BIRMINGHAM CITY University

A Deep Learning Approach to Detecting Skin Lesion

Supervisor: Dr. Besher Alhalabi

By: Samiksha Ingole | 22188113

Dissertation

18th September 2023

Abstract

Skin lesion detection is an important task in dermatology because it has the capacity to help medical practitioners diagnose different skin disorders, such as melanoma and other kinds of skin cancer. In this study, we explored the use of deep learning methods in skin lesion detection. A convolutional neural network (CNN) baseline model was used in our study, along with five prominent pretrained models: VGG16, MobileNet, InceptionV3, ResNet, and DenseNet. This study used the HAM10000 dataset, which contains 10,015 images of skin lesions. Data augmentation techniques are employed to diversify and increase the robustness of the training dataset. The MobileNet model outperforms not only the other pretrained models but also the CNN baseline model. This demonstrates MobileNet's outstanding ability to identify skin lesions. The MobileNet model with a fine-tuning approach achieves an accuracy of 83.23%, while the optimised model further improves the accuracy to 87%. Furthermore, we provided a user-friendly web-based application for skin lesion detection, with MobileNet serving as the backend model. This practical application makes skin lesion detection techniques more accessible to both healthcare professionals and everyone else, resulting in earlier diagnosis and better patient outcomes.

Acknowledgement

I would like to convey my deep gratitude and appreciation to all those who helped contribute to the successful completion of this dissertation.

First and foremost, I would want to express my sincere gratitude to my supervisor, Dr. Besher Alhalabi, for ongoing support, insightful advice, and unwavering guidance during this research project. Your expertise, patience, and commitment to my academic growth have been instrumental in shaping this dissertation. I am grateful to Birmingham City University's teachers and staff, whose commitment to academic achievement presented me with a dynamic learning environment.

Last but not least, I am extremely thankful to all my friends and my family: to my father, whose unwavering belief in my abilities has been a constant source of motivation; to my mother, whose support, encouragement, and understanding made this journey possible; to my sister, for her continuous encouragement and support throughout this research project.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Background and Rationale

Skin cancer is a significant health concern across the globe, with its prevalence increasing rapidly over the last several decades. Over the last ten years, the number of people diagnosed with melanoma skin cancer has increased by about 50%, making it the sixth most common cancer in the United Kingdom (Cancer Research UK, 2023). Skin cancer is classified into three main categories: basal cell skin cancer, squamous cell skin cancer, and melanoma (Waweru et al., 2020). Basal cell and squamous cell skin cancers are collectively referred to as nonmelanoma skin cancers. More than 90% of skin cancer cases are caused by UV light exposure (Narayanamurthy et al., 2018). Skin lesions, characterised by the abnormal growth of skin cells, can indeed lead to skin cancer and potentially harmful effects on an individual's health. The early diagnosis of skin lesions is critical for improving patient outcomes and reducing death rates. In the traditional method of skin cancer detection, a dermatologist visually examines the skin to identify any problematic signs. The dermatologist uses a systematic method of analysing the size, shape, colour, and texture of the lesions. They also analyse other features, such as irregular borders, asymmetry, and variations in elevation (Zhang et al., 2020). Due to the wide variety of features, detecting and differentiating melanoma from non-melanoma lesions is a significant challenge. A biopsy may be conducted if the dermatologist discovers a possibly malignant lesion. The traditional approach to detecting skin cancer mainly depends on the dermatologist's expertise and experience (Kassem et al., 2021). Their visual evaluation, along with a dermatoscope and, in certain cases, a biopsy, is critical to accurately detecting skin cancer. However, standard inspection by dermatologists for skin lesion identification is time-consuming, subjective, and sometimes inaccessible in rural areas, resulting in a delayed or insufficient diagnosis. To overcome these difficulties, researchers are investigating the potential of developing automated systems for skin cancer diagnosis utilising modern technologies such as artificial intelligence and deep learning algorithms. In recent years, deep learning has excelled in several kinds of computer vision applications. Its ability to understand complicated patterns and features from massive data sets makes it a potential technique for skin lesion diagnosis. Leveraging deep learning techniques, particularly convolutional neural networks (CNNs), researchers and medical professionals are investigating the development of automated systems for skin lesion detection. Transfer learning also shows remarkable potential for advancing the field of skin lesion detection. Transfer learning, a technique in which pre-trained models are adapted to specific tasks, plays a pivotal role in enhancing the accuracy and efficiency of skin lesion detection. By training neural networks on large datasets of skin lesion images, these systems can learn to identify trends, features, and patterns that may not be immediately apparent to the human eye. This could change the way we find and identify skin lesions, leading to earlier discovery, better accuracy, and better patient results in the fight against skin cancer. The purpose of this project is to employ deep learning techniques, such as convolutional neural networks (CNNs) and some pretrained models, to create a robust system for classifying skin lesions into specific categories. This research intends to overcome the limitations of the traditional methodology by providing a faster, more objective, and perhaps more broadly accessible method for finding and diagnosing skin lesions, thereby improving patient outcomes in the context of skin cancer diagnosis. Automated skin lesion diagnosis has the potential to revolutionise how medical practitioners approach this crucial area of skin diseases, ultimately protecting lives and minimising the financial and social impact that skin cancer has on both patients and healthcare systems.

1.2 Research Questions

The following research questions must be addressed as part of this project:

- What does the existing literature reveal about the latest deep learning approaches for skin lesion detection, and how can this knowledge inform our project?
- How can a deep learning model be developed and optimised to correctly classify various kinds of skin lesions?
- What are the different techniques and approaches involved in leveraging the transfer learning of pretrained models?
- What difficulties and limitations come with using deep learning models for skin lesion detection?

1.3 Aim and Objectives

Aim

The aim of the project is to develop an accurate web-based application for diagnosing various forms of skin lesions using deep learning models.

Objectives

- Conduct an extensive literature review to identify and analyse state-of-the-art deep learning approaches for skin lesion detection.
- Develop and optimise a robust deep learning model for accurate categorization of skin lesions using a CNN architecture, leveraging transfer learning and data augmentation approaches to improve performance.
- Provide a comprehensive evaluation and discussion of the customised CNN and transfer learning models.

1.4 Overview

This study is an in-depth examination of the potential of deep learning approaches in skin lesion detection. The project is divided into chapters, each of which contributes to a thorough understanding and implementation of our deep learning-based approach:

Chapter 2: **Literature Review**- presents a thorough analysis of the available literature, highlighting the crucial contribution of deep learning to improving skin lesion identification.

Chapter 3: **Methodology**- examines the use of our agile development technique throughout the project's lifespan. Also, provides a foundational understanding of deep learning concepts, methods, and techniques that relate to our field of study.

Chapter 4: **Dataset-** covers dataset gathering, exploratory data analysis (EDA), and image data processing methods, which all work together to set the foundation for effective deep learning model development.

Chapter 5: Experiment setup and results- describes the experiment design, model specifications, and performance assessment for models implemented.

Chapter 6: Model evaluation and optimisation- focuses on assessing our top-performing model and then optimising it to raise its ability to detect skin lesions accurately.

Chapter 7: Web Application Development- explores the process of creating a user-friendly Flask-based web application for detecting skin lesions.

Chapter 8: Discussion- engages in a comprehensive discussion of the results, implications, and significance of the skin lesion detection project.

Chapter 9: Conclusion and Future work- summarises significant results, their implications for skin lesion detection, and possible directions for further research and development.

2. Literature Review

Skin lesions are a major public health concern, and their detection using deep learning and machine learning approaches has received a lot of interest recently. Deep learning, specifically convolutional neural networks (CNNs), has demonstrated exceptional efficiency in extracting discriminative features from skin lesion images and accurately classifying them into various types of skin disorders (Bechelli & Delhommelle, 2022; Khan et al., 2021; Waweru et al., 2020). On the other hand, machine learning algorithms such as support vector machine classifiers, k-nearest neighbors classifiers, and decision tree classifiers have been widely used for skin cancer classification (Bechelli & Delhommelle, 2022; Vidya & Karki, 2020). Previous studies have shown that CNNs outperform standard machine learning algorithms in terms of performance (Bechelli & Delhommelle, 2022; Mazhar et al., 2023). The purpose of this literature review is to provide an overview of existing research and improvements in the use of deep learning methods for skin cancer detection.

2.1 Skin Lesion Detection and Traditional Approaches

Skin lesions, which can range from harmless moles to potentially fatal cancers, present a serious hazard to public health. The importance of early detection cannot be emphasised, as it influences treatment options and prognosis directly. The diagnosis of skin lesions has traditionally been based on the skill of dermatologists and doctors who carefully evaluate visual indications of lesion such as the ABCD criteria (asymmetry, border irregularity, colour change, and diameter) (Ayan & Ünver, 2018). These ABCD criteria give dermatologists a structured framework for making diagnostic choices. In addition, dermatologists consider additional factors in their diagnosis procedure, such as patient history, medical records, and family history. Another technique of identifying lesion is a seven-point check list. It is made up of seven essential criteria, or aspects, that are assessed when a skin lesion is visually inspected.

Even though these conventional methods have been helpful in early skin lesion detection, they are not without limitations. Due to the subjectivity involved in visual inspection, different dermatologists may reach different findings about the same lesion. Traditional treatments may also struggle with the complexities of complicated or unusual lesions. When lesions do not meet the traditional ABCD criteria, diagnosing them can be difficult and may result in delayed or inaccurate results. In reaction to these problems, the field of dermatology has evolved a lot with the help of new technologies and statistical methods. More accurate and data-driven skin lesion identification approaches are now possible because of cutting-edge technologies like machine learning, deep learning, big data, and artificial intelligence.

2.2 Deep Learning in Medical Imaging

Medical image analysis is one of the most promising applications of deep learning in healthcare. In recent times, the use of deep learning methods in medical imaging has revolutionised the field, providing unprecedented abilities in disease detection, diagnosis, and prediction. Deep learning algorithms have proven remarkable abilities in learning and extracting complex properties from images automatically (Lee et al., 2017). Deep learning, notably convolutional neural networks (CNNs), is a powerful method for extracting complicated patterns and features from medical images such as X-rays, MRIs, CT scans, retinal photography, dermoscopy images, and microscopy images (Ker et al., 2017). Their capacity to precisely identify anomalies in medical imaging has proven helpful for the early diagnosis of diseases including cancer, cardiovascular diseases, and neurological problems. The application of

CNN-based tumour detection from brain scan images achieved a high accuracy of 98.67% using the Softmax classifier, beating other common ML classifiers (Basheera & Ram, 2019). In a study conducted by Sakib et al. (2020), a proposed customised CNN-based chest radiograph classification system with data augmentation achieved a COVID-19 detection accuracy of 93.94%, outperforming the scenario without data augmentation (54.55%). In the study (Liu et al., 2014), deep learning was used to create a novel method for the early detection of Alzheimer's disease and mild cognitive impairment, giving improved performance in multi-class classification. The study used deep learning networks and convolutional techniques to analyse breast cancer images, achieving 99% accuracy using GoogleNet and exploring statistical feature-based classifiers for improved recognition (Rajakumari & Kalaivani, 2022). Deep learning algorithms have played a critical role in improving the quality of medical images. They are skilled in noise reduction and super-resolution, resulting in clearer and more diagnostically valuable photographs. For instance, deep learning-based denoising has improved picture quality in brain MRI (Kidoh et al., 2020).

2.3 Skin Lesion Classification Using Deep Learning Approach

In recent years, there has been a growing interest in using deep learning models to detect and classify skin lesions. Several studies have been carried out to investigate the efficiency of deep learning models in appropriately diagnosing various forms of skin lesions. In the study (Fu'adah et al., 2020), a CNN model with three hidden layers using 3 × 3 filter sizes and sequential outputs of 16, 32, and 64 channels was used to classify skin cancer lesions and benign tumour lesions. The CNN model with the Adam optimizer achieved the highest accuracy of 99% in classification. The researcher (Harangi, 2018) advanced the field of skin lesion detection by highlighting the benefits of merging several CNNs into a unified framework for better results in image classification tasks. The generation of an effective lesion depiction is critical for the appropriate categorization of lesions.

While most research focuses on binary classification tasks such as benign and malignant classification, some studies have broadened their focus to multi-class classification based on deep learning, with the goal of classifying skin lesions into different subtypes, such as melanoma, dermatofibroma, nevi, basal cell carcinoma, and more (Fu'adah et al., 2020; Iqbal et al., 2021; Khan et al., 2019; Mishra et al., 2021). This broadening of class categories aids in the early identification and precise classification of a wider variety of skin conditions. The researchers (Zhang et al., 2020) proposed a novel optimised approach for detecting skin cancer from input images. The paper (Khan et al., 2021) describes a novel framework for skin lesion diagnosis on mobile devices based on convolutional neural networks (CNNs). It enhances segmentation accuracy while decreasing computing time by using a 16-layered CNN for lesion segmentation and feature extraction. The whale optimisation technique has been improved and used for the optimal selection of weights and biases in the CNN to improve performance. The study showed (Mishra et al., 2021) that deep convolutional neural networks have superior accuracy and performance, especially when trained on dermoscopic images, with the best results coming from dropout regularisation and the stochastic gradient descent (SGD) optimizer. This highlights the significance of optimising hyperparameters, especially learning rates, in order to produce accurate and high-performing models in the medical field. The study provided a skin lesion segmentation algorithm that utilises CNN in combination with an advanced hair-removal technique, efficiently removing hair structures from dermoscopic images (Zafar et al., 2020). The paper (Shetty et al., 2022) demonstrates that a Convolutional Neural Network (CNN) architecture was developed with a focus on using hyperparameters such as the Adam optimizer, categorical cross-entropy loss function, 150 epochs, a batch size of 32, and an initial learning rate of 0.001. These hyperparameters were selected carefully based on their broad efficacy in training deep neural networks, and they were adjusted through experimentation to produce a well-performing model with minimal loss and less overfitting to the training

data. The study proposed a customised Convolutional Neural Network (CNN) architecture with a six-layered model and batch normalisation for the categorization of dermoscopic images into benign or malignant skin lesions (Jayalakshmi & Kumar, 2019). The suggested batch normalised CNN (BN-CNN) outperformed the custom CNN with an amazing accuracy rate of 89.30%, highlighting the critical importance of batch normalisation in enhancing classification results.

