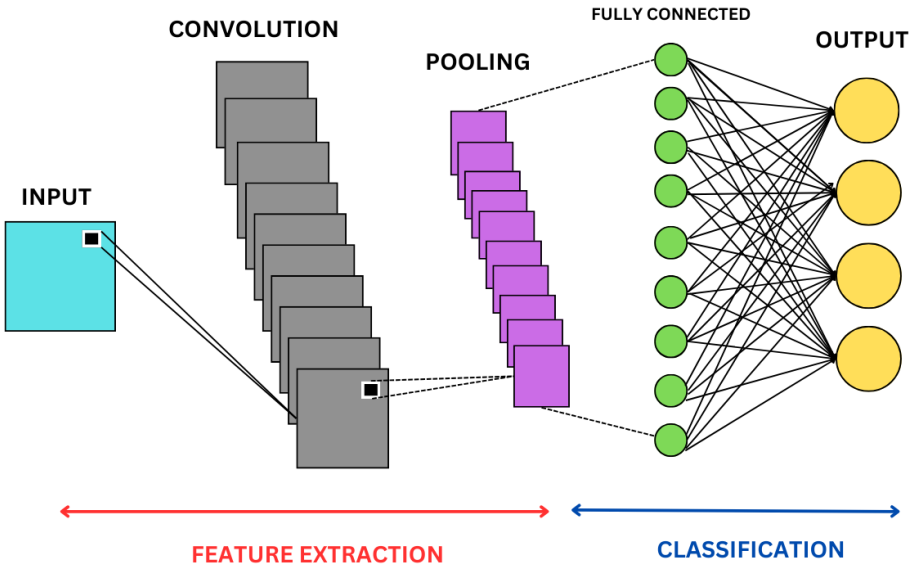

Figure 3.3: CNN Architecture: Feature Extraction and Classification.

Since the network can learn from mistakes and enhance the input-to-output mapping, the non-linearity quality is important.

$ReLU(X) = max(0, X)$

For any input X, it means:

$If X > 0, then ReLU(X) = X.$

$If X <= 0, then ReLU(0) = 0.$

Pooling Layers: CNN use pooling operations, like mean and max pooling, to reduce the spatial dimensions of feature maps. This reduces computation time and prevents overfitting. Max pooling keeps the most prominent feature by choosing the highest value from each region it slides over, while mean pooling averages the values to create a smoother representation. These techniques help downsample feature maps, which lowers the number of parameters and computations and results in a more robust model against variations in feature location. The following Figure 3.4 shows how mean pooling summarises the information in that region while max pooling preserves significant features.

Residual Layers: Residual Layers in CNN introduce a shortcut connection that avoids one or more convolutional layers, allowing the model to learn the residual functions itself. We are attempting to fit a different function, $F(x) = H(x)$ x rather

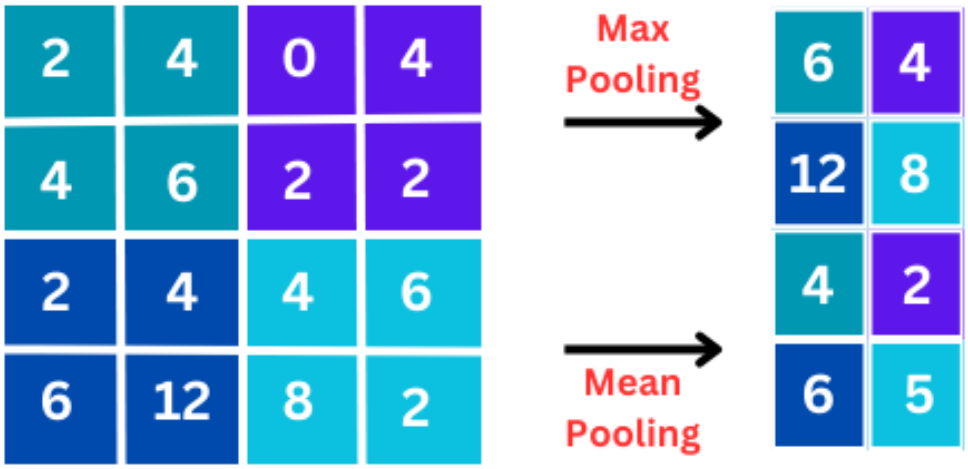

Figure 3.4: Mean And Max Pooling Summary

than learning some mapping $H(x)$. This method solves the vanishing gradient issue and improves performance, making it helpful for training deep CNN. Instead of learning the entire transformation, it allows the network to concentrate on learning differences or residuals for the input

Fully Connected Layers: These convolutional and pooling layers are followed by fully connected layers, which flatten the output. To predict the final output classes, these layers somehow apply all the features that the convolutions have learned. Here, the last layer often uses a variation of softmax activation, which provides the model with the probabilities of each class so it can identify and assign a class label to an input image.

Regularisation Techniques: The convolution layer and fully connected layers, Batch Normalisation and Dropout layers were added to enhance the model's capacity to generalise effectively to previously unknown data. By pushing the model to develop stronger features instead of becoming too dependent on some nodes, dropout helps minimise overfitting by randomly setting some of the input units to zero during training.

3.4.3 Transfer Learning

One important technique used was transfer learning, in addition to training CNN from scratch. This method makes use of a model that has previously been trained on a large dataset such as ImageNet, which includes millions of images in a variety of categories and fine-tuning it to fit a new task. Transfer Learning includes the

following:

Pre-trained Model Selection: InceptionV3, EfficientNet, and VGG19 are common models for classifying images transfer learning. These models have strong feature representations that they can use for new tasks that need less data because they were pre-trained on the ImageNet dataset. Because of its deep architecture and recent successes in medical image classification, VGG19 in particular is frequently chosen for the classification of skin cancer.

Freezing the Base Layers: To capture general features like edges and textures, this normally happens by freezing the first few layers of the pre-trained model. Thus, information about the ImageNet dataset is well saved during fine-tuning as the weights of those layers are not changed.

Customise the Top Layers: The final layers of the pre-trained model are swapped out for new fully connected layers tailored specifically to the skin cancer classification task. These new layers are trained using the skin cancer dataset, enabling the model to learn features essential for differentiating between various skin conditions.

Fine tuning: After the top layers have been trained, the model may go through a process called fine-tuning in which some of the pre-trained model's deeper layers are unfrozen and retrained at an extremely low learning rate. This is an important stage because it helps the model to fine-tune the base features it learnt from ImageNet to better match the unique properties of images showing skin cancer.

3.4.4 Model Training and Validation

TensorFlow and Keras were used in the development of the transfer learning model as well as the customised CNN. The models were trained using an optimizer called Adam, which is well-known for its effectiveness and popularity in training deep learning models, and a categorical cross-entropy loss function, which is the conventional choice for multi-class classification problems.

Training Validation Split: The dataset was split into training and validation sets, in an 80:20 ratio. The validation set was used to evaluate the model after each epoch, assisting in ensuring that the model did not overfit the training set, while the training set was used to modify the model weights.

Batch Processing: Each batch with 32 images was used to process the images. Batch processing allows parallel computations, which speeds up training and helps manage memory usage.

Early Stopping: To avoid overfitting and reduce training duration, an early criterion was added. If, after a certain amount of epochs (patience), the validation accuracy did not improve, the training was stopped. By using this method, the model will continue to maintain the optimal weights that equal the highest validation accuracy.

Learning Rate Scheduling: During training, a learning rate scheduler was most likely used to change the learning rate. By decreasing the learning rate as the model converges and helps it settle into a local minimum, this technique helps fine-tune the model

3.5 Evaluation Metrics

The performance of the transfer learning and the CNN model was evaluated on a different test dataset. provide an improved understanding of the model performance, this evaluation also included metrics like accuracy, precision, recall, and F1 score in addition to more advanced evaluations like confusion matrix.

Accuracy: Accuracy, which evaluates the percentage of accurate predictions the model makes, was the primary metric used to evaluate the model's performance.

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

Precision, Recall and F1 score: These metrics were calculated to provide a more thorough analysis of the model's success, especially in cases where there may be imbalances in the dataset, i.e., classes with a higher sample count than others. The F1 score balances the difference between precision and recall by providing a harmonic mean of both measures. Precision counts the proportion of true positives among the presented positives, recall measures the proportion of true positives among the actual positives.

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

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

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Confusion Matrix A detailed breakdown of the model's performance across different classes was provided by a confusion matrix. This matrix provides important insights into areas in which the model might need to be improved by helping to identify which classes the model mistakes with other classes.

Chapter 4

Results

4.1 CNN Base Model

Figure 4.1 shows the model's performance steady improvement in training accuracy across 60 epochs, while validation accuracy changes and peaks at 0.54, which may indicate overfitting. The validation loss differs, showing difficulties with generalisation, while the training loss decreases slowly. Nevus is the most properly diagnosed class.

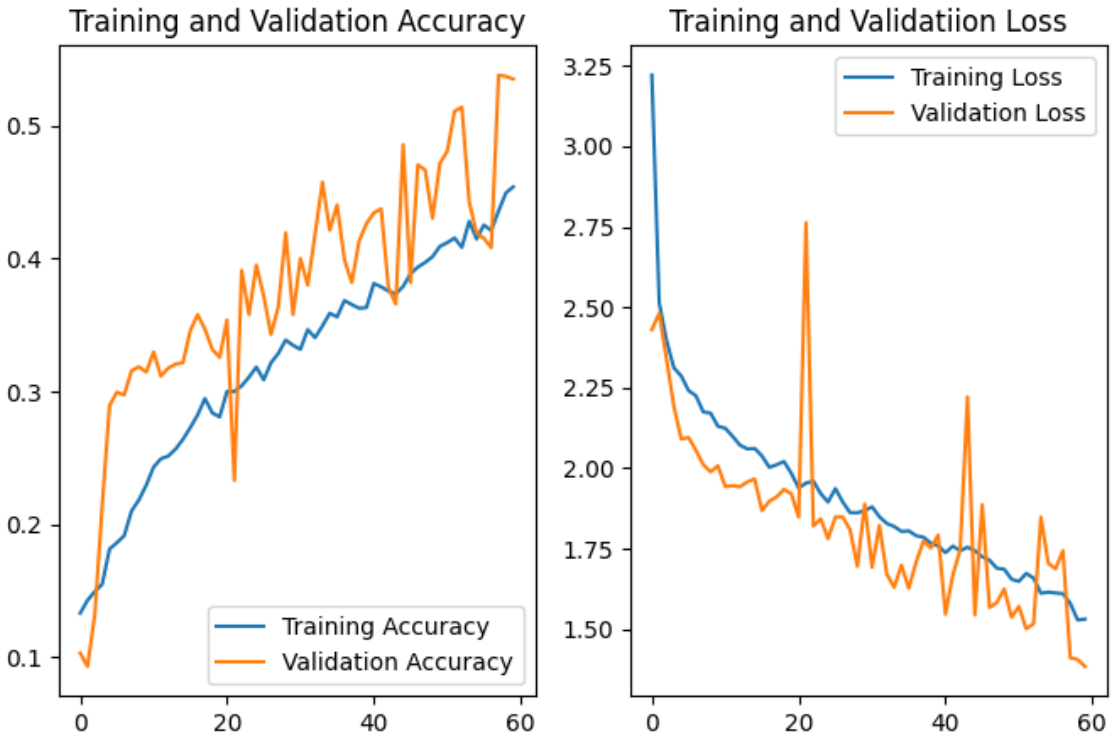

Figure 4.1: CNN Base Model Training and Validation Accuracy & Loss

according to the confusion matrix in Figure 4.2, Squamous Cell Carcinoma and Seborrheic Keratosis show high rates of misclassification, mainly with cancer types that show similarly. Nevus has the greatest F1 score (0.95) and Squamous Cell Carcinoma has the lowest (0.27), according to the Figure 4.3 classification report with a macro average F1 score of 0.54, the model achieves an overall accuracy of 54%.