2.4 Skin Lesion Dataset and Challenges

Skin lesion datasets play an important role in the development and evaluation of skin lesion detection systems. These datasets typically contain a wide variety of images representing different types of skin lesions. The HAM10000 is a benchmark dataset commonly used for skin lesion analysis and detection using deep learning (Bechelli & Delhommelle, 2022; Khan et al., 2019; Sae-Lim et al., 2019; Waweru et al., 2020). It includes 10,015 dermoscopic images of seven different types of skin lesions. The International Skin Imaging Collaboration (ISIC) 2017 dataset is another well-known and commonly used dataset in the field of skin lesion detection (Harangi, 2018; Mazhar et al., 2023; Murphree & Ngufor, 2017). The ISIC dataset includes a wide range of skin lesion images, including melanoma, nevi, and basal cell carcinoma. The Dermofit Image Library is another notable dataset that offers high-quality pictures of various skin disorders. It has been commonly used for benchmarking skin lesion classification models (Fan et al., 2019; Abhishek & Hamarneh, 2021). These datasets are an important source for developing and testing deep learning models, ultimately advancing skin lesion detection systems. However, medical datasets often pose several challenges. One major challenge is small datasets. Additionally, an imbalanced dataset is another challenge, leading to biased classification models (Al-Asadi & Altun, 2022; Sae-Lim et al., 2019). To address the issue of data imbalance, many approaches have been proposed, including over-sampling and under-sampling. These approaches, however, have downsides such as information loss and overfitting. To address these issues, researchers (Emara et al., 2019) implemented a weighted random sampling strategy, which helps balance the representation of different classes in the dataset during training. With the rising usage of digital health technology, patient data privacy and security are becoming another challenge. The collection and distribution of medical images, especially photos of skin lesions, raises issues about patient privacy (Pacheco & Krohling, 2019).

These challenges are critical in determining the direction of future improvements in the field of skin lesion detection. Addressing these issues is critical not only for enhancing the accuracy and reliability of classification models but also for ensuring their usability and efficacy in clinical contexts.

2.5 Integration of Pretrained Models

Transfer learning, an area of deep learning, has become well-known for its capacity to enhance the performance of skin lesion detection models by using pre-trained neural network architectures on huge data sets. Several studies employ pre-trained models such as VGG, ResNet, Inception, and DenseNet for feature extraction and classification tasks in skin lesion detection (Khan et al., 2019; Waweru et al., 2020). Pre-trained models such as GoogleNet and MobileNet were used to compare performance to customised CNN models (Junayed et al., 2021). The study proposed a deep learning approach for skin lesion classification by combining and fine-tuning three pre-trained deep learning architectures that include Xception, Inception-ResNet-V2, and NasNetLarge (Ahmed et al., 2020). Due to the limited number of images in the skin lesion dataset, the approach employs transfer learning by initialising the model with weights from the VGG16 pre-trained on a large dataset, such as ImageNet, to exploit

learned features (Romero et al., 2017). The research (Hosny et al., 2018) describes a transfer learning-based methodology using a pre-trained deep learning network, particularly AlexNet, which outperforms previous techniques with an outstanding classification rate. The study provided an approach for classifying skin lesions that uses MobileNet, a lightweight pretrained CNN model (Sae-Lim et al., 2019). Another study illustrates MobileNet's efficiency in classifying skin lesions, with accuracy rates of 81.52% without augmentation and an increase to 82% with augmentation (Salian et al., 2020). The work (Hassan et al., 2020) highlights the rising importance of transfer learning approaches, in particular DenseNet-121, in classifying skin lesions, using the 'HAM10000' dataset and data augmentation to improve classifier performance. These studies demonstrate the efficacy of several pre-trained architectures and approaches for reaching high classification accuracy for skin lesions.

3. Methodology

3.1 System Development Methodology

Agile is a project management methodology that prioritises flexibility, collaboration, and iterative delivery. Agile methods outperform traditional approaches like waterfall and spiral by providing enhanced adaptability and supporting direct collaboration between customers and developers (Al-Saqqa et al., 2020). Agile planning focuses on short-term, specific characteristics, streamlining the process as compared to traditional methods. There are several kinds of agile approaches. The Agile Scrum approach will be followed in this project. Agile Scrum is a collaborative and iterative method that promotes adaptation, flexibility, and continual progress (Day, 2020). A sprint is the smallest scrum block, usually lasting 1–3 weeks. Each sprint will have its own set of objectives, tasks, and outputs (Srivastava et al., 2017). Figure 1 shows the sprint cycle.

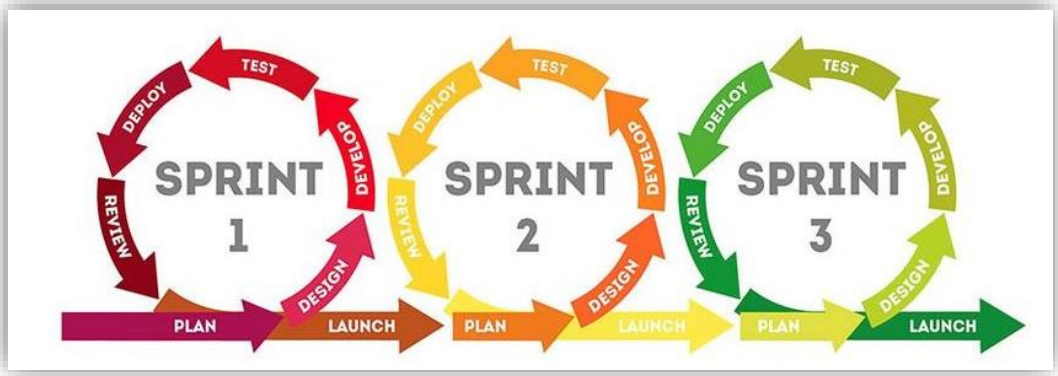


Figure 1. Sprint Cycle (Făgărășan et al., 2021)

A sprint review meeting will be held at the end of each sprint to present the finished work to supervisors and receive feedback. This feedback will be utilised to fine-tune and adapt the project's scope and priorities. There will also be a sprint retrospective meeting to reflect on the performance and recommend areas for improvement in future sprints. The following sprints are planned for this project:

Sprint 1: During the first sprint, we will concentrate on reviewing prior work on skin lesion detection using deep learning approaches. Additionally, we will collect the skin lesion dataset, which include images and corresponding labels. The HAM10000 dataset, a huge collection of multi-sources dermatoscopic images of common pigmented skin lesions released by Harvard Dataverse (Tschandl, 2018), will be used in this study.

Sprint 2: The second sprint will entail performing exploratory data analysis (EDA) on the skin lesion dataset. This involves using various Python libraries, such as Numpy, Pandas, Matplotlib, Seaborn, etc. The primary goal of this activity is to get insights into data, discover data imbalances, and determine preprocessing and data augmentation requirements.

Sprint 3: CNN design and implementation will take place during this sprint. The CNN model is designed to learn hierarchical patterns as well as characteristics from the raw input data. The major components of CNN model are convolutional layers, pooling layers, and fully connected layers (Zhang et al., 2020). The architecture and layers of CNN will be determined based on prior studies. A deep

learning framework, such as TensorFlow, will be used to implement the CNN model.

Sprint 4: This sprint will focus on training and evaluation of CNN model. The training procedure includes feeding the pre-processed images from HAM10000 dataset into the CNN model, optimising the model parameters using backpropagation, and updating the weights to minimise the loss function. After training, performance of CNN model will be evaluated using appropriate method.

Sprint 5: Several pretrained transfer learning models will be investigated during Sprint 5. Transfer learning is a type of supervised learning approach in which knowledge obtained from one task is used to improve performance on another similar task. Based on the literature study, a trained model such as VGG, ResNet, Inception, and DenseNet, among others, will be chosen. These models have already learned properties from large-scale datasets like ImageNet, which they might use to detect skin lesions. The pretrained models will be trained and evaluated on the HAM10000 dataset.

Sprint 6: This sprint will compare the CNN model built in Sprint 3 to the pretrained models implemented in Sprint 5. The purpose of this sprint is to determine the best performing model employing suitable evaluation metrics and validation methodologies. Following that, it will optimise the best-performing model and prepare it for integration into a web-based application. Various methodologies and approaches will be investigated in order to fine-tune and optimise the chosen model further, with the goal of improving its performance and efficiency.

Sprint 7: The objective of Sprint 7 is to develop a user-friendly web-based application that enables users to input images of skin lesions and obtain predictions for the presence of skin lesions. To create a web application, the flask framework will be utilised. This framework offers the infrastructure required for developing web applications and handling user interactions. Flask makes it easier to integrate the deep learning model into the web application.

By implementing an agile scrum approach, this project will aim to avoid failures and ensure flexibility during the project.

3.2 Theoretical Background

This section provides a comprehensive overview of the theoretical background and techniques employed in the design and development of this project.

3.2.1 Convolutional Neural Network

A convolutional neural network (CNN) is a deep learning method created primarily for image processing and recognition. CNNs are effective for image classification because they can automatically learn spatial hierarchies of features (Mohan, 2014; Zhao & Du, 2016). This hierarchical learning enables CNNs to recognise complicated patterns, edges, textures, and styles in images, making them very proficient in detecting objects and patterns within the images on which they have been trained (Mall et al., 2023). In the case of skin lesion detection, CNN's ability to distinguish patterns and textures is important. It is essential to distinguish between several skin disorders that have similar appearances. A basic convolutional neural network (CNN) architecture (Figure 2) consists of the following layers:

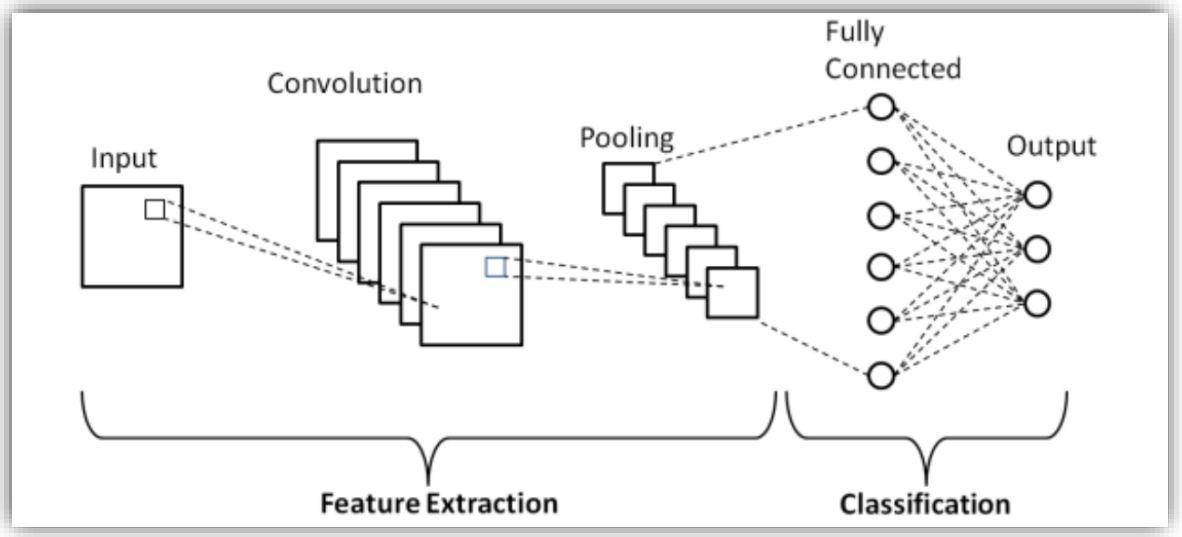


Figure 2. CNN Architecture (Phung & Rhee, 2019)

- **Input Layer**
  The input layer receives an image's pixel values as its input. The characteristics of the input images, such as their height, breadth, and number of colour channels, determine the size of the input layer in a CNN.

- **Convolutional Layers:**
  The convolutional layers serve as an important component of a CNN. The purpose of this layer is to extract various attributes from the input images. It consists of a set of filters, which are small windows that pass over the input image to conduct convolutional operations (Dertat, 2017). Each filter is designed to identify distinct characteristics, such as edges, patterns, corners, or textures, by using a dot product between the filter weights and the localised regions of the input picture (Ramprasath et al., 2018). The output of this layer is commonly referred to as the feature map (Phung & Rhee, 2019). This feature map is then passed on to subsequent layers, allowing them to learn additional characteristics from the input picture.

- **Pooling Layers:**
  In CNN, convolutional layers are followed by pooling layers. The primary aim of this layer is to minimise the dimensions of the convolved feature map, resulting in minimising the computational cost (Al-Saffar et al., 2017). Pooling layers can also help to achieve translation invariance in feature maps (Nahata & Singh, 2020). Translation invariance implies that the network can recognise the same features independent of their position or location in the input image. By setting up translation invariance, the CNN can handle small changes in where objects are placed in the image. This makes it better for real-world uses where objects are not always in the exact centre of the picture.

  There are various types of pooling methods. Maximum pooling and average pooling are the two widely used forms of pooling (Yani et al., 2019). The max pooling method returns the highest value from the region of the image covered by the kernel (Dertat, 2017). It aids in capturing the region's most essential feature. The average pooling provides an average of all the values in the region covered by the kernel in the input image (Hijazi et al., 2015). It provides

a way to summarise the knowledge in that area. Figure 3 illustrates the maximum pooling and average pooling operations.

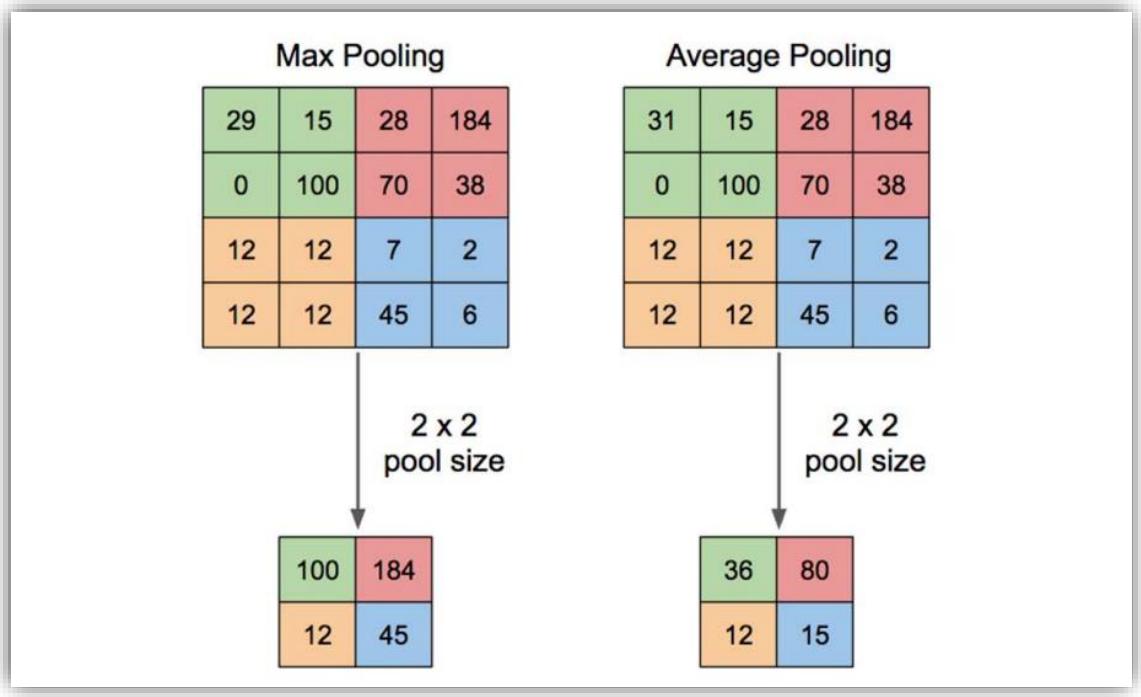


Figure 3. Max and Average Pooling Operation (Yani et al., 2019)

- **Fully Connected Layers:**

  The fully connected layer, also known as the dense layer, is the final layer of a CNN that conducts the classification task. It is made up of weights and biases as well as neurons and is used to link neurons from various layers. Before being sent to the fully connected layer, the output of the final convolutional and pooling layers is usually flattened into a one-dimensional vector (Dertat, 2017). This process changes the spatially organised features into a continuous series of values. Each neuron in the fully connected layer uses an activation function to modify the weighted sum of its inputs. The SoftMax function is the most often used activation function for classification problems, converting raw scores into class probabilities (Mall et al., 2023). The results of the CNN are determined by the outputs of the neurons in the fully connected layer. In classification, the neuron with the highest probability corresponds to the predicted label for the input image.

3.2.2 Transfer Learning models

In the domain of skin lesion detection, transfer learning is a useful strategy for using the capabilities of deep learning models that have been pre-trained on large datasets. The following transfer learning models were selected because of their proven success in image recognition applications, as evidenced by their utilisation in prior studies.

3.2.2.1 VGG 16

The VGG16 is a foundational model renowned for its uncomplicated structure and exceptional performance. The VGG16 model has a test accuracy of 92.7% in ImageNet, a dataset with over

14 million training pictures in over 1000 object classes (Lopez et al., 2021). VGG16, as the name implies, is a 16-layer deep neural network (Choi et al., 2021). It is consisting of thirteen convolutional layers, followed by three fully connected layers (Figure 4). The VGG16 model is remarkable for its use of small 3x3 convolutional filters. It expects images with a fixed dimension of 224x224 pixels (Romero et al., 2017). The rectified linear unit (ReLU) is used as the activation function in both convolutional layers and fully linked layers.

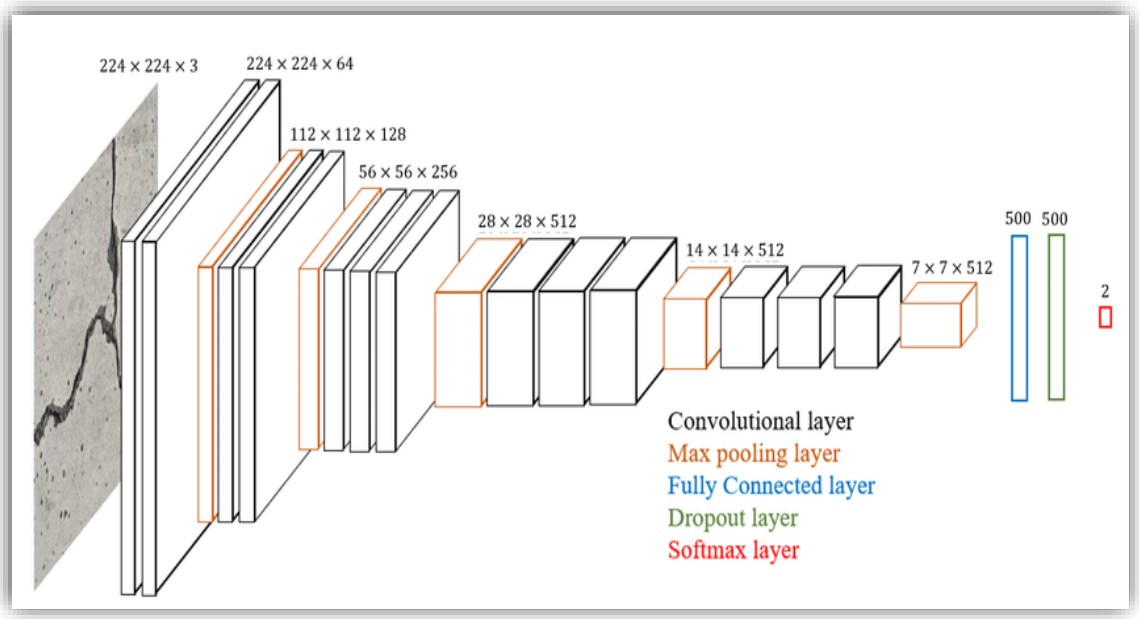


Figure 4. VGG16 Architecture (Choi et al., 2021)

3.2.2.2 MobileNet

MobileNet was created by Google researchers as part of their attempts to design efficient CNN architectures for mobile and embedded devices (Ansar et al., 2020). These models have a focus on computational efficiency while still retaining a high standard of accuracy for different computer vision tasks. MobileNet models employ depth-wise separable convolutions, a technique that divides conventional convolutions into two distinct layers, such as depth-wise convolution and point-wise convolution (Han et al., 2022) (Figure 5). Depth-wise Convolution applies a single convolutional filter per input channel. This layer records spatial data inside each channel. Pointwise convolution combines the outputs of the depth-wise convolution using a 1x1 convolution. This layer records cross-channel correlations. This methodology effectively decreases the quantity of parameters and calculations needed, enhancing the efficiency of the model.

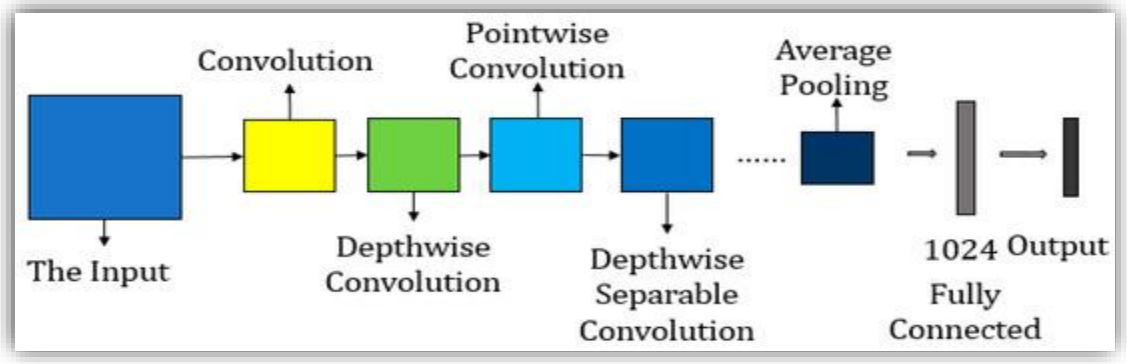


Figure 5. MobileNet Basic Architecture (Han et al., 2022)

3.2.2.3 DenseNet

DenseNet, also known as densely connected convolutional networks, was first introduced in 2017 by Gao Huang et al. in their article titled Densely Connected Convolutional Networks (Huang et al., 2017). It is a convolutional neural network (CNN) architecture with a feed-forward design that connects every layer to every other layer. This lowers the number of parameters and improves gradient flow during training by enabling the network to learn more effectively by reusing features. There are several variants of the Densenet architecture that vary in size and complexity to accommodate a variety of requirements and tasks. DenseNet-121 is a specific variant of the DenseNet architecture. It has 121 layers and is renowned for maintaining a balance between computational speed and model capacity (Radwan, 2019). DenseNet-121 employs four dense blocks (Figure 6), each of which consists of numerous convolutional layers with dense connectivity, bottleneck layers for efficiency, and transition layers for downsampling (Albelwi, 2022).

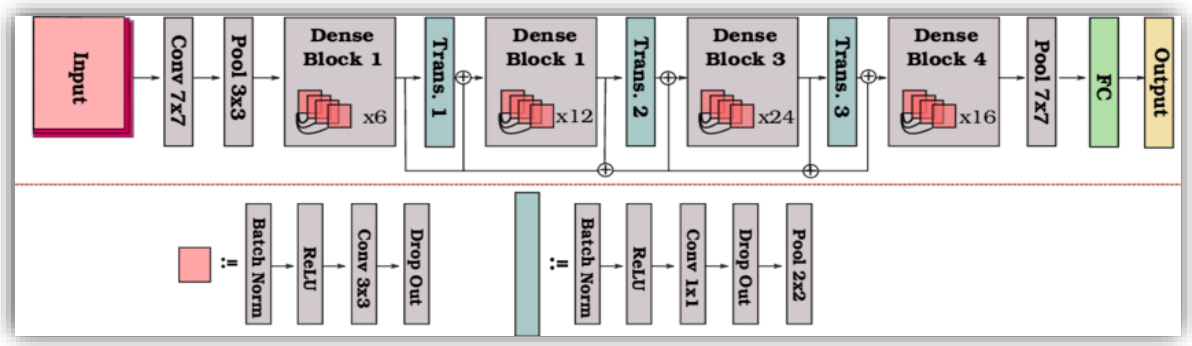


Figure 6. DenseNet121 Architecture (Radwan, 2019)

3.2.2.4 Inception

The Inception model, first developed by the Google Research team, is a CNN architecture for image recognition (Saber et al., 2021). The original Inception design has changed over time, giving rise to several variations that each expand upon the fundamental concepts while addressing specific flaws. This study uses the InceptionV3 variant for transfer learning. The InceptionV3 design is known for its innovative utilisation of inception modules. These modules

are the network's building blocks, and they are meant to collect characteristics at multiple scales by applying concurrent convolutional processes of varying sizes, such as 1x1, 3x3, and 5x5 convolutions, as well as pooling layers (Ali et al., 2021). The architecture is made up of a sequence of these inception modules, as well as auxiliary classifiers placed at strategic spots to aid with the vanishing gradient problem during training (Figure 7).

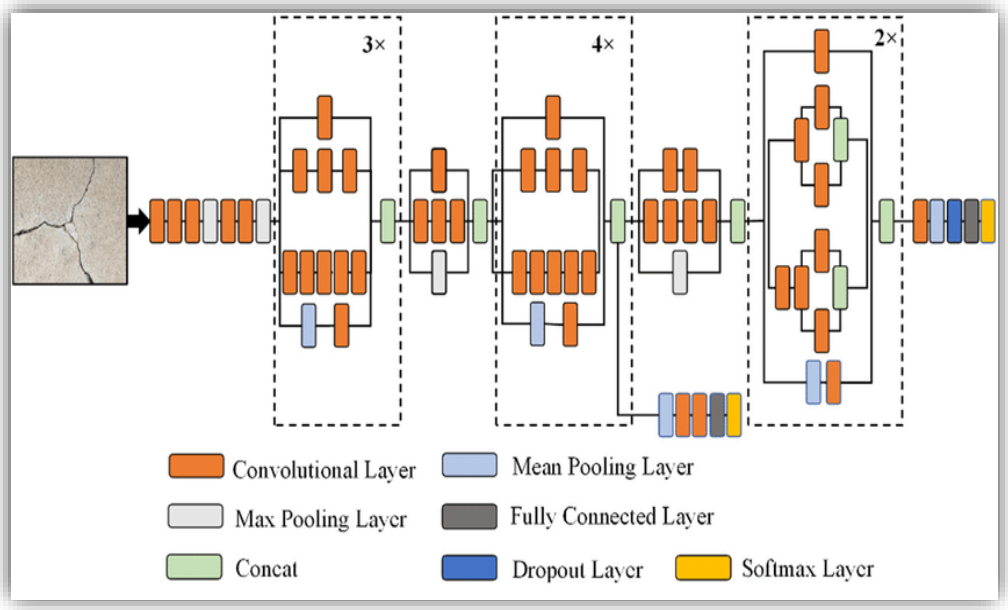


Figure 7. Inception V3 Architecture (Ali et al., 2021)

3.2.2.5 ResNet

The ResNet (Residual Network) architecture is well recognised in the field of deep learning. Microsoft Research introduced ResNet in 2015 to solve deeper network degradation problem (Wei et al., 2023). The degradation problem is the fact that as CNNs go deeper, their accuracy starts to saturate or even decline because of the vanishing gradient problem. ResNet designs are categorised according to their depth, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152. ResNet-50 is a 50-layer ResNet architecture that includes both convolutional and fully connected layers (Mukherjee, 2022). The architecture of ResNet-50 is made up of building components called residual blocks, each of which has several convolutional layers (Figure 8). The skip connections in these blocks aid in the propagation of gradients during training, allowing the network to learn complicated features and patterns. The inclusion of these skip connections enables ResNet-50 to attain extraordinary depth while being trainable and effective (Bhattacharjee et al., 2023).

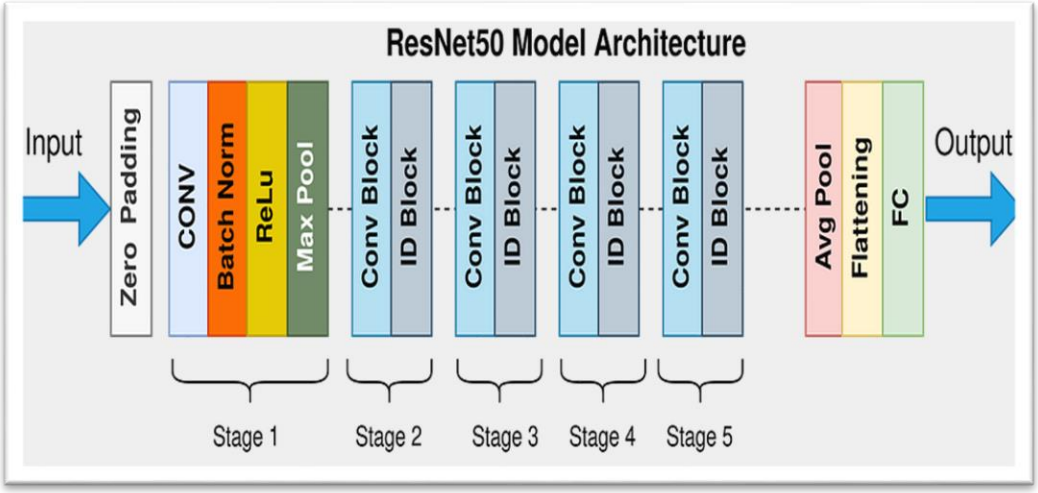


Figure 8. ResNet50 Architecture (Mukherjee, 2022)

3.2.3 Transfer Learning approach

There are several techniques for implementing transfer learning models. The following two approaches were employed in this project:

3.2.3.1 Feature Extraction

Feature extraction is a transfer learning approach that leverages complex feature extractors from pre-trained models. In this technique, high-level features are extracted from input images using the initial layers of a pre-trained model (Mueller, 2020). Then, these features are sent into a new set of layers that are generally made up of fully connected layers and are trained to identify the target classes.

The main advantage of feature extraction is its capacity to build on the generalizable information acquired by the initial layers of the pre-trained model. These features, which may include edges, textures, and basic shapes, are shared across a wide range of tasks related to image recognition. Feature extraction provides an effective approach to changing a pre-trained model for a new task with small datasets by retaining the pre-trained layers and training only the classifier layers.

3.2.3.2 Fine-Tuning

Fine-tuning is an extension of transfer learning that involves changing not just the classifier layers but also some of the previous layers of the pre-trained model. Instead of keeping these layers fixed, they are "unfrozen," and the whole model is trained on the new dataset. This procedure seeks to modify the feature representations of the pre-trained model to reflect the nuances of the target task (Alhadhrami et al., 2018).