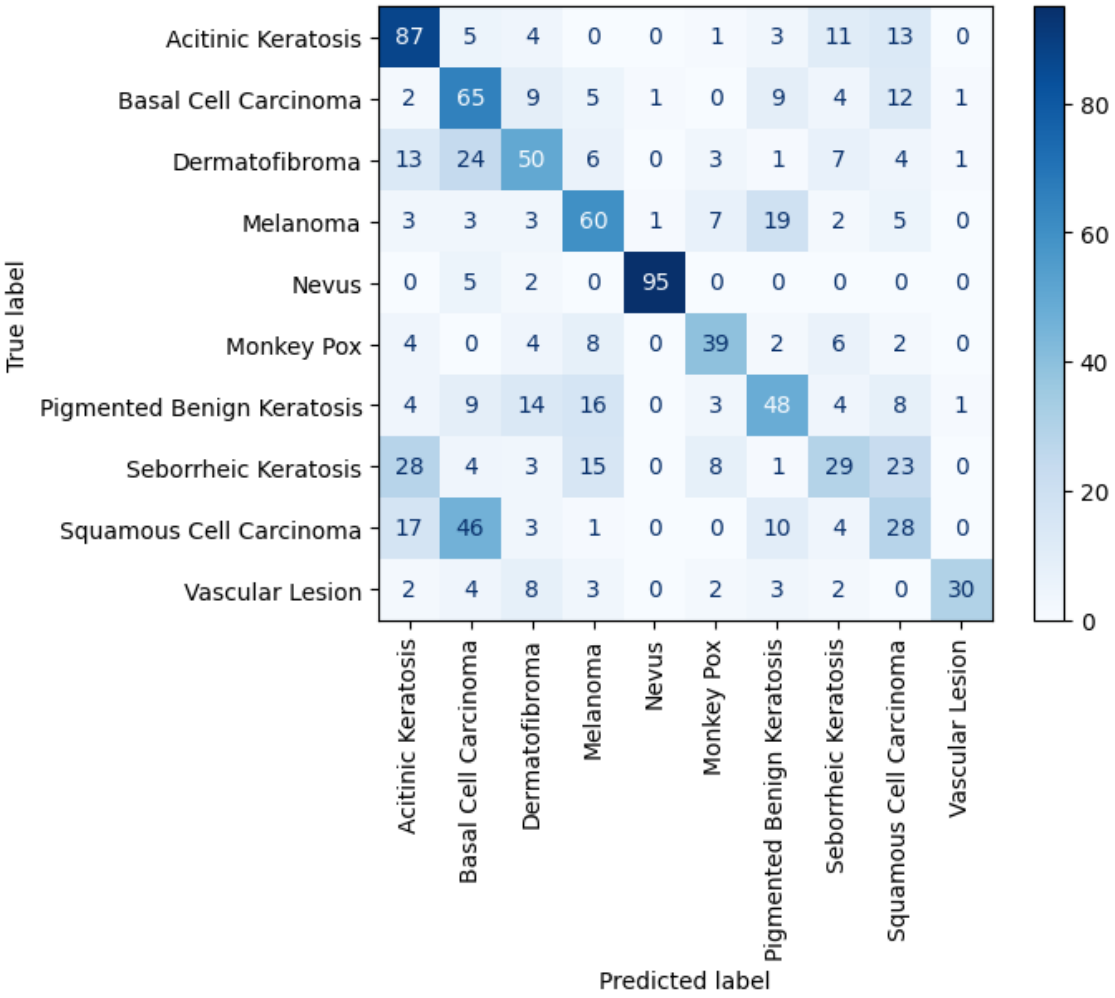

Figure 4.2: CNN Base Model Confusion Matrix

	precision	recall	f1-score	support
Acitinic Keratosis	0.54	0.70	0.61	124
Basal Cell Carcinoma	0.39	0.60	0.48	108
Dermatofibroma	0.50	0.46	0.48	109
Melanoma	0.53	0.58	0.55	103
Nevus	0.98	0.93	0.95	102
Monkey Pox	0.62	0.60	0.61	65
Pigmented Benign Keratosis	0.50	0.45	0.47	107
Seborrheic Keratosis	0.42	0.26	0.32	111
Squamous Cell Carcinoma	0.29	0.26	0.27	109
Vascular Lesion	0.91	0.56	0.69	54
accuracy			0.54	992
macro avg	0.57	0.54	0.54	992
weighted avg	0.54	0.54	0.53	992

Figure 4.3: CNN Base Model Classification Metrics

Discussion: The existing model achieves an average level of overall accuracy, showing good performance for certain classes like Nevus and Vascular Lesion. Overfitting, confusion between visually similar cancer types such as Squamous Cell Carcinoma and Basal Cell Carcinoma, and difficulty in generalising to unseen data are some of the major challenges it experiences. Various approaches can be used to address these problems and improve performance. While regularisation methods like L2 regularisation and dropout may decrease overfitting, data augmentation can improve generalisation by expanding the training set. The learning process can also be balanced by using class-weighted loss functions or oversampling to address class imbalance. The model's existing problems can be addressed by putting these changes into practice, which will improve the model's performance across all skin cancer classifications.

4.2 CNN Base Model (Residual Layers)

By epoch 60, In Figure 4.4, the model had 60% accuracy on the training and validation sets, but the validation accuracy was unreliable, indicating that generalisation problems would remain. Overall better generalisation was indicated by the training loss continuously declining and the validation loss improving with less split than the previous model. With 101 accurate predictions, Nevus was the category with the best classification while Squamous Cell Carcinoma and Seborrheic Keratosis were frequently misclassified as shown in Figure 4.5, However, 40 cases of Squamous Cell Carcinoma were misidentified as Basal Cell Cancer. Squamous Cell Carcinoma had

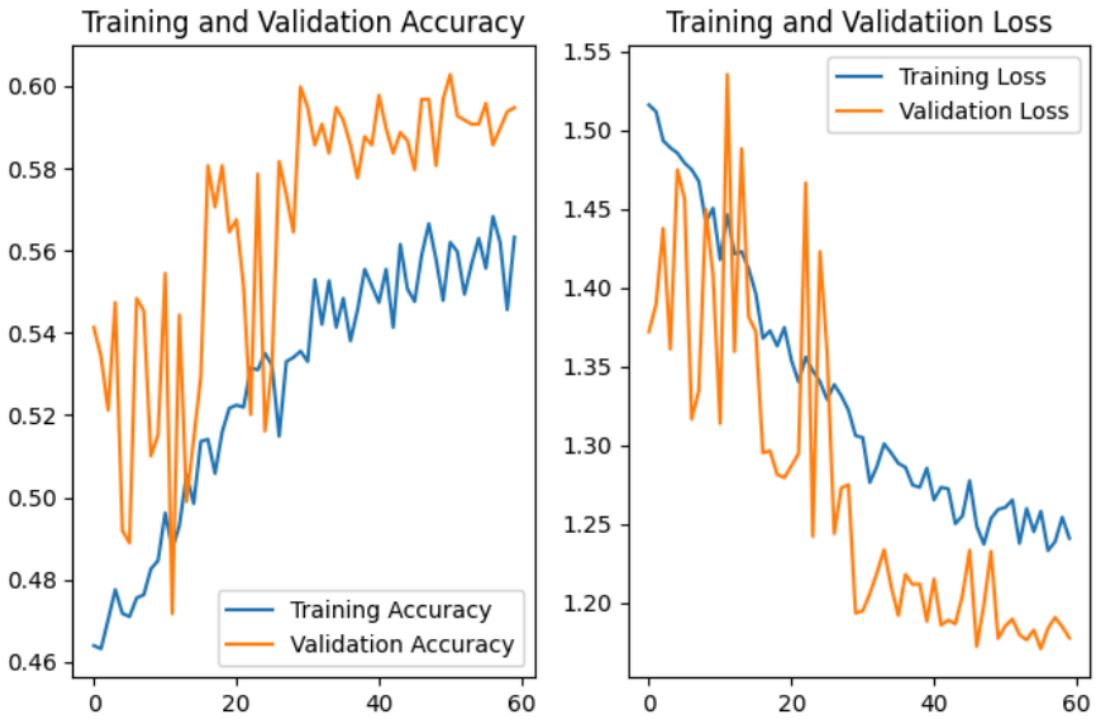

Figure 4.4: CNN Base Model (Residual Layers) Training and Validation Accuracy & Loss

the lowest F1 score of 0.32, whereas Nevus also had the highest F1 score of 0.99. In Figure 4.6 with a macro average F1 score of 0.60, the model's overall accuracy was 60%.