Fine-tuning increases the adaptability of a model, making it suitable for scenarios with a larger data set. By allowing prior layers to acquire task-specific properties, the model may capture higher-level abstractions linked to the new task. However, fine-tuning needs thoughtful consideration of learning rates, regularization, and the potential for overfitting, as the model's original weights are modified to fit the new data distribution.

4. Dataset

The dataset used in the skin lesion detection study will be explored in this section.

4.1 Data Source

In recent years, the field of skin lesion detection has grown significantly, resulting in several publicly available datasets. These datasets typically contain a wide variety of images representing different types of skin lesions. Among the publicly accessible datasets, the HAM10000 dataset from Kaggle's open data repository was chosen for this study. The decision to utilise the HAM10000 dataset was influenced by various factors, including dataset size, diversity of skin lesions, quality of the images, and relevance to the objectives of the study. In addition, the HAM10000 is a benchmark dataset in this field; utilising it assures that our work is comparable to existing studies.

4.2 Dataset Descriptions

The HAM10000 dataset is a collection of 10015 dermatoscopic images of common pigmented skin lesions released by Harvard Dataverse (Tschandl, 2018). The dataset contains a wide variety of skin lesion types, including actinic keratoses and intraepithelial carcinoma (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc). In addition to the dermatoscopic images, the HAM10000 dataset includes a metadata file that contains necessary details related to each image. The properties of metadata are shown in Table 1.

Attribute	Description
lesion_id	A unique identifier for each lesion.
image_id	A unique identifier for each dermatoscopic image in the dataset.
Dx	The diagnostic category of the skin lesion. Categories include Actinic keratoses and intraepithelial carcinoma (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc).
dx_type	The method of lesion diagnostic used. It specifies whether histopathology (histo), follow-up examination (follow_up), expert consensus (consensus), or in-vivo confocal microscopy (confocal) confirm the diagnostic category (dx).
Age	The patient's age at the time the image was taken.
Sex	The gender of the patient.
localization	The specific slocation on the body where the skin lesion is located.

Table 1. Attributes of HAM10000 metadata

4.3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an important phase in this study. The HAM10000 dataset will be thoroughly examined in this section, including its properties, distribution, and potential difficulties. Understanding the nature and distribution of the dataset is critical for developing an accurate classifier. Through this analysis, we seek to gain valuable insights that will guide the implementation of our deep learning models for skin lesion detection.

- **Frequency of each lesion type**

The HAM10000 dataset comprises seven distinct categories of skin lesions, as seen in the accompanying Figure 9. The Melanocytic Nevi lesion type accounts for around 65% of the dataset, making it the most prevalent lesion type. The remaining 35% of the data consists of other categories of skin lesions. It is worth noting that the dataset reveals a significant absence of images related to dermatofibroma and vascular lesions. The discovery of this insight is crucial in the development of deep learning models. In order to ensure precise and reliable predictions for all categories of lesions, it may be necessary to consider the uneven distribution of certain lesion types during the training of the model.

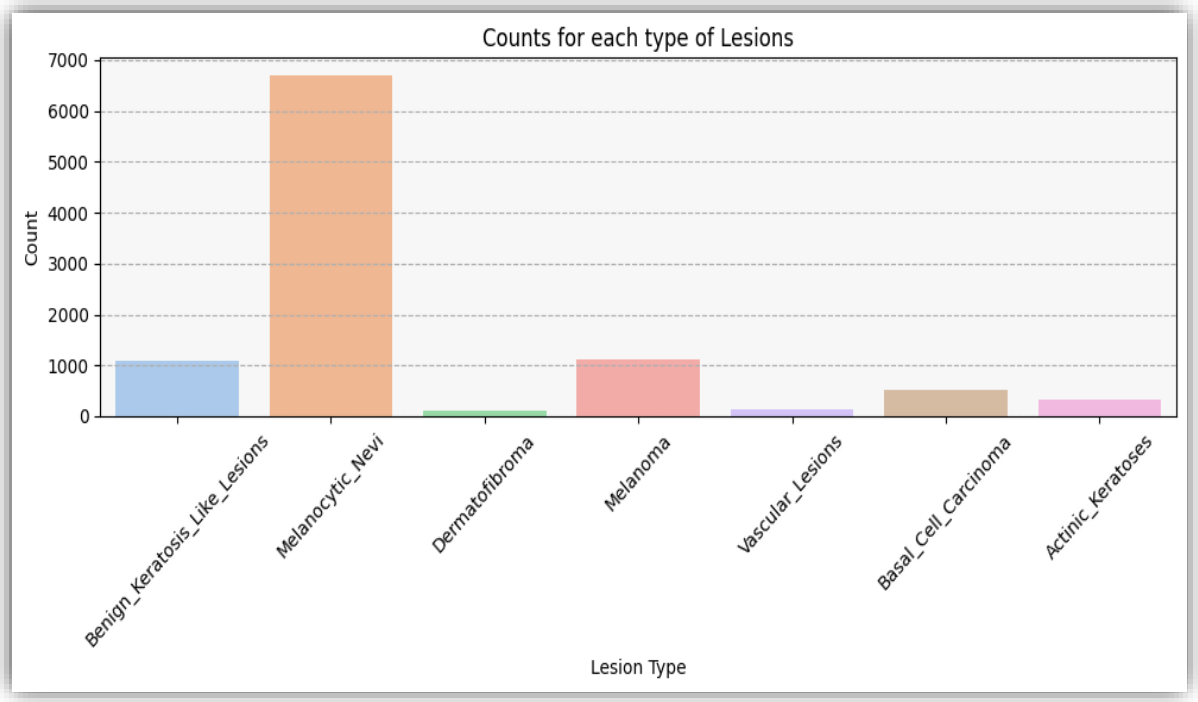


Figure 9. Frequency of Lesions

- **Gender Distribution**

The dataset is divided into three gender categories: female, male, and unknown. According to the bar chart in Figure 10, females account for 45.5% of the dataset, while males account for 54%. There is a small percentage, about 0.6%, where the gender information is labelled as unknown.

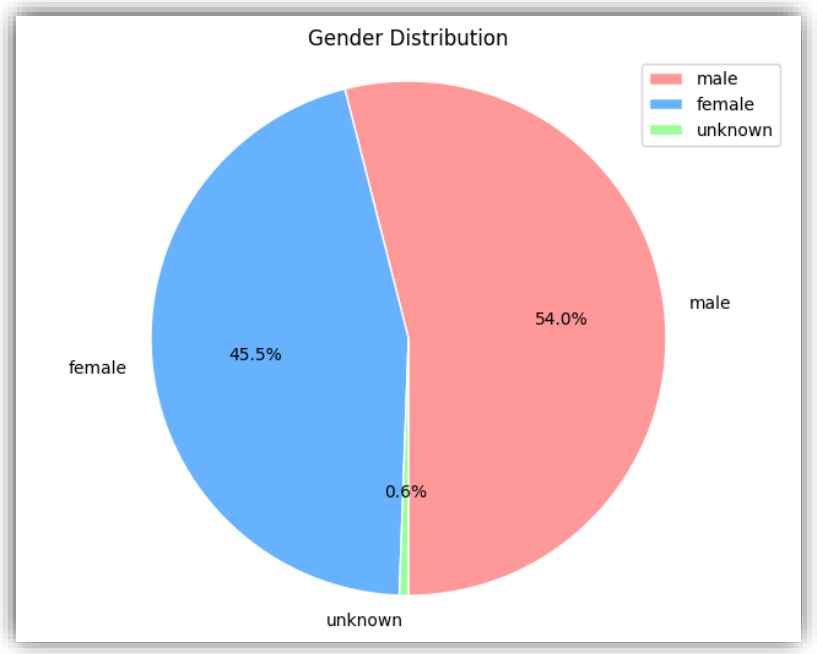


Figure 10. Gender Distribution

- **Age Distribution**

Skin lesions affect a substantial number of people aged 40 to 60 (Figure 11). Within this age range, the histogram showed a prominent peak, indicating a higher prevalence of skin lesions.

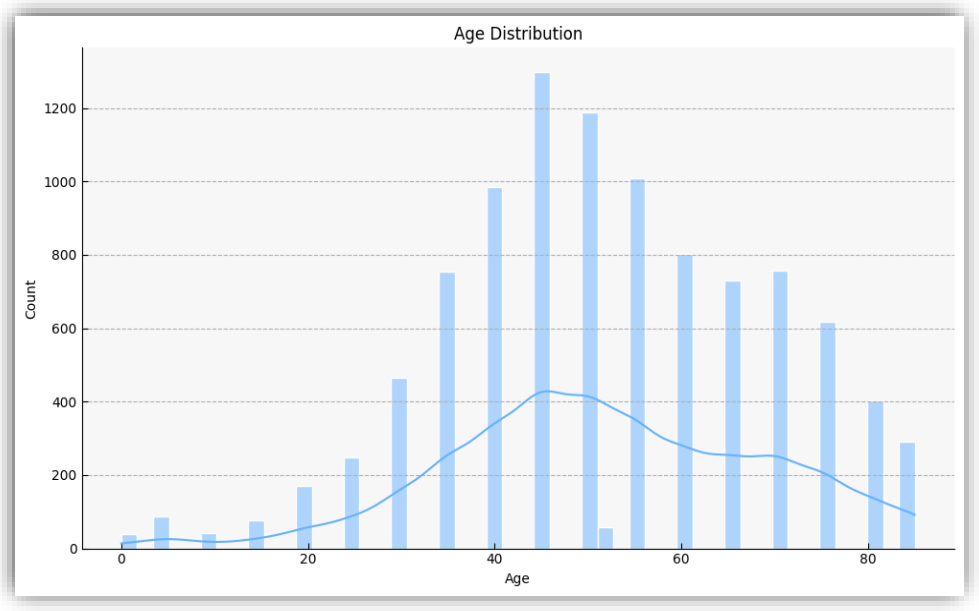


Figure 11. Age Distribution

- **Distribution of diagnosis method used**

The bar plot in Figure 12 depicts the distribution of skin lesion diagnosis approach used for skin lesion detection. The highest bar represents the 'histopathology' (histo) diagnosis category, suggesting that histopathological investigation was the most used diagnostic procedure in the dataset. It is used on 5340 skin lesion detection cases. 'Follow-up' diagnosis method is used in 3704 cases, 'consensus' in 902 cases, and 'confocal' in 69 cases.

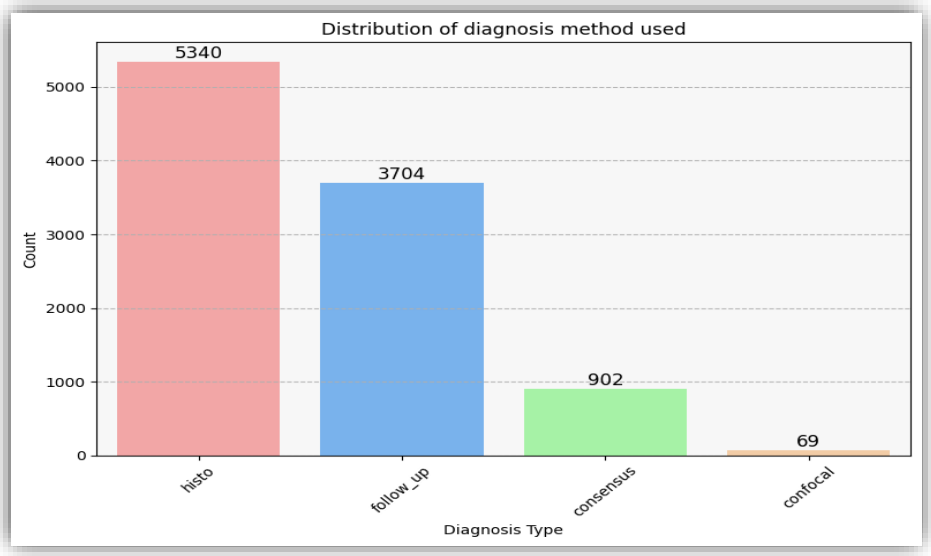


Figure 12. Distribution of Diagnosis methods

- **Location of Lesions**

The bar plot in Figure 13 depicts the relative frequency of skin lesions in various parts of the body. According to the findings, in many cases skin lesions are seen on the patient's backs. This makes the back the most prevalent region for skin lesions among the patients in the dataset. Following the back, other locations, such as the trunk, lower extremity, abdomen, and upper extremity, have a significant incidence of skin lesions. On the other hand, skin lesions appear to be less common in areas such as the ear, genital, hand, and acral.

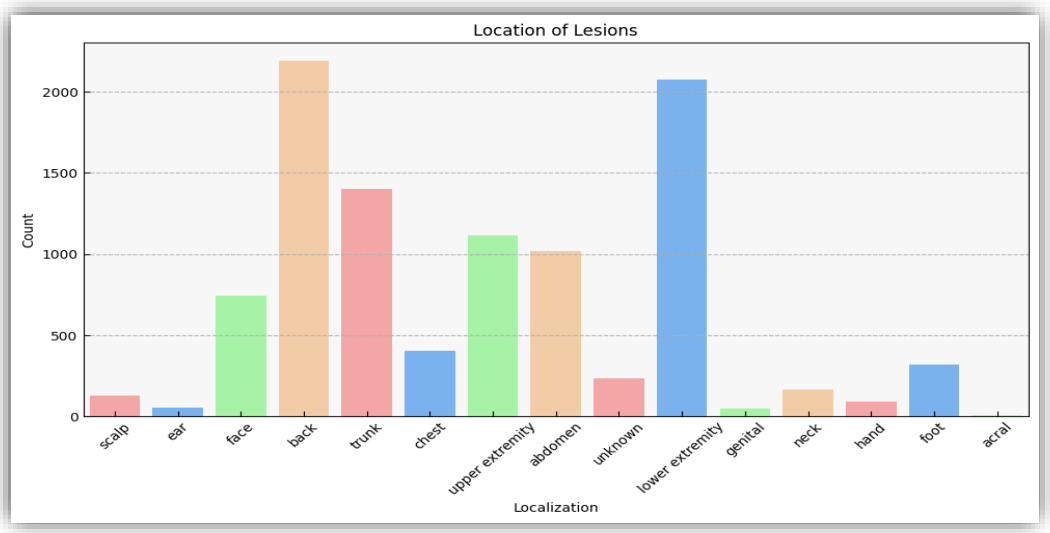


Figure 13. Lesion Location

- **Localization VS Gender**

Figure 14 demonstrates that the location of skin lesions differs between male and female individuals. The back is the most prevalent region for skin lesions in men, whereas the lower extremities are more commonly affected in female.