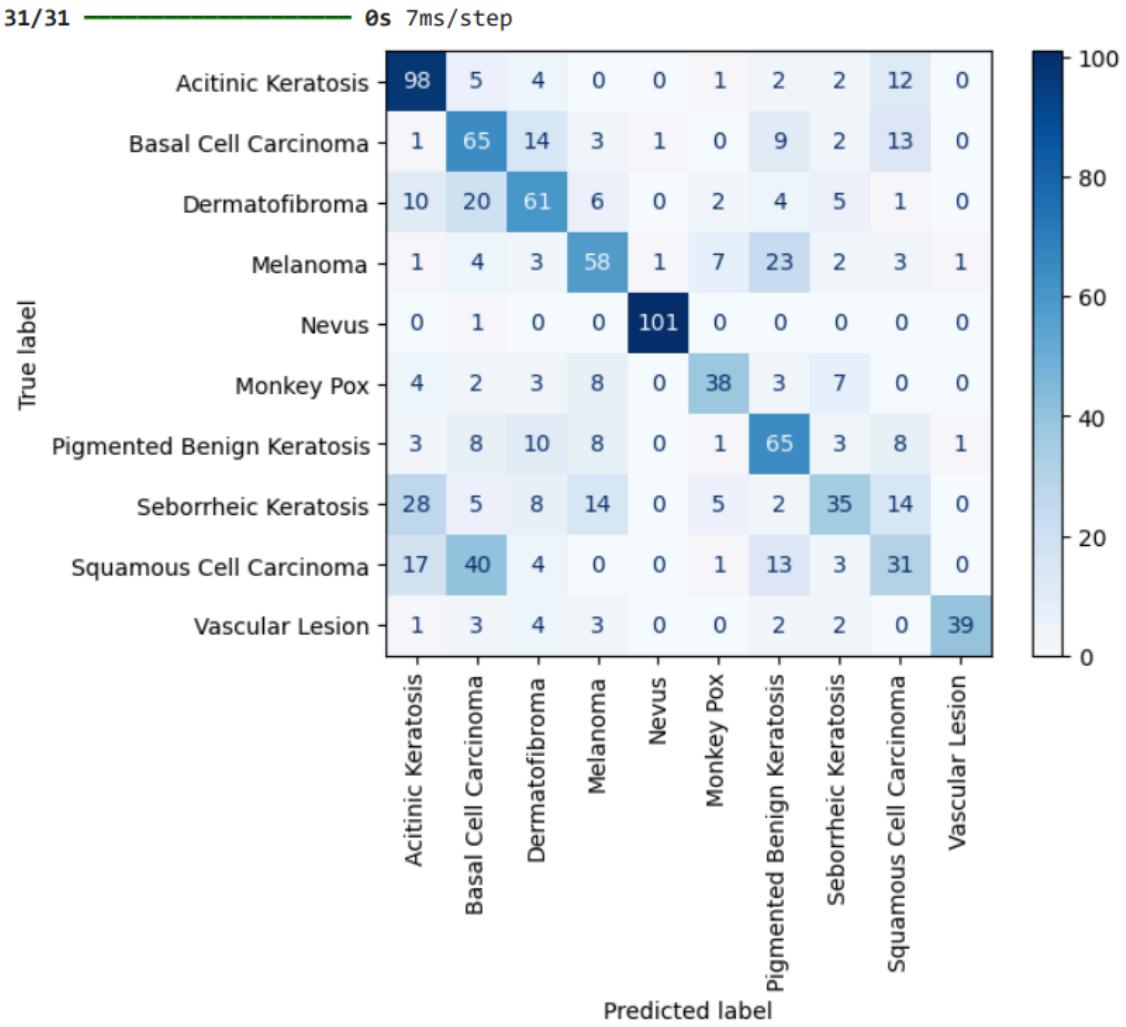

Figure 4.5: CNN Base Model (Residual Layers) Confusion Matrix

	precision	recall	f1-score	support
Acitinic Keratosis	0.60	0.79	0.68	124
Basal Cell Carcinoma	0.42	0.60	0.50	108
Dermatofibroma	0.55	0.56	0.55	109
Melanoma	0.58	0.56	0.57	103
Nevus	0.98	0.99	0.99	102
Monkey Pox	0.69	0.58	0.63	65
Pigmented Benign Keratosis	0.53	0.61	0.57	107
Seborrheic Keratosis	0.57	0.32	0.41	111
Squamous Cell Carcinoma	0.38	0.28	0.32	109
Vascular Lesion	0.95	0.72	0.82	54
accuracy			0.60	992
macro avg	0.63	0.60	0.60	992
weighted avg	0.60	0.60	0.59	992

Figure 4.6: CNN Base Model (Residual Layers) Classification Metrics

Discussion: Actinic Keratosis and Pigmented Benign Keratosis are two classes where the current model performs better in classification, indicating increased feature learning and slightly improved generalisation. The 60% shows this improved overall accuracy over previous models and the decreased variations in validation loss and accuracy. The model still has a lot of problems despite these advancements, such as ongoing changes in validation performance that point to overfitting and stability problems. Class confusion continues, especially between Squamous Cell Carcinoma and Seborrheic Keratosis. Addressing class imbalance with techniques like oversampling or class-weighted loss functions may help improve the model even more. Furthermore, using regularisation strategies like L2 regularisation or increased dropout could help in reducing validation instabilities. While hyperparameter adjustment may stabilise training and increase validation accuracy, incorporating various data augmentation could improve generalisation.

4.3 EfficientNetB0 Model

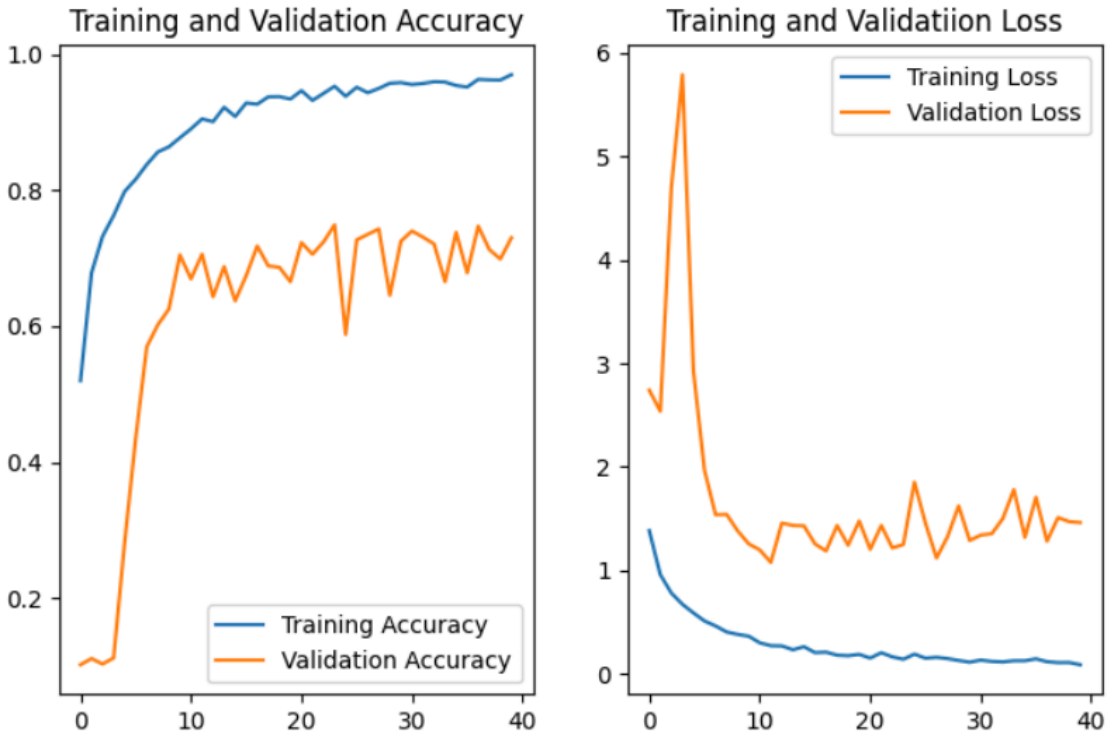

Figure 4.7: EfficientNetB0 Model Training and Validation Accuracy & Loss

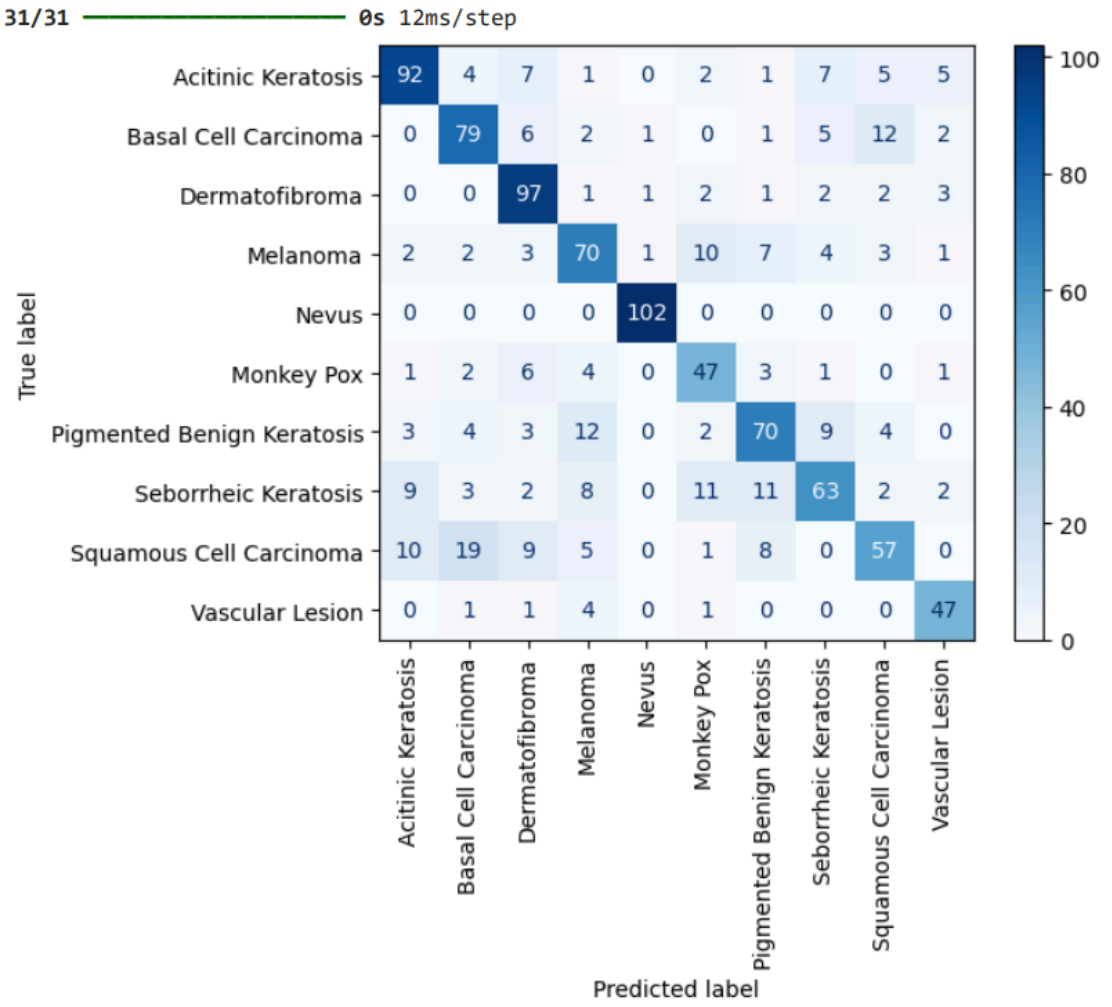

Figure 4.8: EfficientNetB0 Model Confusion Matrix

	precision	recall	f1-score	support
Acitinic Keratosis	0.79	0.74	0.76	124
Basal Cell Carcinoma	0.69	0.73	0.71	108
Dermatofibroma	0.72	0.89	0.80	109
Melanoma	0.65	0.68	0.67	103
Nevus	0.97	1.00	0.99	102
Monkey Pox	0.62	0.72	0.67	65
Pigmented Benign Keratosis	0.69	0.65	0.67	107
Seborrheic Keratosis	0.69	0.57	0.62	111
Squamous Cell Carcinoma	0.67	0.52	0.59	109
Vascular Lesion	0.77	0.87	0.82	54
accuracy			0.73	992
macro avg	0.73	0.74	0.73	992
weighted avg	0.73	0.73	0.73	992

Figure 4.9: EfficientNetB0 Model Classification Metrics

In Figure 4.7, the training accuracy of the model improves quickly, approaching 100%, while the validation accuracy improves at roughly 73%, showing good but still improvable generalisation. Training loss generally declines, although there may be some slight overfitting based on the difference between training and validation losses. Nevus is the best-classified class, according to the Figure 4.8 confusion matrix, with 102 correct predictions, Vascular Lesion and Dermatofibroma perform equally well. Even so, there are still a few minor misclassifications. such as in the relationship between seborrheic keratosis and squamous cell carcinoma were incorrectly classified as basal cell carcinoma. Squamous Cell Carcinoma got the lowest F1 score (0.59) but Nevus the highest (0.99). With a weighted F1 score of 0.73 and a macro-average of 73% as shown in Figure 4.9, the model's overall accuracy shows balanced performance across all classes.