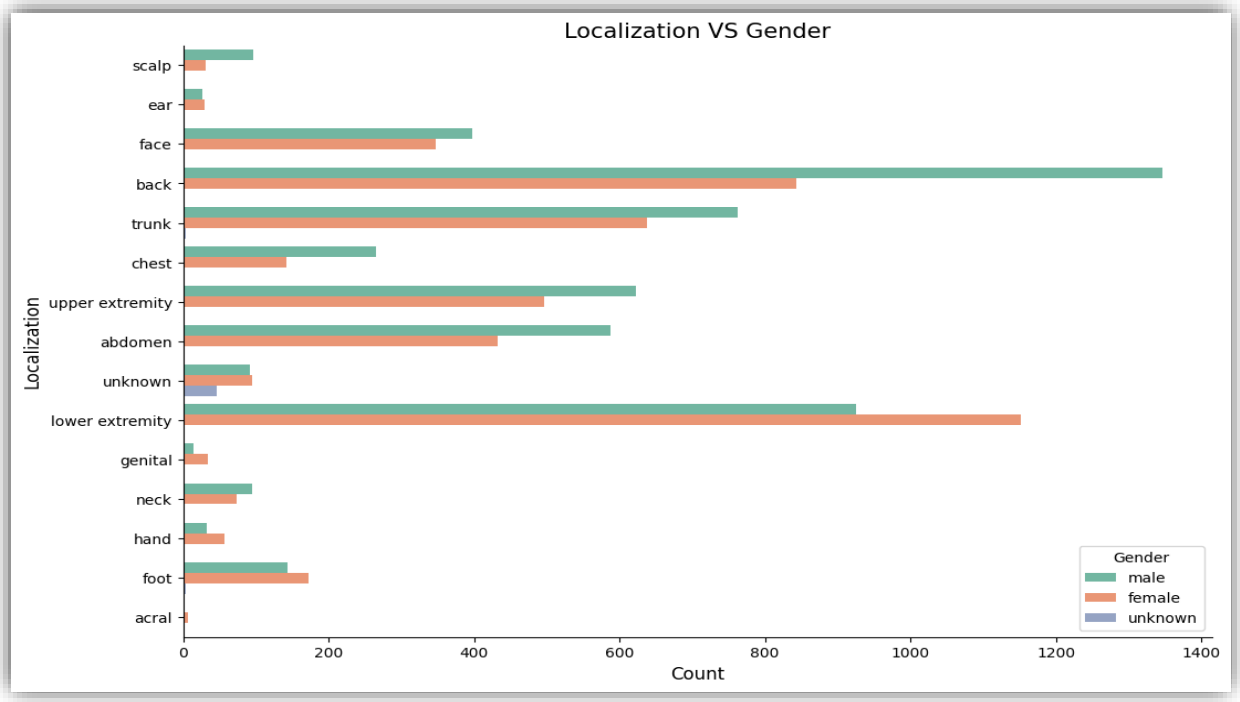


Figure 14. Localization and Gender

- **Image Dataset**

The HAM10000 dataset contains a large number of skin lesion photos from seven different types of skin lesions. In Figure 15, five representative lesion images from each of the seven classes in the dataset are represented. By investigating these lesion images, we can obtain insight into the visual differences between distinct classes.

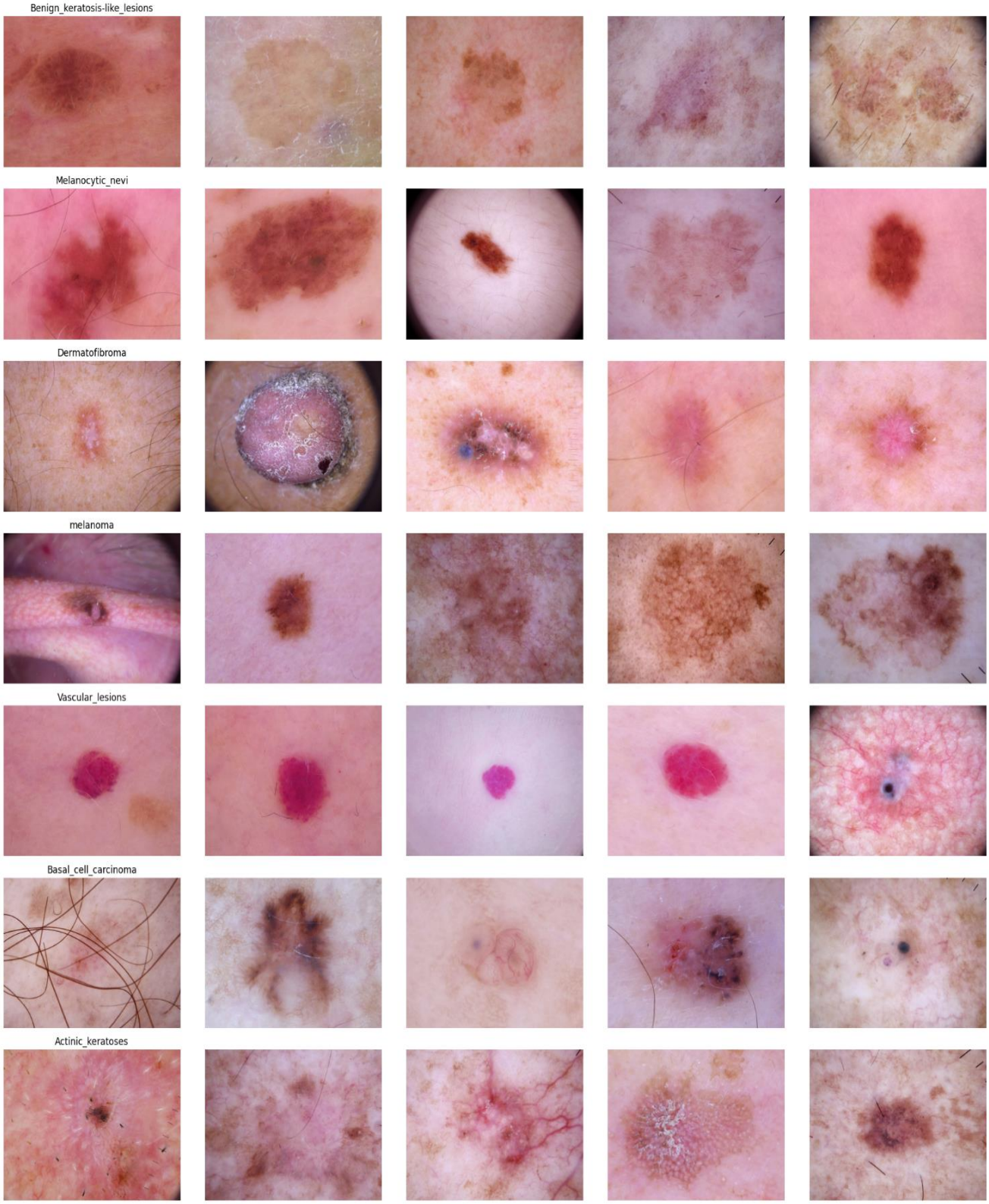


Figure 15. Skin Lesion Image Dataset

4.4 Data Augmentation and Normalization

In order to improve the efficacy and robustness of our skin lesion detection model, we used two key preprocessing techniques: Data Augmentation and Data Normalisation.

Data Augmentation:

Since different types of skin lesions are represented at different frequencies in our data, we are aware of the class imbalance problem. To tackle this, we implemented augmentation techniques to balance underrepresented classes, includes applying several transformations to current data samples to produce new, slightly changed samples while preserving the fundamental properties of the original data (Ayan & Ünver, 2018). This approach not only expands the dataset but also strengthens the model's capacity to adapt to various conditions. In the context of skin lesion recognition, data augmentation techniques include rotation, flipping, cropping, zooming, altering brightness and contrast, and introducing random noise. It uses the ImageDataGenerator class from TensorFlow Keras to perform data augmentation on a HAM10000 dataset. The Figure 16 depicts the original image in the dataset as well as five augmented images derived from it. The training set is enriched by these augmented examples, giving the model a deeper comprehension of various skin lesion appearances.

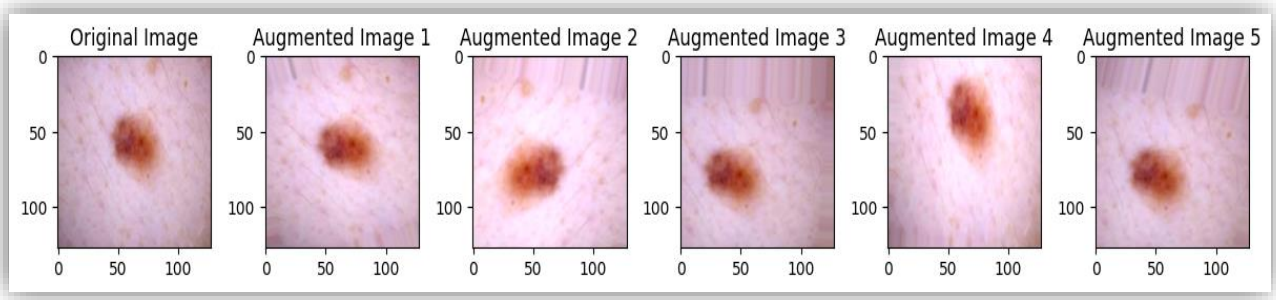


Figure 16. Data Augmentation

Data Normalisation:

Data normalisation involves transforming image pixel values to a standardised scale, typically between 0 and 1, in order to improve model training. This assures consistent learning across various images and avoids biases caused on by different pixel intensity ranges. Pixel values are centred on zero and adjusted for unit variance by computing the mean and standard deviation. This stage stabilises training, protects against gradient problems, and speeds up precise model convergence. Overall, normalisation makes a skin lesion detection model more accurate and reliable.

4.5 Image Resizing

Image resizing is a fundamental preprocessing step in working with deep learning model. The HAM10000 dataset, which contains skin lesion images, with the original size of image 450 pixels in height, 600 pixels in width, and three colour channels. However, this large size causes difficulties because of limited computing resources. To address this problem, a strategic choice

has been made to downsize the images to a more manageable and standardised dimension of 128 pixels by 128 pixels, while keeping the original three colour channels. This is especially important when dealing with huge datasets like HAM10000. This resized picture size of 128x128x3 is used as an input size for both the basic Convolutional Neural Network (CNN) model and transfer learning-based models. This standardised size enables fair comparisons and effective model training, regardless of the model architecture.

5. Experiment setup and Results

This section provides details about the experiment setup used for building the models. The baseline model and several transfer learning models were built with the well-known TensorFlow Keras framework. This section delves into the specifics of these models, describing their distinctive architectural details and configuration. Also, this section provides the performance of implemented models.

5.1 Dataset splitting

Before creating and deploying deep learning models, dataset splitting methods were used to assure the integrity and efficacy of our model's development and evaluation. The dataset is divided into three parts: training, testing, and validation, in the ratio of 70:20:10, respectively. The training set, which is the largest subset, provides the foundation of the model's learning process. Here, the model learns to identify significant trends and features from the data and modifies its parameters. After the model has been trained, the testing set is used to evaluate its performance. By evaluating the model's predictions on unobserved data, we can determine its capacity for making accurate predictions on new instances. The validation set is used for fine-tuning the model during the training phase. It improves the choice of hyperparameters and the architectural modification of the model. This set prevents overfitting by providing an unbiased assessment of the model's performance on new data.

5.2 Parameters

The efficacy of skin lesion detection utilising CNN and transfer learning models is strongly reliant on the careful selection and adjustment of several parameters. To undertake a fair and meaningful comparison of several models, it is important to maintain consistency across certain parameters. By maintaining the same values for these parameters across all models, any variations in performance may be attributed to the distinct architectural differences amongst the models rather than variations in the training process. The important parameters addressed in the experiment are detailed in the following subsections:

5.2.1 Model Compilation Parameters

• **Optimizer:** The optimizer decides how to update the model's weights during training to reduce the loss function. The 'Adam' optimizer was selected; it adjusts the learning rate based on previous gradient updates, frequently leading to faster convergence and improved performance.

• **Loss Function:** The loss function calculates the discrepancy between real labels and predicted values. The 'categorical_crossentropy' loss function was used since it is the best option for multi-class classification tasks with one-hot encoded labels.

• **Metrics:** The model's effectiveness was assessed using the 'accuracy' metric, which counts the percentage of properly categorised samples.

5.2.2 Model Training Parameters

- **Number of Epochs**: There were 20 epochs used for the training, which means the full dataset will be iterated through 20 times. The same maximum number of epochs has been defined for all models to ensure fair possibilities for convergence.

- **Batch Size**: The term ‘batch size’ describes the quantity of training samples used in a single iteration. The batch size was set to 32. In order to keep memory use and training performance consistent across all models, the same batch size is employed for training.

- **datagen.flow**: The datagen.flow creates batches of augmented data from the training set. This allows the model to receive batches of augmented data throughout each training iteration (epoch), and the weights will be modified in response to the performance of these batches.

5.3 Model Performance

The classification report was used to get an in-depth analysis of the model's performance across several classes. The following are the metrics from the classification report:

- **Precision**: Precision is the ratio of correctly anticipated positive results to the total predicted positives.

- **Recall**: The recall is calculated as the ratio of correctly classified positive samples to the total number of positive samples. The recall metric measures how well the model detects positive samples. The greater the recall, the more positive samples are found.

- **F1-score:** The harmonic mean of accuracy and recall is the F1-score. It gives a more balanced perspective of the model's performance, which is especially useful when the class distribution is unequal. A higher F1 score suggests a better balance of precision and recall.

- **Accuracy**: Accuracy is defined as the ratio of accurately predicted observations to total observations. Even though it is a general indicator of model performance, it could not be the optimal metric in cases when the distribution of classes is unbalanced.

- **Loss**: Loss value is a metric used to assess how well a model's predictions match the actual target values found in the training set.

- **Confusion Matrix:** A confusion matrix is an important tool for evaluating deep learning models since it provides a visual representation of their performance across different classes. It enables to assess how well a model's predictions correspond to the actual class labels. In the context of skin lesions multiclass classification, where there are more than two classes, the confusion matrix becomes a matrix with rows and columns representing each class.

- **ROC and AUC curve:** ROC (Receiver Operating Characteristics) and AUC (Area Under the Curve) are evaluation techniques used for analysing the performance of classification models in both binary and multiclass scenarios. An ROC curve displays the trade-off between true positive rate (sensitivity) and false positive rate across multiple categorization thresholds visually. AUC measures the overall performance of the model, with higher values suggesting greater discrimination.

5.4 CNN Implementation and Result

This section delves into the architecture of Convolutional Neural Network (CNN) model and efficiency of model for detecting skin lesions.

CNN Model Architecture:

CNN architecture is made up of three convolutional layers followed by a max-pooling layer. Using 3x3 filters in each convolutional layer permits the model to capture local patterns. Subsequent max-pooling layers gradually lower the spatial dimensions, improving the model's capacity to capture essential features at various scales. Following the convolutional layers, the model employs a flatten operation to convert the 2D feature mappings to a 1D vector. This continues with by a dense layer with 512 units and ReLU activation, which helps in capturing more abstract representation from input image. To reduce overfitting during training, a dropout layer with a rate of 0.5 is used. The final dense layer has 7 units, which corresponds to the number of skin lesion classes. Each class is assigned probabilities using the softmax activation function, resulting in a multi-class classification output. The total number of parameters in the baseline CNN model is 4,684,871 (Table 2). It comprises both trainable and non-trainable characteristics. The model modifies trainable parameters during training via backpropagation in order to enhance its performance on the task for which it was designed. These parameters are learned from the data and contribute to the model's capacity to generate accurate predictions. The non-trainable parameters are not modified during training. The number of non-trainable parameters in our model is zero, which signifies that baseline model has no non-trainable parameters. All parameters have been configured to be trainable.

Total number of parameters	4,684,871
Trainable parameters	4,684,871
Non-trainable parameters	0

Table 2. Parameters of Baseline CNN model

Model Performance:

The performance of baseline CNN model is evaluated using test dataset and validation dataset. During the testing phase, the model performed 63 evaluation steps, each step with a duration of an average time of 21.333 milliseconds. The computed average loss on the test dataset was around 0.6426, indicating how far the model's predictions differed from the actual values. Furthermore, the test dataset accuracy was 74.79%, indicating the proportion of accurately predicted cases. This suggests that the model displayed a decent level of ability to make predictions. In the validation assessment phase, the model completed 26 evaluation steps, each of which took an average of 10.366 seconds. The computed average loss on the validation dataset was roughly 0.6324, indicating the model's ability to generalise its predictions to unknown validation data. Furthermore, the validation dataset accuracy was roughly 76.93%, indicating that the model's prediction ability remained constant when applied to new data. The classification report is used to evaluate the model's accuracy in detecting lesion cases across several classes. We examined precision, recall, and F1-score for each class separately (Table 3). The class ‘Melanocytic Nevi’ has a high precision (0.86), meaning that the model's positive predictions are practically good. This shows that the model correctly predicts and catches most cases for this class. The interpretation of these metrics enables us to modify our optimisation strategies to each class's distinctive characteristics. We consider the support metric, which represents the number of instances in the test dataset that correspond to each class. The macro average, with a precision of 0.58, recall of

0.56, and F1-score of 0.54, provides a comprehensive view of all classes. Weighted average takes class distribution into account, resulting in accuracy of 0.76, recall of 0.76, and F1-score of 0.74. These averages provide a comprehensive assessment of the model's overall performance. Figure 17 display the ROC curve and AUC of the baseline CNN model, while Figure 18 shows the confusion matrix.

	Precision	Recall	F1-score	Support
Actinic Keratoses	0.27	0.27	0.27	60
Basal Cell Carcinoma	0.40	0.66	0.50	97
Benign Keratosis Like_Lesions	0.45	0.49	0.47	224
Dermatofibroma	0.35	0.38	0.41	27
Melanocytic Nevi	0.86	0.91	0.88	1320
Melanoma	0.75	0.41	0.53	246
Vascular Lesions	0.61	0.33	0.43	29
Accuracy		0.76		2003
Macro avg	0.58	0.56	0.54	2003
Weighted avg	0.74	0.76	0.74	2003

Table 3. Classification report of baseline CNN model

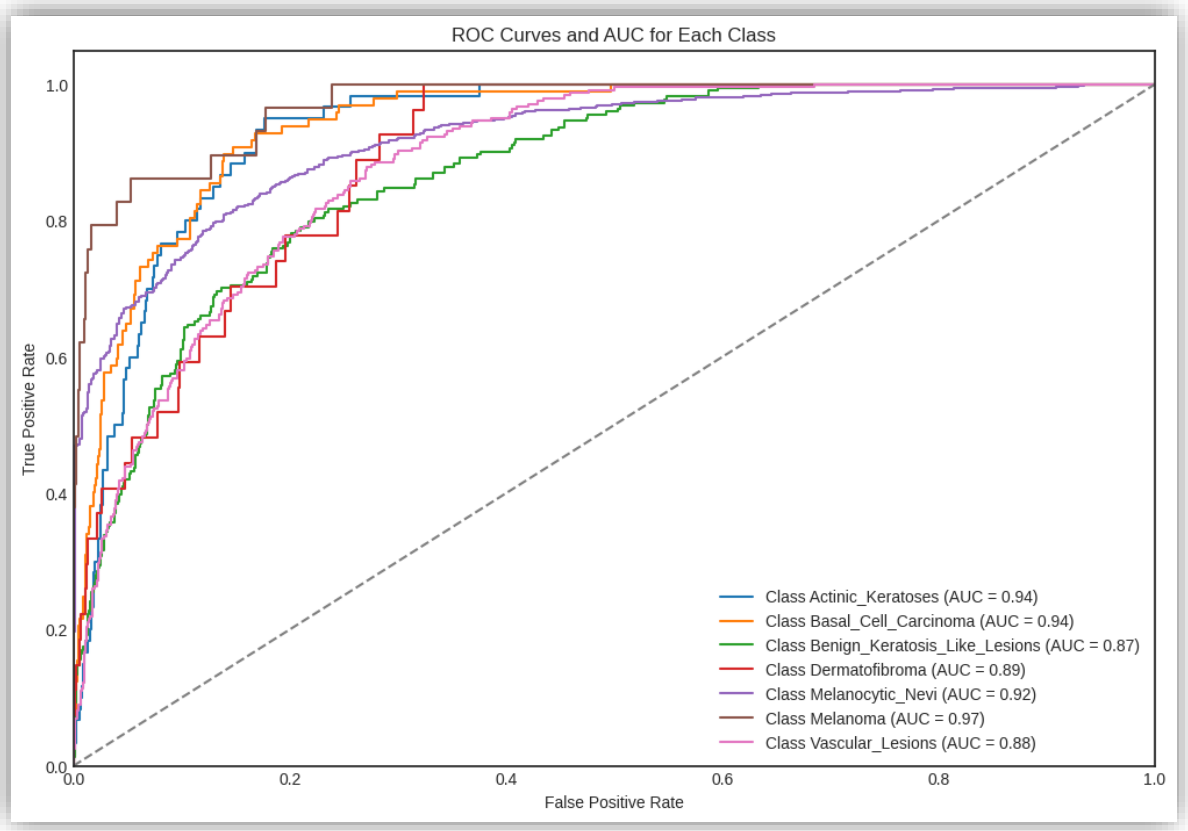


Figure 17. ROC curve and AUC of baseline CNN model

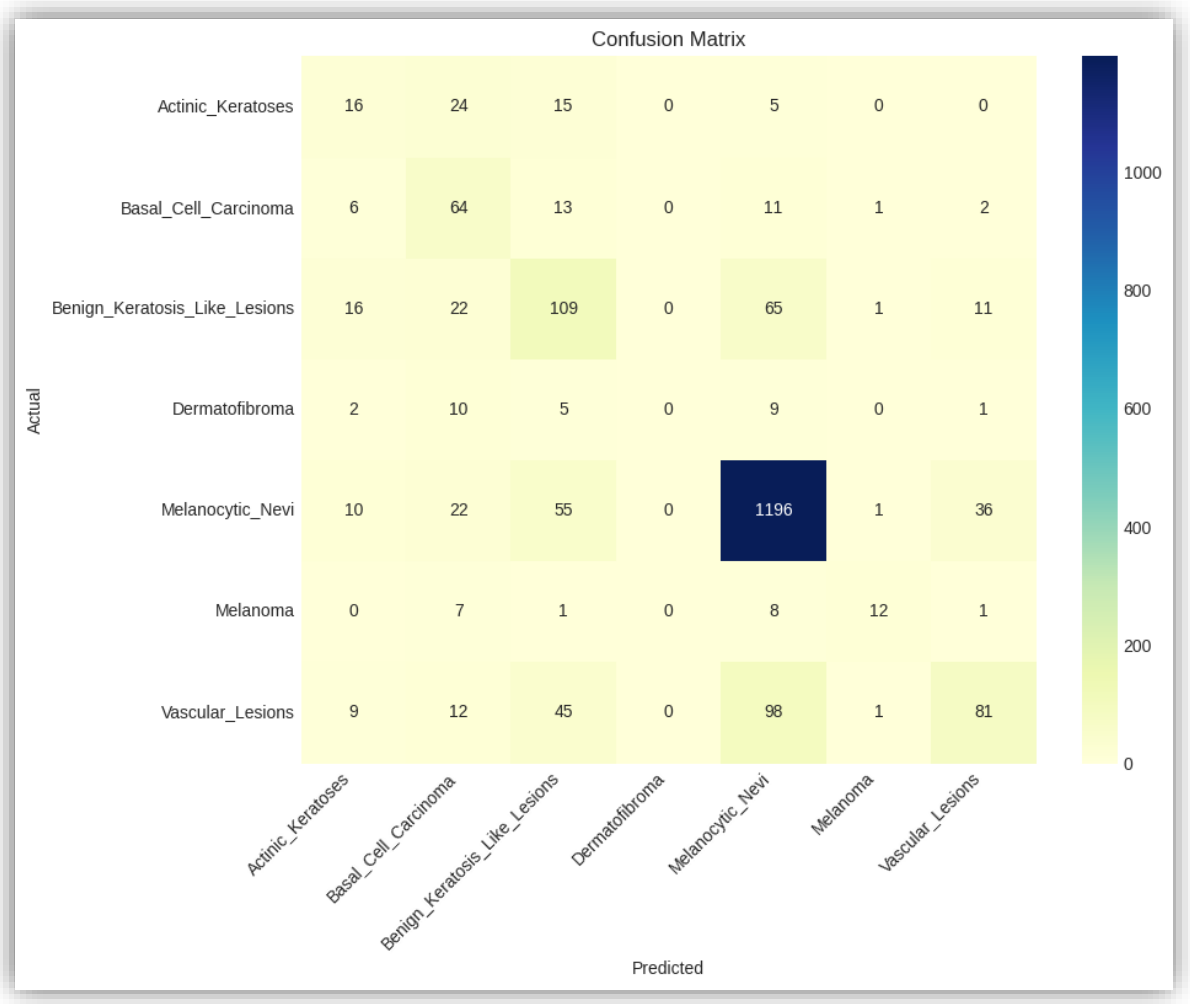


Figure 18. Confusion matrix of baseline CNN model

5.5 Transfer Learning Models Implementation

This section explores the designed structural components of the transfer learning models explained in 3.2.2 as well as their effectiveness in identifying skin lesions.

5.5.1 VGG 16 implementation and result

Model Architecture:

- **Feature extraction approach**
  In this method, the pretrained VGG16 base model is employed as a feature extractor, and a custom classification head is put on top to adjust the features for skin lesion detection. The VGG16 model serves as the architecture's foundation. It is loaded with pre-trained weights but does not include the top classification layer. Following that, custom layers are added to the base model. These layers are composed of a Global Average Pooling layer, which decreases the spatial dimensions of the retrieved features, and a Dense layer with Rectified Linear Unit

(ReLU) activation, which captures non-linear correlations. Finally, an output layer with a Softmax activation function is included to categorise skin lesion images into one of seven distinct classes.

- **Fine Tuning approach**
  The fine-tuning method is used to improve the model's capacity to recognise complicated patterns unique to the skin lesion identification task. Following the first feature extraction approach, in which the VGG16 base layers are frozen, a fine-tuning approach is introduced. In this approach, particular layers in the VGG16 model are selectively unfrozen to allow them to be modified during training. While the underlying model is being modified to capture more precise information from the skin lesion dataset, the custom classification layers continue to exist to transform these enhanced features into meaningful class predictions.

Model Performance:

The following findings show the performance of two techniques for detecting skin lesions using the VGG 16 model (Table 4 and 5).

Test Accuracy	75.73%
Test Loss	0.704
Validation Accuracy	72.31%
Validation Loss	0.736

Table 4. Performance of VGG16 feature extraction approach

Test Accuracy	77.28%
Test Loss	0.671
Validation Accuracy	74.67%
Validation Loss	0.708

Table 5. Performance of VGG16 fine tuning approach

The evaluation of the VGG16 model shows that it is successful in detecting skin lesions, with the fine-tuning technique outperforming the feature extraction approach. The accuracy of fine-tuning approach is higher than that of the baseline CNN model.

5.5.2 MobileNet implementation and result

Model Architecture:

- **Feature extraction approach**

The pretrained MobileNet base model is used as a feature extractor in the feature extraction technique, and a customised classification head is put on top to tailor the features for skin lesion detection. The foundation of the architecture is the base MobileNet model. It includes pre-trained

weights but does not include the upper classification layer. In the implementation of the MobileNet model, the same customised classification head layers are utilised as in the VGG16 model. The MobileNet basic model adds around 89 layers to the architecture. Furthermore, the custom classification layers include 3 layers, increasing the total model complexity. In this model, the total number of trainable parameters is 3,757,255.

- **Fine Tuning approach**

The fine-tuning method, which is also used on the MobileNet model, is meant to improve the model's efficiency. This approach involves deliberately unfreezing certain layers inside the MobileNet model, enabling them to be changed during training. The number of layers and parameters evolves under the fine-tuning technique. Unfreezing certain layers enables them to be adjusted, increasing model complexity. At the end of the model, custom classification layers are added to convert upgraded information into meaningful class predictions.

Model Performance:

The following Table 6 and 7 show the performance of two techniques for detecting skin lesions using the MobileNet model.

Test Accuracy	76.49%
Test Loss	0.704
Validation Accuracy	77.31%
Validation Loss	0.708

Table 6. Performance of MobileNet feature extraction approach

Test Accuracy	83.23%
Test Loss	0.561
Validation Accuracy	81.80%
Validation Loss	0.608

Table 7. Performance of MobileNet fine tuning approach

Both models showed respectable performance in the categorization of skin lesions in terms of accuracy and loss. The model that uses a technique for feature extraction performs quite similarly to the baseline CNN model. The fine-tuning technique outperformed feature extraction, yielding higher accuracy and reduced loss on both test and validation sets. This highlights the effectiveness of fine-tuning pre-trained models to adjust to specific tasks like skin lesion detection.

5.5.3 DenseNet implementation and result

Model Architecture:

- **Feature extraction approach**

The approach uses DenseNet121 as its base in this method. Adapting the retrieved characteristics for skin lesion identification requires the addition of custom classification layers on top of the base DenseNet121 model. In order to allow for a fair comparison, the custom layers include the same layers as those used in the prior transfer learning models.

- **Fine Tuning approach**

The fine tuning approach modifies the underlying Densenet121 model with a purposeful layer alteration. The DenseNet121 model's particular layers are deliberately unfrozen, allowing them to be retrained using the skin lesion dataset. In our model, the first 70 layers are unfrozen. Following the unfreezing of the specified layers, a custom classification head is added to the basic model.

Model Performance:

The following findings show the performance of two techniques for detecting skin lesions using the DenseNet model (Table 8 and 9).

Test Accuracy	76.49%
Test Loss	0.704
Validation Accuracy	77.31%
Validation Loss	0.708

Table 8. Performance of DenseNet feature extraction approach

Test Accuracy	78.63%
Test Loss	0.584
Validation Accuracy	73.79%
Validation Loss	0.718

Table 9. Performance of DenseNet fine tuning approach

Feature extraction achieved 76.49% test accuracy and 0.704 test loss, along with 77.31% validation accuracy and 0.708 validation loss. In comparison, fine-tuning produced somewhat lower validation accuracy (73.79%) and greater validation loss (0.718), but better test accuracy (78.63%) and lower test loss (0.584). Fine-tuning resulted in improved test performance but small generalisation trade-offs in validation.