Discussion: The latest model performs very well on important classes including Nevus, Dermatofibroma, and Vascular Lesion and has better generalisation to unknown data. The highest accuracy is 73%, mainly because of the usage of transfer learning. There are still difficulties when it comes to class confusion between types that appear similar, such as basal cell carcinoma and squamous cell carcinoma. Furthermore, there are still signs of overfitting, as seen by the difference between validation and training losses. Class-specific fine-tuning using class-weighted loss functions or focused data augmentation could help the model overcome these problems. Techniques like early stopping, L2 regularisation, or greater dropout could be used to reduce overfitting. Moreover, increasing feature extraction using advanced pre-trained models such as EfficientNet may improve the model's performance. By decreasing variations and enhancing generalisation even more, hyperparameter adjustment can also help in stable validation performance. These improvements, mainly for the harder classes, could increase accuracy and reduce misclassifications.

4.4 EfficientNetB0 (Freeze Layers):

Figure 4.10 shows the validation accuracy of the model settled at 78%, while fluctuations suggest possible overfitting. The model reached close to perfect training accuracy, close to 100%. Training loss decreased over time, but this difference in the validation loss indicates problems with unseen data.

Figure 4.11 shows, that Nevus is perfectly classified in the confusion matrix, while the misclassification rates for Actinic Keratosis and Pigmented Benign Keratosis were equally reduced. 13 Squamous Cell Carcinoma samples were incorrectly classi-

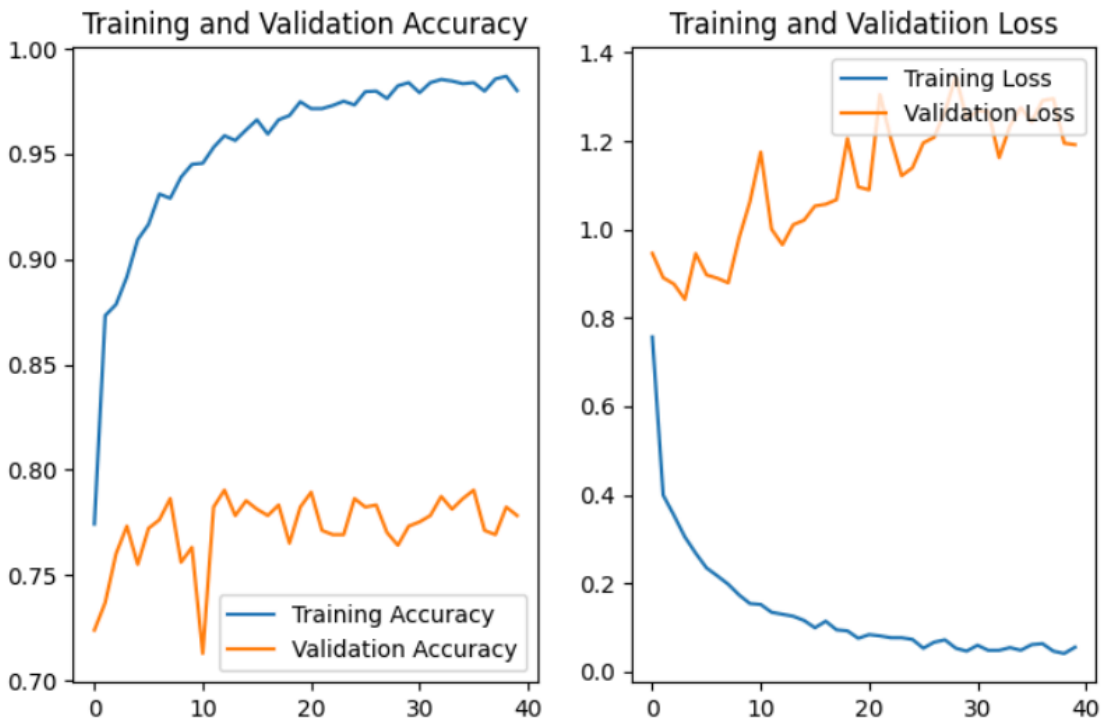

Figure 4.10: EfficientNetB0 Model (Freeze Layer) Training and Validation Accuracy & Loss

fied as Basal Cell Carcinoma, indicating that there is still a bit of confusion between Seborrheic Keratosis and Squamous Cell Carcinoma.

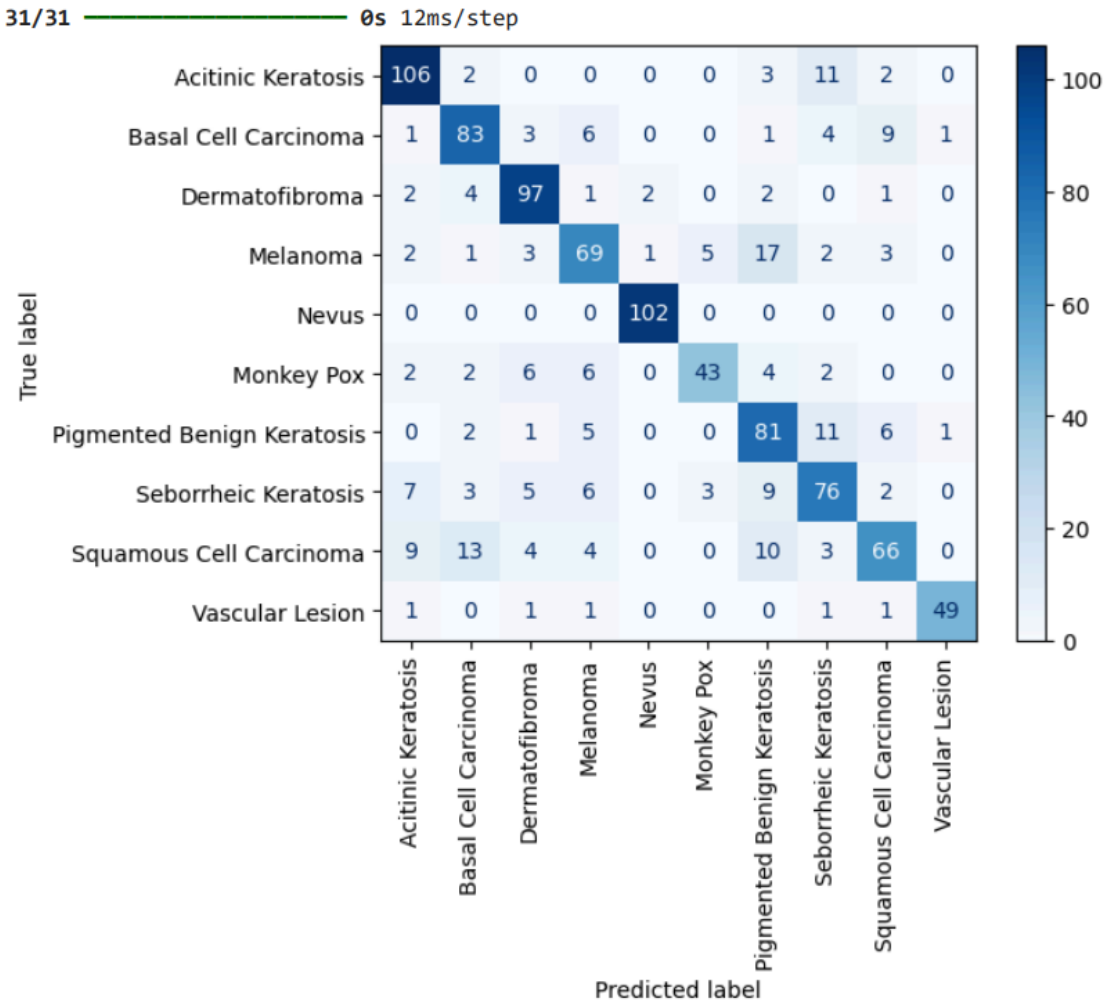

Figure 4.11: EfficientNetB0 Model (Freeze Layer) Confusion Matrix

Figure 4.12 shows that Nevus has the highest F1 score (0.99), with important contributions from Dermatofibroma (0.85) and Vascular Lesion (0.93). On the other hand, misclassification was a problem for seborrheic keratosis (0.68) and squamous cell carcinoma (0.66). With a macro-average F1 score of 0.78, the model's accuracy of 78% achieved overall, which is an important increase over previous methods.

	precision	recall	f1-score	support
Acitinic Keratosis	0.82	0.85	0.83	124
Basal Cell Carcinoma	0.75	0.77	0.76	108
Dermatofibroma	0.81	0.89	0.85	109
Melanoma	0.70	0.67	0.69	103
Nevus	0.97	1.00	0.99	102
Monkey Pox	0.84	0.66	0.74	65
Pigmented Benign Keratosis	0.64	0.76	0.69	107
Seborrheic Keratosis	0.69	0.68	0.69	111
Squamous Cell Carcinoma	0.73	0.61	0.66	109
Vascular Lesion	0.96	0.91	0.93	54
accuracy			0.78	992
macro avg	0.79	0.78	0.78	992
weighted avg	0.78	0.78	0.78	992

Figure 4.12: EfficientNetB0 Model (Freeze Layer) Classification Metrics

Discussion: The model performs better than previous models in classes such as Nevus, Vascular Lesion, and Dermatofibroma, reaching its maximum accuracy of 78%. But class confusion between Squamous Cell Carcinoma and Basal Cell Carcinoma, as well as overfitting signs caused by an error between training and validation loss, are still present. To reduce these issues, class-specific changes, enhanced regularisation, complex data augmentation, and hyperparameter tuning can assist in reduced overfitting, normalising validation outcomes, and enhancing generalisation for difficult classes.

4.5 VGG19 Model

Figures (4.13, 4.14, and 4.15) show the validation accuracy changes and then stabilises at around 60%, indicating instability when working with new, unseen data, while the training accuracy of the model continues to improve to 70%.

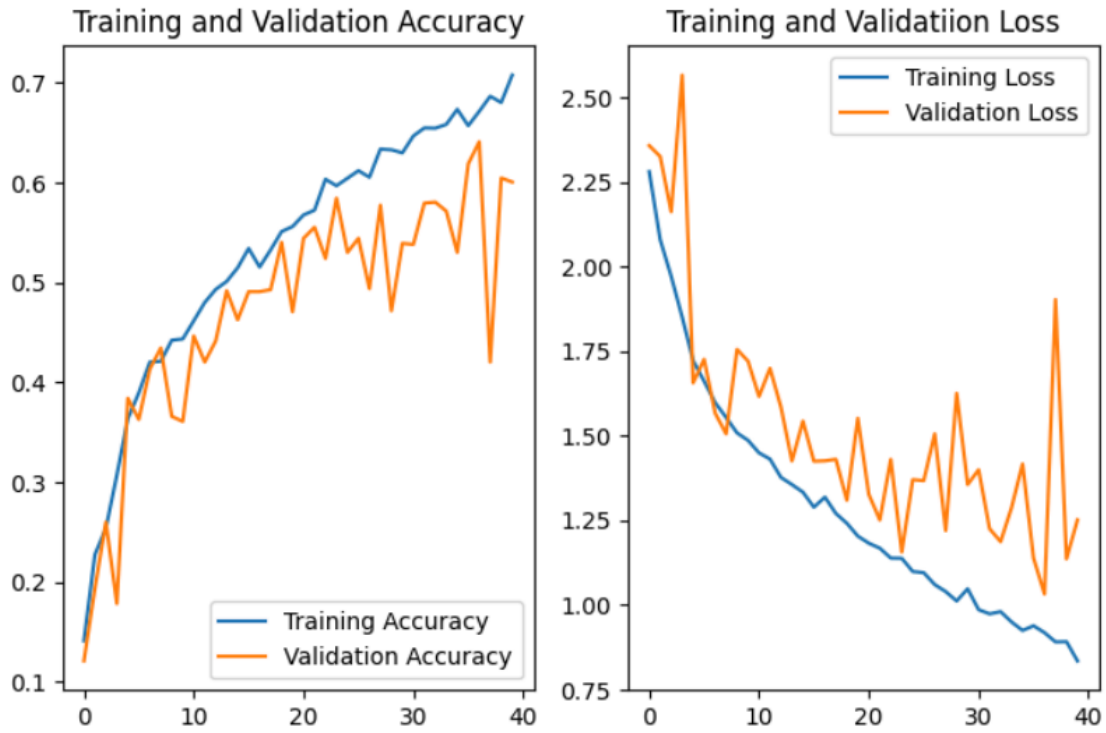

Figure 4.13: VGG19 Model Training and Validation Accuracy & Loss

The validation loss varies strongly, indicating the expected overfitting and poor generalisation, while the training loss smoothly falls, indicating effective learning. Nevus (101 right predictions, with only 1 misclassified) and Melanoma (85 correct predictions) show good classification results in the confusion matrix.

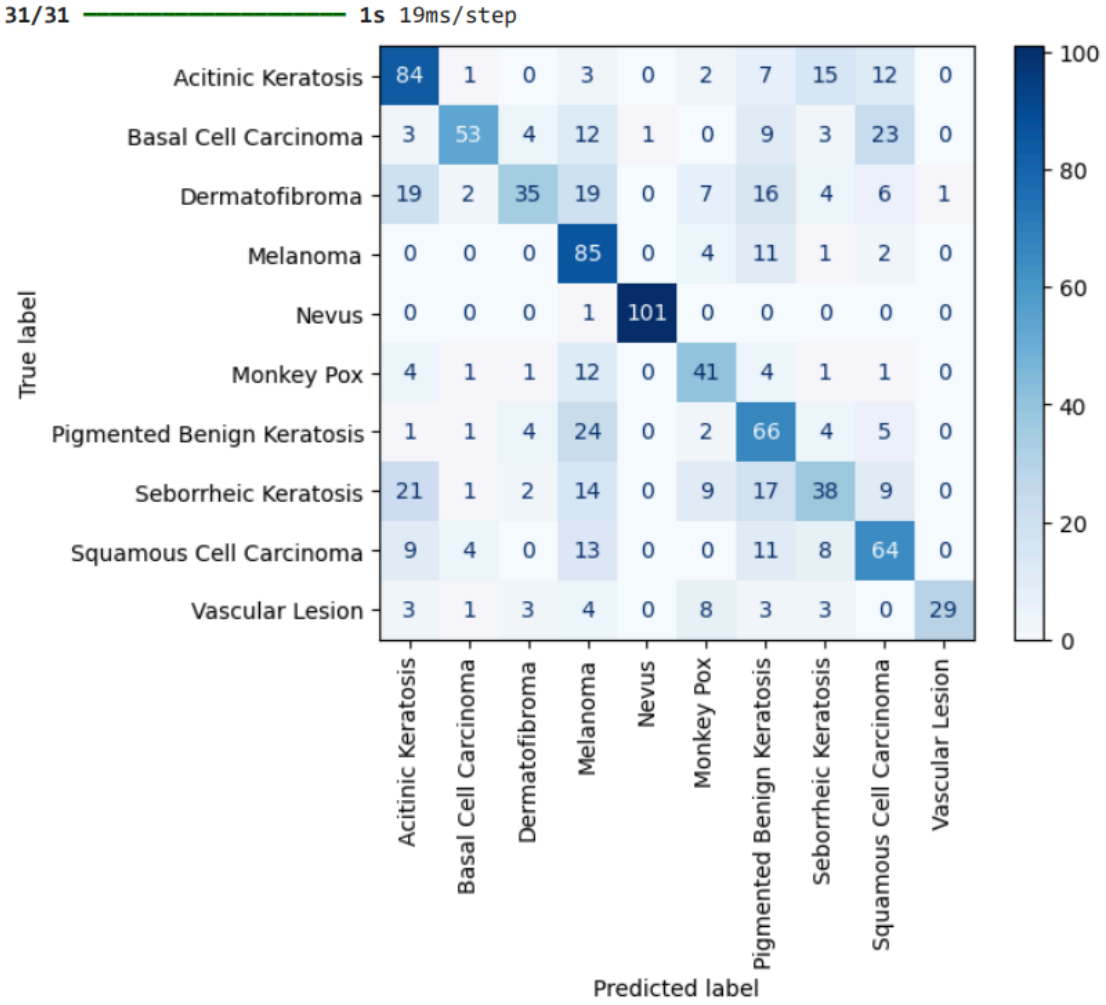

Figure 4.14: VGG19 Model Confusion Matrix

However, there is a high rate of misclassification in certain classes, such as Seborrheic Keratosis and Dermatofibroma. At 0.99, Nevus has the highest F1 score, followed by Dermatofibroma and Seborrheic Keratosis, at 0.44 and 0.40, respectively. With a macro-average F1 score of 0.60 and an accuracy of 60% overall, the model performs well in all classes.

	precision	recall	f1-score	support
Acitinic Keratosis	0.58	0.68	0.63	124
Basal Cell Carcinoma	0.83	0.49	0.62	108
Dermatofibroma	0.71	0.32	0.44	109
Melanoma	0.45	0.83	0.59	103
Nevus	0.99	0.99	0.99	102
Monkey Pox	0.56	0.63	0.59	65
Pigmented Benign Keratosis	0.46	0.62	0.53	107
Seborrheic Keratosis	0.49	0.34	0.40	111
Squamous Cell Carcinoma	0.52	0.59	0.55	109
Vascular Lesion	0.97	0.54	0.69	54
accuracy			0.60	992
macro avg	0.66	0.60	0.60	992
weighted avg	0.64	0.60	0.60	992

Figure 4.15: VGG19 Model Classification Metrics

Discussion: For important classes such as Nevus and Melanoma, the present model performs very well, getting excellent precision and recall, and showing slow drops in training loss, which indicates effective learning. Its difficulties with generalisation, however, can be seen from the unstable validation loss and the large gap between training and validation accuracy, which point to overfitting. The model's performance continues to decline in this area due to misclassifications, particularly between Dermatofibroma and Actinic Keratosis and Seborrheic Keratosis and Squamous Cell Carcinoma. To improve this, the model may be able to more accurately capture complex features unique to skin cancer by selectively fine-tuning the deeper layers through a transfer learning method. Additionally, by reducing overfitting and providing more consistent performance measures across various data subsets, k-fold cross-validation would provide a more accurate analysis of the model's generalisation.

4.6 VGG19 Model (Freeze Layer)

At epoch 40, the model's validation and training accuracies both reach 50%, showing no sign of overfitting and suggesting that the learning is fast but limited in scope. Figure 4.16 shows the training and validation losses drop slowly and level out at higher levels, showing difficulties in further improvement.