5.5.4 Inception implementation and result

Model Architecture:

- **Feature extraction approach**

  The feature extraction strategy employs the InceptionV3 model, which has been pretrained on a broad dataset, to extract relevant characteristics from skin lesion pictures. The pretrained layers of InceptionV3 are frozen, restricting further weight alterations during training. Custom classification layers are built on top of the frozen layers, just like in previous transfer learning models.

- **Fine Tuning approach**
  Like the feature extraction method, the fine-tuning method utilises the pretrained InceptionV3 model as its foundation. In contrast to feature extraction, a fraction of the basic model's layers is unfrozen, enabling them to be updated during training. The custom classification layers on top of the underlying model stay unchanged and continue to refine the additional features into useful skin lesion predictions.

Model Performance:

The results below (Table 10 and 11) demonstrate performance of two techniques for detecting skin lesions using the InceptionV3 model.

Test Accuracy	71.59%
Test Loss	0.799
Validation Accuracy	72.06%
Validation Loss	0.838

Table 10. Performance of InceptionV3 feature extraction approach

Test Accuracy	75.88%
Test Loss	0.616
Validation Accuracy	79.30%
Validation Loss	0.573

Table 11. Performance of InceptionV3 fine tuning approach

The test accuracy for the feature extraction method was around 71.59%, although the validation set accuracy was closer to 72.07%. The results of the fine-tuning method, on the other hand, were better, with a test accuracy of around 75.89% and a validation accuracy of about 79.30%. This shows that, compared to the feature extraction method, accuracy was enhanced by fine-tuning the layers of the pre-trained model.

5.5.5 ResNet implementation and result

Model Architecture:

- **Feature extraction approach**

In the case of the ResNet model, feature extraction model implementation begins by using the ResNet-50 architecture as a base. Custom classification layers are placed on top of the ResNet-50 model to modify this architecture for the goal of identifying skin lesions. These new layers are designed to be like the architecture used in the previously described transfer learning models, guaranteeing a fair and consistent comparison.

- **Fine Tuning approach**

In the fine-tuning method, we take the ResNet-50 model and purposefully change a few layers for retraining with HAM10000 dataset. In particular, the ResNet-50 model's some top layers are unfrozen, enabling them to be adjusted to work best with our dataset. Following that, a custom classification head is added to the underlying model.

Model Performance:

The findings below show how two strategies for identifying skin lesions using the ResNet-50 model performed (Table 12 and 13).

Test Accuracy	71.09%
Test Loss	0.8289
Validation Accuracy	70.44%
Validation Loss	0.849

Table 12. Performance of ResNet-50 feature extraction approach

Test Accuracy	72.24%
Test Loss	0.7754
Validation Accuracy	71.94%
Validation Loss	0.7950

Table 13. Performance of ResNet-50 fine tuning approach

In the feature extraction method, ResNet-50 achieved a test accuracy of 71.09% with a test loss of 0.8289, along with a validation accuracy of 70.44% and a validation loss of 0.849. ResNet-50 showed increased test accuracy (72.24%) and decreased test loss (0.7754) after being fine-tuned, and validation accuracy of 71.94% and validation loss of 0.7950. Comparatively, the baseline CNN model achieved higher accuracy than both models.

6. Model Evaluation and Optimisation

	Baseline CNN	VGG16	MobileNet	DenseNet	InceptionV3	ResNet-50
Test Accuracy	74.79%	75.73%	83.23%	78.63%	75.88%	72.24%
Test Loss	0.6426	0.704	0.561	0.584	0.616	0.7754
Validation Accuracy	76.93%	72.31%	81.80%	73.79%	79.30%	71.94%
Validation Loss	0.6324	0.736	0.608	0.718	0.573	0.7950

Table 14. Performance of all implemented models for comparison

After building and assessing the performance of the baseline CNN model and other transfer learning models, it was discovered that the MobileNet model performs very well according to the performance matrix (Table 14). The MobileNet model with the fine tuning method achieves an accuracy of 83.23%. Various optimisation strategies were used to improve the accuracy and performance of the model. In the earlier model version, a simpler training process was used without callbacks such as ReduceLROnPlateau and EarlyStopping. These callbacks are used to obtain the best results and make training more efficient. During model training, we used these callbacks to optimise the MobileNet model. The ReduceLROnPlateau callback changes the learning rate during training based on the validation loss. By monitoring the validation loss, this callback lowers the learning rate by 20% if no improvement is seen for three consecutive epochs, improving convergence, and assisting the model in escaping local minima. The EarlyStopping callback is very helpful in stopping overfitting and reducing training time. It keeps track of the validation loss and stops training if there is no improvement after five successive epochs. It also helps to save computational power while deep learning models are being trained. During the MobileNet model training with these callbacks, the number of epochs is cut from 20 to 14, and the test accuracy (83.97%) is slightly improved. Also, the validation accuracy improved from 81.79% to 83.79%.

Another optimisation method is incorporated into the model by adding a batch normalisation layer. Batch normalisation, often known as BatchNorm, is an effective approach for improving the convergence and generalisation of deep neural networks. BatchNorm works by normalising the activations in each mini-batch during training, which has several important effects. First, it stabilises the training process, making it less sensitive to how the weights are set up at the beginning and how fast the learning rate is set. This stabilisation leads to faster convergence, which makes it easier for the model to learn. By adding batch normalisation to our model, we were able to improve the stability of training, speed up convergence, and improve the model's performance. The accuracy of the optimised MobileNet model is 86.96%. The classification report is used to evaluate the model's performance in each class (Table 15). Also, we assessed the performance indicators, such as the macro average and weighted average. The macro average represents the overall accuracy and balance of the model, with values of 0.80 for precision, 0.76 for recall, and 0.77 for the F1-Score. When class imbalance is taken into account, the weighted average produces excellent accuracy and recall values of 0.87, demonstrating the model's efficacy in a variety of classes. These findings support the model's precision and efficiency in handling class imbalances in skin lesion detection. The confusion matrix and ROC curve of the analysed model are shown in Figures 19 and 20, respectively. Figure 21 depicts the accuracy and loss trends seen in our optimised model's training data throughout epochs, as well as the accuracy and loss patterns observed in validation data.

	Precision	Recall	F1-score	Support
Actinic Keratoses	0.74	0.57	0.64	60
Basal Cell Carcinoma	0.88	0.74	0.80	97
Benign Keratosis Like_Lesions	0.72	0.80	0.76	224
Dermatofibroma	0.68	0.78	0.72	27
Melanocytic Nevi	0.91	0.96	0.94	1320
Melanoma	0.80	0.56	0.66	246
Vascular Lesions	0.90	0.90	0.90	29
Accuracy	0.87			2003
Macro avg	0.80	0.76	0.77	2003
Weighted avg	0.87	0.87	0.86	2003

Table 15. Classification report of optimised MobileNet model

Figure 19. Confusion Matrix of optimised MobileNet model

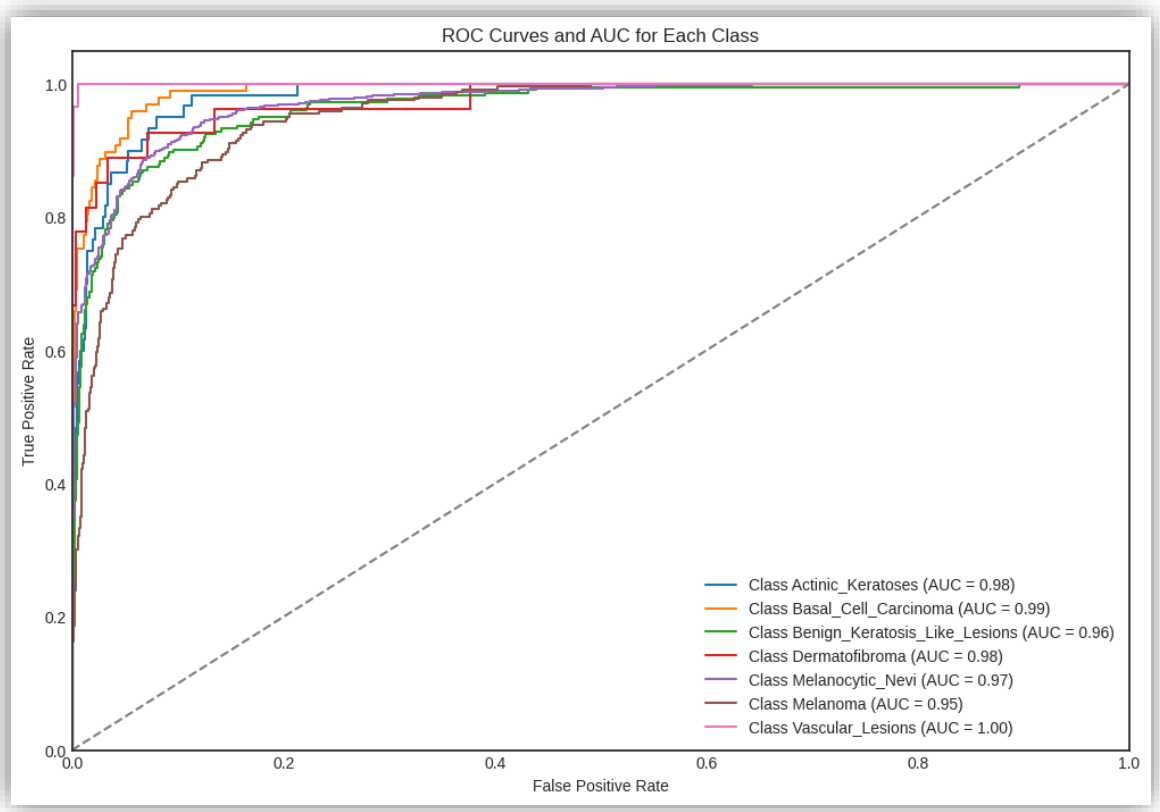


Figure 20. ROC curve and AUC of optimised MobileNet model

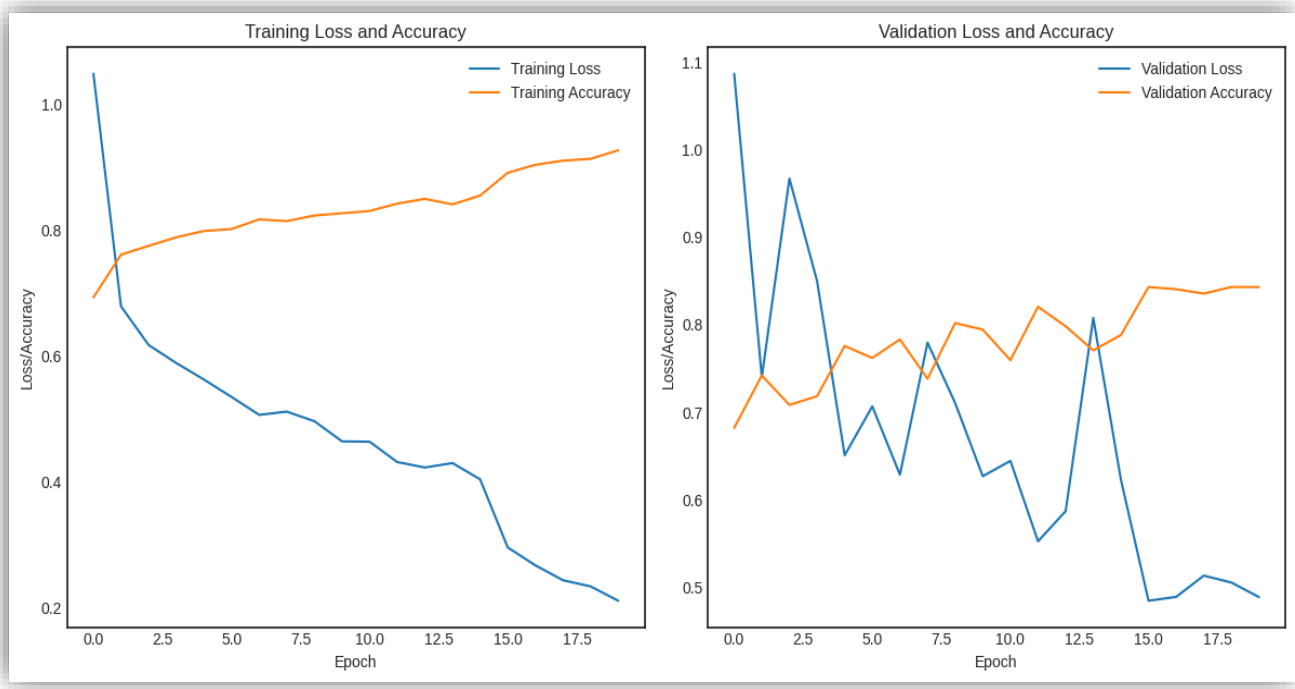


Figure 21. Accuracy and loss of optimised MobileNet model

7. Web Application Development

The web application is created to enable interaction with an implemented deep learning model to detect skin lesions. This section focuses on the development of a web application that detects skin lesions using the Flask framework. For medical professionals and individuals concerned about skin health, the incorporation of powerful deep learning models into a user-friendly interface provides an accessible and efficient solution.

7.1 Flask Framework

Flask is a microweb framework used for building web applications using the Python programming language (Vyshnavi & Malik, 2019). It is a flexible and lightweight web framework popular for its simplicity and adaptability (Brower, 2023). The architecture of Flask enables the smooth integration of machine learning and deep learning models, user interfaces, and backend capabilities, resulting in a strong foundation for the skin lesion detection application (Figure 22).

7.2 User Interface

The user interface of the web application has been made easy to understand and use. Users can upload skin lesion images directly or provide links to images through the interface. This adaptable approach accommodates a wide range of user preferences, making the web app accessible and user-friendly for medical practitioners and individuals. The web application's user interface (UI) was created with a user-friendly approach and cross-device compatibility in mind, guaranteeing a seamless experience for users accessing the platform from multiple devices. This user-friendly design allows users to engage with the web app effortlessly, whether they are using a desktop computer, laptop, mobile phone, tablet, or other devices. The UI is built using responsive design principles, which means that the layout and elements of the interface automatically adapt to different screen sizes and resolutions. The simple and user-friendly design of the web app allows individuals to use it effortlessly.