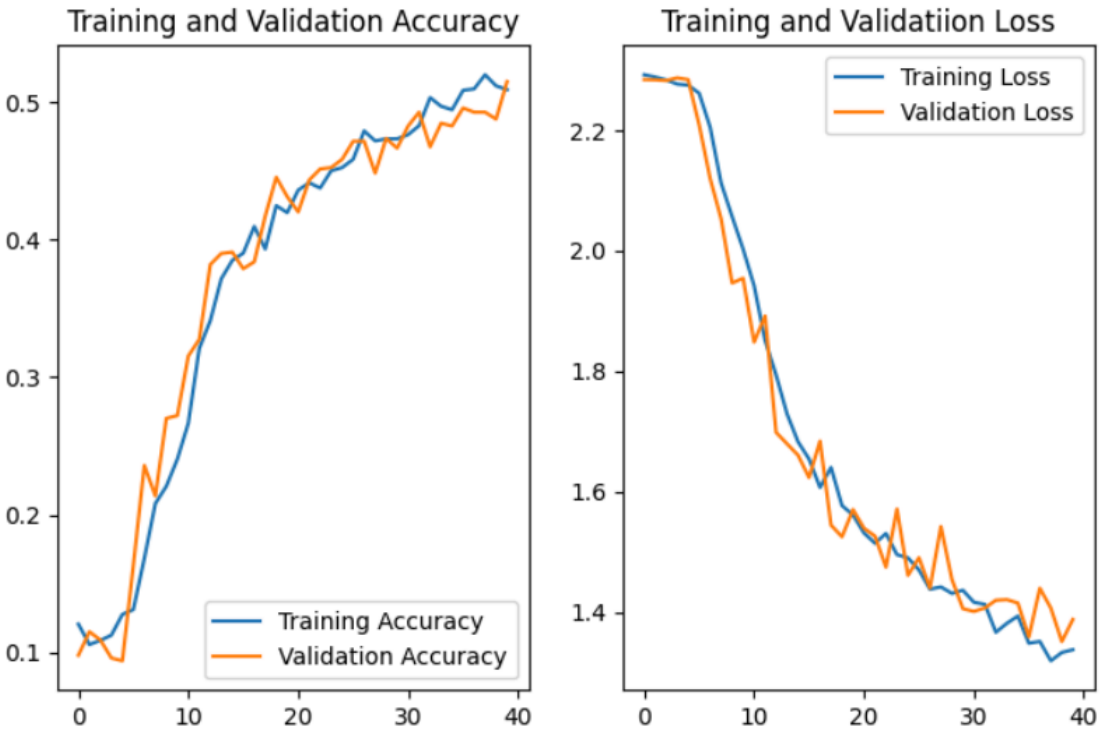

Figure 4.16: VGG19 Model (Freeze Layer) Training and Validation Accuracy & Loss

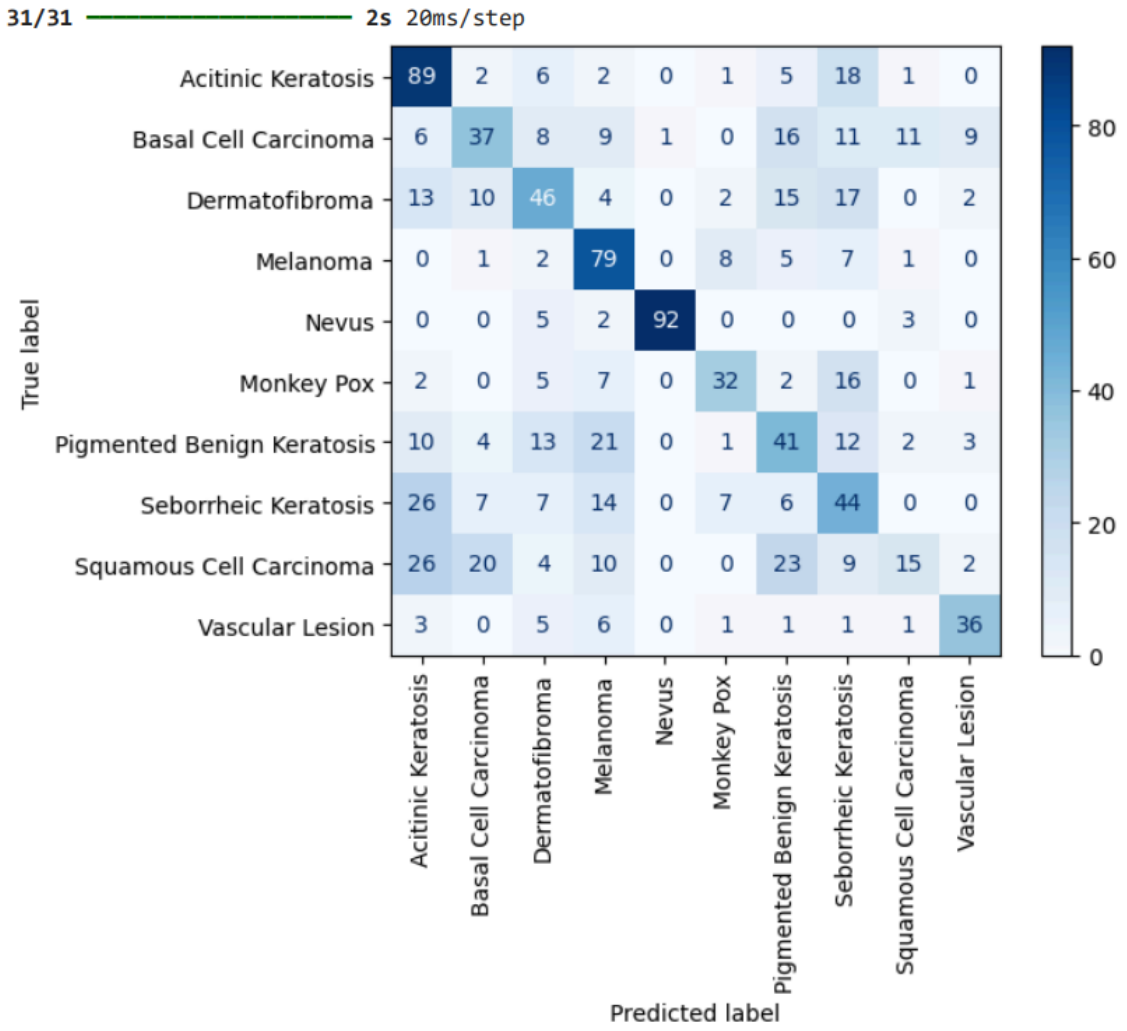

Figure 4.17: VGG19 Model (Freeze Layer) Confusion Matrix

With 92 accurate predictions, the Nevus classification is strong. However, the model shows high rates of misclassification for Squamous Cell Carcinoma, Seborrheic Keratosis, and Dermatofibroma. With an F1-score of 0.94, Nevus is the most successful, followed by Squamous Cell Carcinoma and Seborrheic Keratosis, which have F1-scores of 0.21 and 0.36, respectively. The macro-average F1-score of the model is 0.52 and its overall accuracy is 52%, showing a moderate performance across various categories as shown in Figures 4.17 and 4.18.

	precision	recall	f1-score	support
Acitinic Keratosis	0.51	0.72	0.60	124
Basal Cell Carcinoma	0.46	0.34	0.39	108
Dermatofibroma	0.46	0.42	0.44	109
Melanoma	0.51	0.77	0.61	103
Nevus	0.99	0.90	0.94	102
Monkey Pox	0.62	0.49	0.55	65
Pigmented Benign Keratosis	0.36	0.38	0.37	107
Seborrheic Keratosis	0.33	0.40	0.36	111
Squamous Cell Carcinoma	0.44	0.14	0.21	109
Vascular Lesion	0.68	0.67	0.67	54
accuracy			0.52	992
macro avg	0.53	0.52	0.51	992
weighted avg	0.52	0.52	0.50	992

Figure 4.18: VGG19 Model (Freeze Layer) Classification Metrics

Discussion: The model performs well in important classes like Nevus and Vascular Lesion and has high generalisation with minimal overfitting, as seen by the small difference between training and validation loss curves. But because of their visual similarity, it has trouble identifying Squamous Cell Carcinoma and Seborrheic Keratosis accurately, which can lead to misreading with Actinic Keratosis. The total accuracy of 52

4.7 InceptionV3 Model

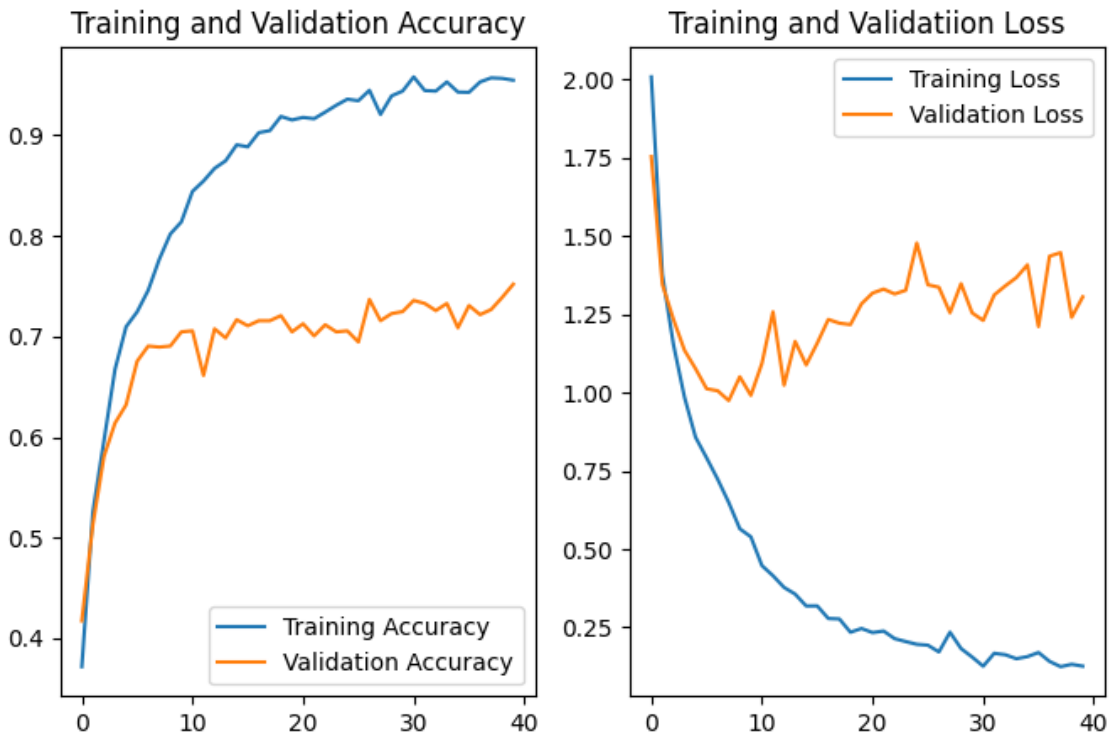

Figure 4.19: InceptionV3 Model Training and Validation Accuracy & Loss

Figures 4.19, 4.20, and 4.21 The model shows not much overfitting and good generalisation with a training accuracy of 95% and a validation accuracy maintained at 75%. Validation loss settles around 1.25, indicating a small gap, while training loss falls smoothly.

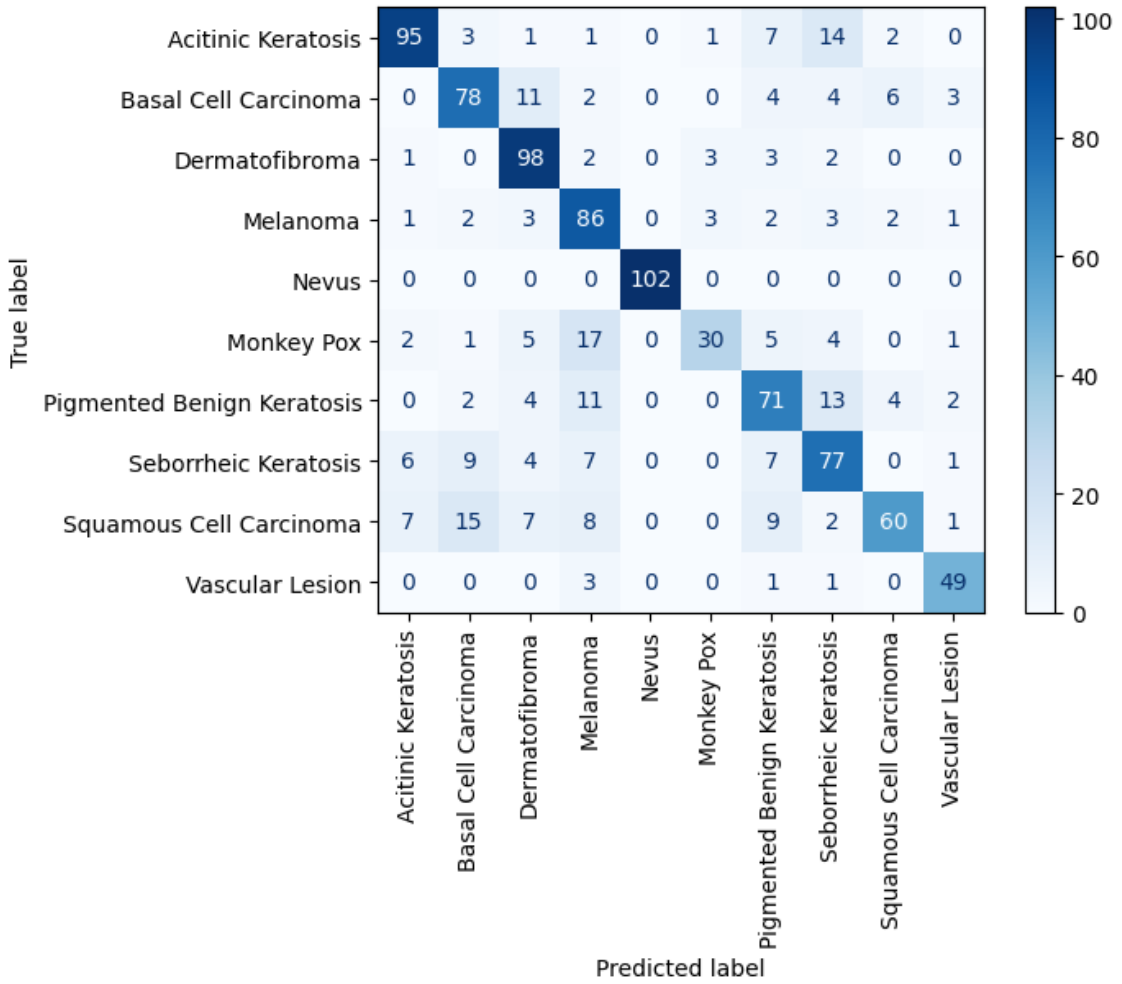

Figure 4.20: InceptionV3 Confusion Matrix

While nevus, vascular lesions, and dermatofibroma are all accurately classified, squamous cell carcinoma is highly misclassified, particularly when combined with basal cell carcinoma. For nevus, the model achieves an F1-score of 1.00, but it has problems with melanoma (0.67) and squamous cell carcinoma (0.66). F1 scores and overall accuracy reach 75%.

	precision	recall	f1-score	support
Acitinic Keratosis	0.85	0.77	0.81	124
Basal Cell Carcinoma	0.71	0.72	0.72	108
Dermatofibroma	0.74	0.90	0.81	109
Melanoma	0.63	0.83	0.72	103
Nevus	1.00	1.00	1.00	102
Monkey Pox	0.81	0.46	0.59	65
Pigmented Benign Keratosis	0.65	0.66	0.66	107
Seborrheic Keratosis	0.64	0.69	0.67	111
Squamous Cell Carcinoma	0.81	0.55	0.66	109
Vascular Lesion	0.84	0.91	0.88	54
accuracy			0.75	992
macro avg	0.77	0.75	0.75	992
weighted avg	0.76	0.75	0.75	992

Figure 4.21: InceptionV3 Classification Metrics

Discussion: A low difference in accuracy between training and validation data indicates the model's strong performance, which minimises overfitting, with 75% accuracy and good generalisation. With excellent precision and F1 scores, it performs well in classes such as Nevus, Vascular Lesion, and Dermatofibroma. Performance is impacted by lower F1 scores for melanoma, monkeypox, and squamous cell carcinoma, as well as confusion between these two types of cancer. Small variations in validation loss show areas for improvement. Class-weighted loss functions, improving InceptionV3 layers, and modifying hyperparameters to stabilise validation loss and improve performance are some possible improvements.

4.8 InceptionV3 Model (Freeze Layers)

When it comes to Nevus, the model does a good job of classifying it. It struggles, however, with Squamous Cell Carcinoma, frequently mistaken for Actinic Keratosis or Basal Cell Carcinoma.

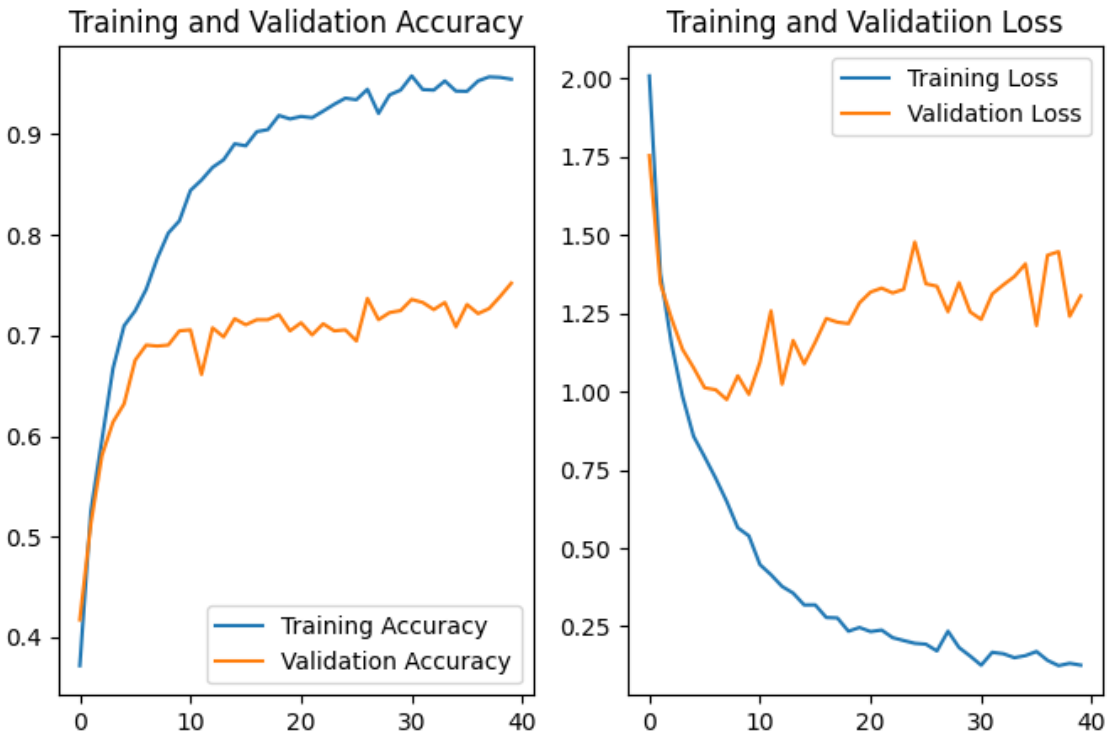

Figure 4.22: InceptionV3 Model (Freeze Layer) Training and Validation Accuracy & Loss

Figure 4.22 shows that the training accuracy reaches 90%, but an important difference with validation accuracy of 65% suggests that overfitting may have occurred. This is confirmed by changing validation loss. Nevus performs well for Pigmented Benign Keratosis (0.56), Seborrheic Keratosis (0.46), and Monkeypox (0.48), but weakly for F1-scores above 0.97. Figures 4.23 and 4.24 macro-average F1-score of 0.64 and an accuracy of 65% overall, the model performs well but in line across the different classes.

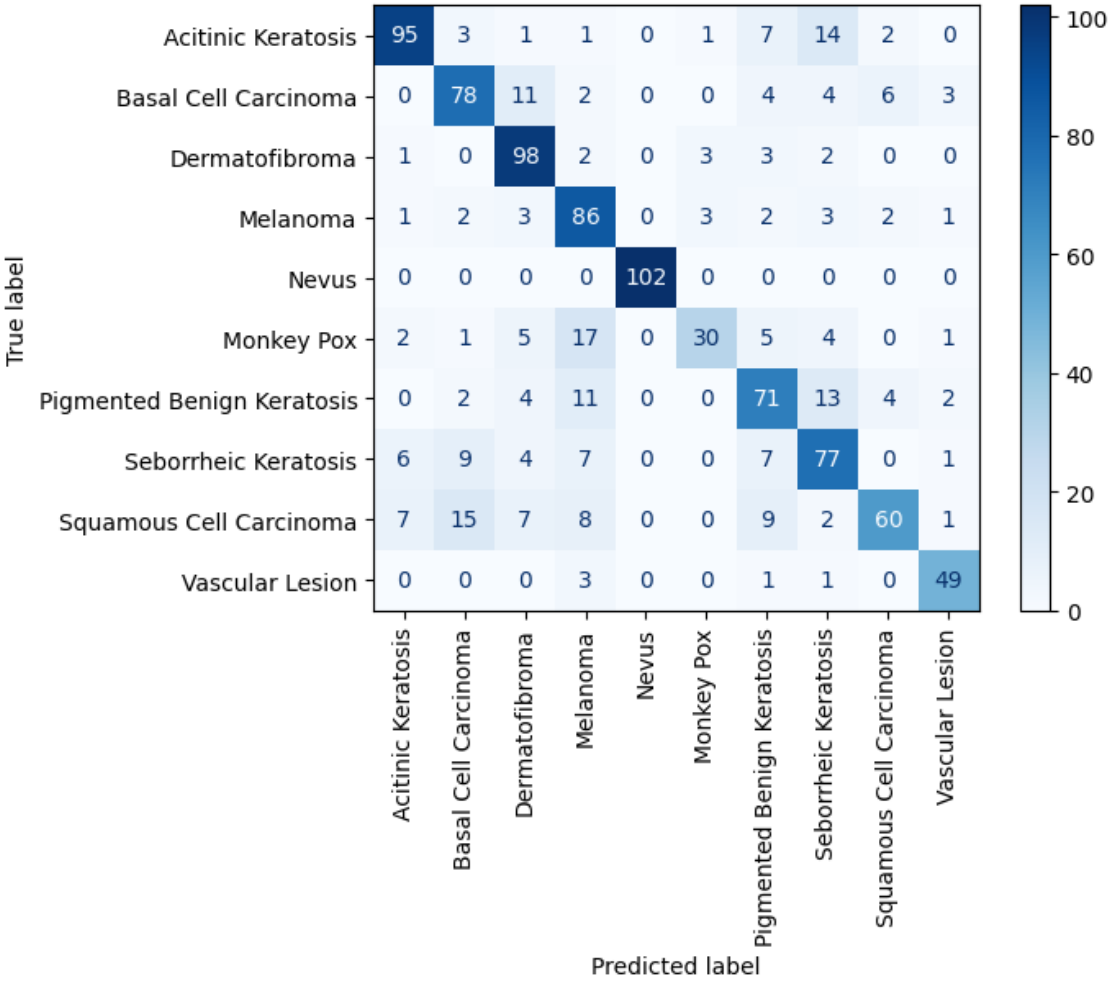

Figure 4.23: InceptionV3 Model (Freeze Layer) Confusion Matrix

	precision	recall	f1-score	support
Acitinic Keratosis	0.85	0.77	0.81	124
Basal Cell Carcinoma	0.71	0.72	0.72	108
Dermatofibroma	0.74	0.90	0.81	109
Melanoma	0.63	0.83	0.72	103
Nevus	1.00	1.00	1.00	102
Monkey Pox	0.81	0.46	0.59	65
Pigmented Benign Keratosis	0.65	0.66	0.66	107
Seborrheic Keratosis	0.64	0.69	0.67	111
Squamous Cell Carcinoma	0.81	0.55	0.66	109
Vascular Lesion	0.84	0.91	0.88	54
accuracy			0.75	992
macro avg	0.77	0.75	0.75	992
weighted avg	0.76	0.75	0.75	992

Figure 4.24: InceptionV3 Model (Freeze Layer)Classification Metrics

Discussion: Actinic Keratosis, Nevus, and Vascular Lesion show great precision, and the current model works well for important classes such as Nevus, Dermatofibroma, and Vascular Lesion. It performs poorly in classes like Seborrheic Keratosis and Monkeypox and struggles with class confusion, particularly between Actinic Keratosis, Basal Cell Carcinoma, and Squamous Cell Carcinoma. Possible overfitting is indicated by an important difference in accuracy between training and validation data. To decrease these problems, techniques such as class-weighted loss functions, advanced data augmentation, deeper InceptionV3 layer processing, SMOTE-addressed class imbalance, and ensemble learning methods may improve performance and reduce misclassification of visually similar cancer types.

4.9 Comparsion Of Models

In the comparison Table 4.1, Base Model 2 (Residual) performs better than Base Model 1 (CNN), showing improved F1-score for Squamous Cell Carcinoma (0.32 vs. 0.27) as well as higher training and validation accuracy (56.8% vs. 45.6% and 59.5% vs 53.5% respectively). Freezing the layers in the EfficientNetB0 models increases validation accuracy (78% vs.73%) but causes overfitting when training accuracy approaches 100%. While Nevus is well handled by both versions (0.99 F1-score), Squamous Cell Carcinoma is better handled by the frozen version (0.66 F1-score). Freezing layers for VGG19 reduces performance; validation accuracy falls from 60% to 50%, and the F1-score for Squamous Cell Carcinoma is worse (0.21 vs. 0.40). InceptionV3 also outperforms its frozen earlier version, obtaining better classification for challenging classifications such as Squamous Cell Carcinoma (0.66 F1-score) and higher validation accuracy (75% vs. 65%). InceptionV3 (non-freeze) finds the perfect balance between accuracy (75%) and generalisation, while EfficientNetB0 (Freeze) suffers from overfitting and gets the highest validation accuracy (78%) overall. The worst-performing model is VGG19 (Freeze), which has a poor classification for challenging classes and low accuracy. In conclusion, EfficientNetB0 (Freeze) needs changes to avoid overfitting despite its excellent accuracy, while InceptionV3 (non-freeze) delivers the best overall performance.

Table 4.1: Model Comparison Summary

Model	**Training Accuracy**	**Validation Accuracy**	**F1-Score (Best Class)**	**F1-Score (Worst Class)**
Base Model 1 (CNN)	**45.6%**	**53.5%**	**Nevus (0.95)**	Squamous Cell (0.27)
Base Model 2 (Residual)	56.8%	59.5%	Nevus (0.99)	Squamous Cell (0.32)
EfficientNetB0	96.5%	73%	Nevus (0.99)	Squamous Cell (0.59)
EfficientNetB0 (Freeze)	100%	78%	Nevus (0.99)	Squamous Cell (0.66)
VGG19	70%	60%	Nevus (0.99)	Seborrheic Keratosis (0.40)
VGG19 (Freeze)	50%	50%	Nevus (0.94)	Squamous Cell (0.21)
InceptionV3	95%	75%	Nevus (1.00)	Squamous Cell (0.66)
InceptionV3 (Freeze)	90%	65%	Nevus (0.97)	Monkeypox (0.48)

Chapter 5

Conclusion

This study's conclusion highlights the differences between the practical use of experimental CNN-based skin cancer detection models in clinical settings and their weaknesses, especially regarding overfitting, class confusion, and generalisation for new data. Findings show that models with high accuracy and F1 scores, like InceptionV3 and EfficientNetB0 (Freeze), are effective for Nevus and Dermatofibroma classes. However, problems still exist, such as overfitting, especially with EfficientNetB0 (Freeze), and incorrect classification of cancers that look identical in appearance (such as Squamous Cell Carcinoma and Basal Cell Carcinoma). Future research should concentrate on implementing complex data preprocessing techniques, like DullRazor algorithms (for hair removal) and histogram equalisation, addressing class imbalance using techniques like class-weighted loss functions or SMOTE, and improving generalisation with robust data augmentation techniques to address these challenges. In addition, more studies and additional clinical validation on various datasets are necessary to close the gap between model development and practical use.

Bibliography

Ali, M. S., Miah, M. S., Haque, J., Rahman, M. M. & Islam, M. K. (2021), 'An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models', *Machine Learning with Applications* **5**, 100036.

Baig, R., Bibi, M., Hamid, A., Kausar, S. & Khalid, S. (2020), 'Deep learning approaches towards skin lesion segmentation and classification from dermoscopic images-a review', *Current medical imaging* **16**(5), 513–533.

Chaturvedi, S. S., Tembhurne, J. V. & Diwan, T. (2020), 'A multi-class skin cancer classification using deep convolutional neural networks', *Multimedia Tools and Applications* **79**(39), 28477–28498.

Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O. & Mahnashi, M. H. (2021), 'Skin cancer detection: a review using deep learning techniques', *International journal of environmental research and public health* **18**(10), 5479.

Gezimati, M. & Singh, G. (2023), 'Advances in terahertz technology for cancer detection applications', *Optical and Quantum Electronics* **55**(2), 151.

Haggenmüller, S., Maron, R. C., Hekler, A., Utikal, J. S., Barata, C., Barnhill, R. L., Beltraminelli, H., Berking, C., Betz-Stablein, B., Blum, A. et al. (2021), 'Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts', *European Journal of Cancer* **156**, 202–216.

Hosny, K. M., Kassem, M. A. & Foaud, M. M. (2018), Skin cancer classification using deep learning and transfer learning, *in* '2018 9th Cairo international biomedical engineering conference (CIBEC)', IEEE, pp. 90–93.

Imran, T., Alghamdi, A. S. & Alkatheiri, M. S. (2024), 'Enhanced skin cancer classification using deep learning and nature-based feature optimization', *Engineering, Technology & Applied Science Research* **14**(1), 12702–12710.

Janda, M., Cust, A. E., Neale, R. E., Aitken, J. F., Baade, P. D., Green, A. C., Khosrotehrani, K., Mar, V., Soyer, H. P. & Whiteman, D. C. (2020), 'Early detection of melanoma: a consensus report from the australian skin and skin cancer research centre melanoma screening summit', *Australian and New Zealand journal of public health* **44**(2), 111–115.

Kausar, N., Hameed, A., Sattar, M., Ashraf, R., Imran, A. S., Abidin, M. Z. u. & Ali, A. (2021), 'Multiclass skin cancer classification using ensemble of fine-tuned deep learning models', *Applied Sciences* **11**(22), 10593.

MacLellan, A. N., Price, E. L., Publicover-Brouwer, P., Matheson, K., Ly, T. Y., Pasternak, S., Walsh, N. M., Gallant, C. J., Oakley, A., Hull, P. R. et al. (2021), 'The use of noninvasive imaging techniques in the diagnosis of melanoma: a prospective diagnostic accuracy study', *Journal of the American Academy of Dermatology* **85**(2), 353–359.

Mahbod, A., Tschandl, P., Langs, G., Ecker, R. & Ellinger, I. (2020), 'The effects of skin lesion segmentation on the performance of dermatoscopic image classification', *Computer Methods and Programs in Biomedicine* **197**, 105725.

Rezaoana, N., Hossain, M. S. & Andersson, K. (2020), Detection and classification of skin cancer by using a parallel cnn model, *in* '2020 IEEE international women in engineering (WIE) conference on electrical and computer engineering (WIECONECE)', IEEE, pp. 380–386.

Saleh, N., Hassan, M. A. & Salaheldin, A. M. (2024), 'Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making', *Scientific Reports* **14**(1), 17323.

Samushonga, W., Kumar, R., Gupta, M. & Obaid, A. J. (2024), A framework for skin cancer prediction using cnn ensembles, *in* '2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)', IEEE, pp. 1–6.

Shete, A. S., Rane, A. S., Gaikwad, P. S. & Patil, M. H. (2021), 'Detection of skin cancer using cnn algorithm', *International Journal* **6**(5).

Strzelecki, M., Kociołek, M., Strakowska, M., Kozłowski, M., Grzybowski, A. & Szczypiński, P. M. (2024), 'Artificial intelligence in the detection of skin cancer: state of the art', *Clinics in Dermatology* .

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**(1), 1–9.

Tuncer, T., Barua, P. D., Tuncer, I., Dogan, S. & Acharya, U. R. (2024), 'A lightweight deep convolutional neural network model for skin cancer image classification', *Applied Soft Computing* p. 111794.

Wen, D., Khan, S. M., Xu, A. J., Ibrahim, H., Smith, L., Caballero, J., Zepeda, L., de Blas Perez, C., Denniston, A. K., Liu, X. et al. (2022), 'Characteristics of publicly available skin cancer image datasets: a systematic review', *The Lancet Digital Health* **4**(1), e64–e74.

Zafar, K., Gilani, S. O., Waris, A., Ahmed, A., Jamil, M., Khan, M. N. & Sohail Kashif, A. (2020), 'Skin lesion segmentation from dermoscopic images using convolutional neural network', *Sensors* **20**(6), 1601.