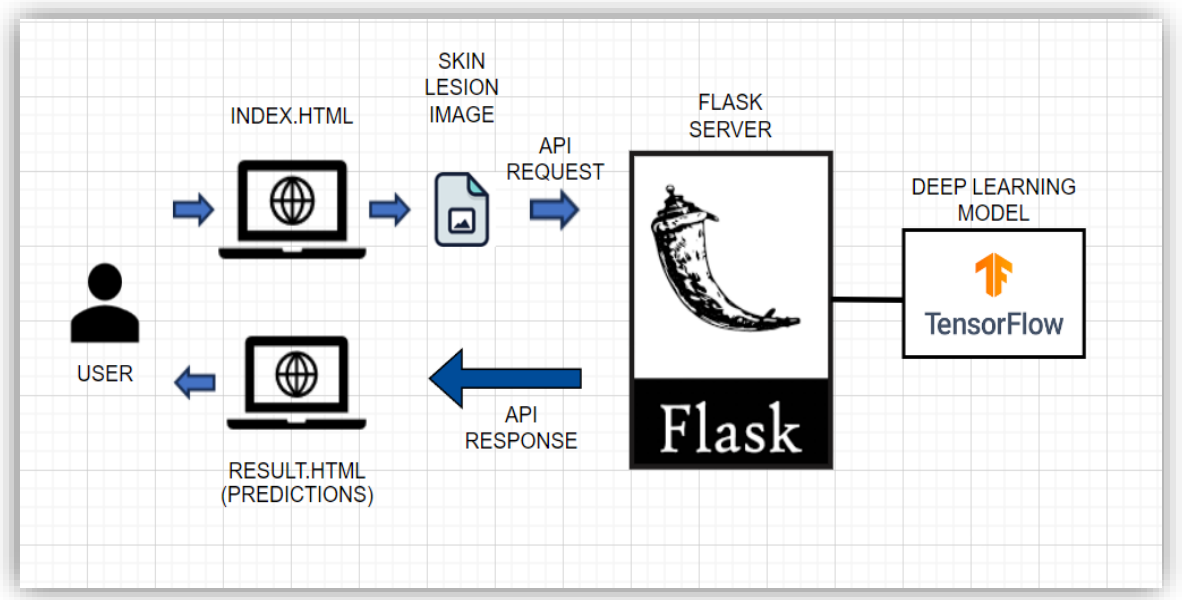


Figure 22. Web application architecture

7.3 Web Backend

As previously demonstrated in experiments, the transfer learning model employing the MobileNet model with a fine tuning approach outperforms baseline CNN and other transfer learning models. So, a finely tuned, optimised MobileNet model is integrated into the backend of the web app. It allows a seamless connection between the model and the web app's UI. When users upload their skin lesion image or provide an image URL, the backend processes the incoming image. This image is then sent to the MobileNet model for classification. Based on the model's learning, it analyses the features of a skin lesion and then classifies it correctly into one of the defined classes. Once the model has completed its analysis, the backend generates and displays the findings to the user. The predicted class of the skin lesion is displayed on the interface, along with a confidence score. This score provides insight into the model's level of confidence in its predicted class, providing users with useful information about the reliability of the results.

7.4 Web App Demonstration

Demo 1:

- The skin lesion picture is taken from the HAM10000 dataset, and it is loaded into the web interface (Figure 23).

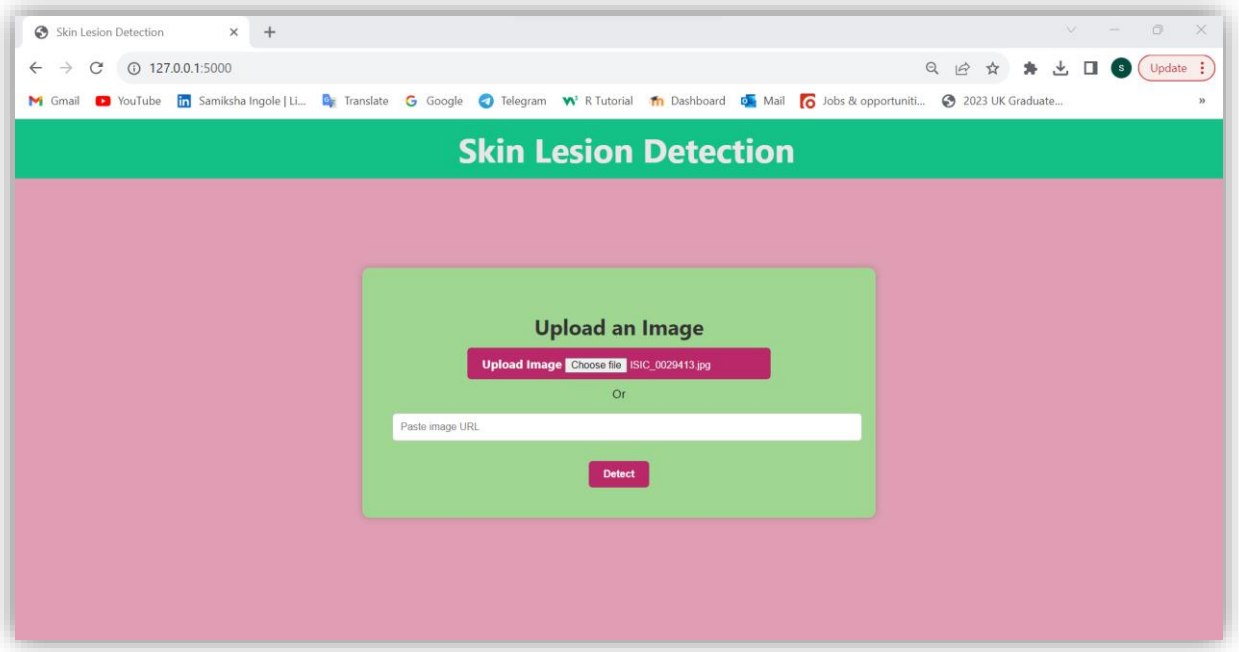


Figure 23. Image ID ISIC_0029413.jpg uploaded

- The backend processes the picture after it has been loaded and predicts the lesion class as benign keratosis-like lesions with a confidence of 0.918127 (Figure 24). The Figure 25 displays the actual label for the image from the HAM10000 metadata, where bkl stands for benign keratosis-like lesions.

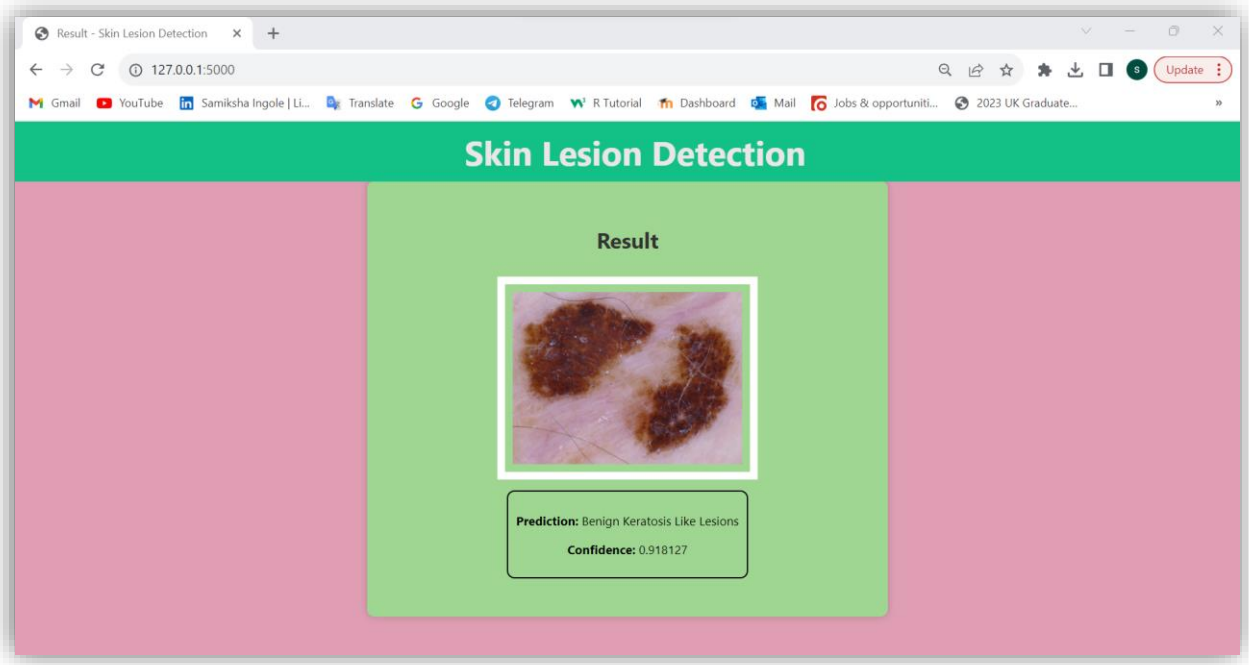


Figure 24. Output of Lesion Image uploaded

69	HAM_0004641	ISIC_0025099	bkl	histo	60	male	abdomen
70	HAM_0004641	ISIC_0031485	bkl	histo	60	male	abdomen
71	HAM_0005801	ISIC_0029413	bkl	histo	70	female	abdomen
72	HAM_0005801	ISIC_0029576	bkl	histo	70	female	abdomen
73	HAM_0004341	ISIC_0031967	bkl	histo	70	female	scalp

Figure 25. Image Metadata from HAM10000 dataset

Demo 2:

- In this demo, we evaluate the model's capacity to make predictions on unseen data, specifically focusing on its ability to handle data that falls outside the dataset. To do this, we obtain a melanocytic nevi lesion image URL from the Google Chrome. Then, for analysis, we paste this picture URL into our web app (Figure 26).

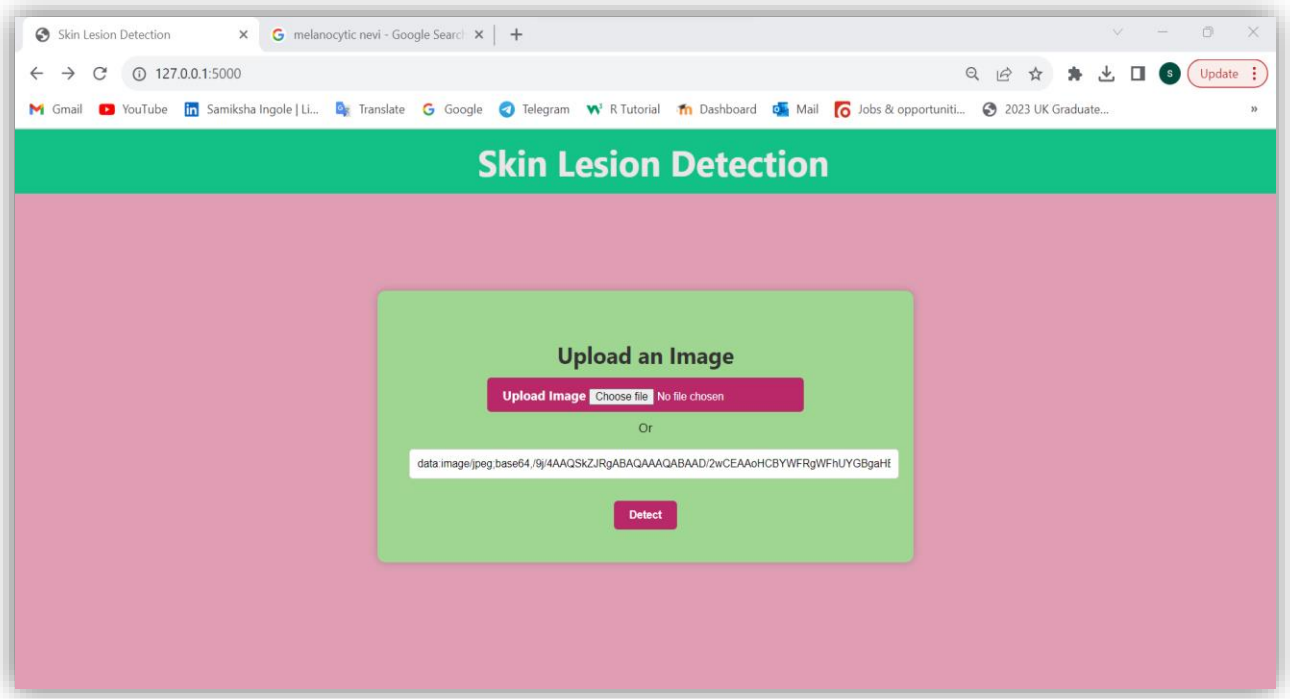


Figure 26. Image URL provided

- The backend system has successfully retrieved the image using the provided URL and then processed it. With a high confidence score of 0.9446023, the system's predictive capacity clearly classified the image as a melanocytic nevi lesion (Figure 27).

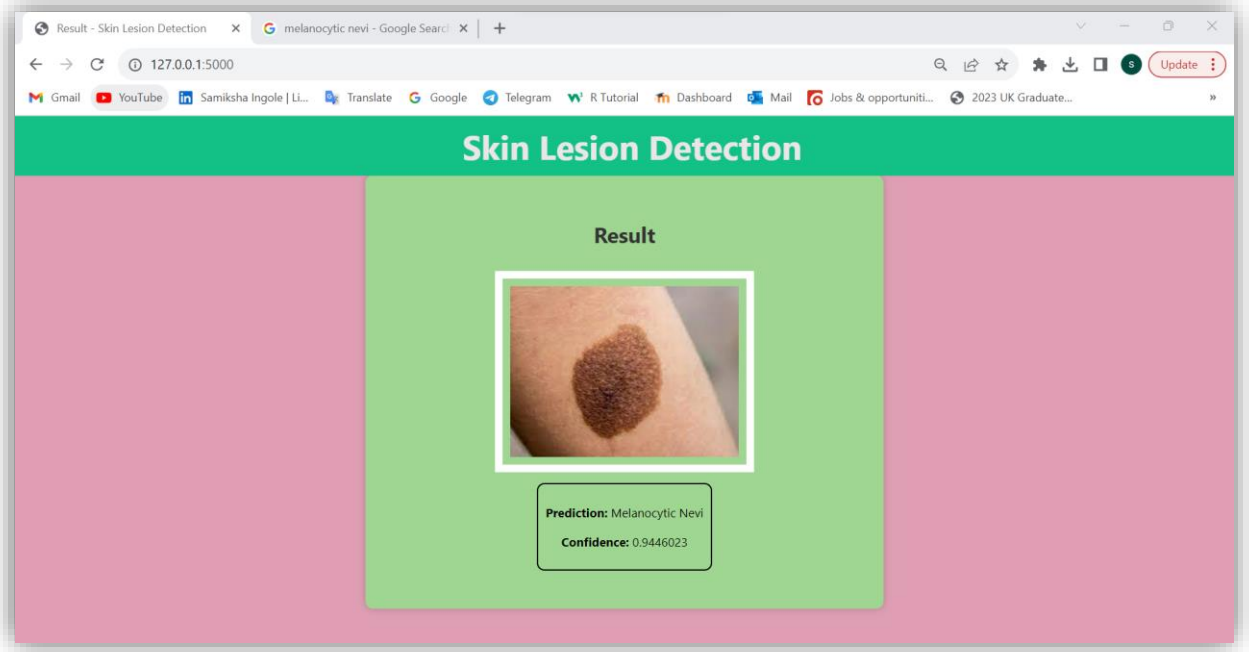


Figure 27. Output of provided image URL

7.5 Web App Testing

Manual testing is performed to evaluate the functionality of the web app. The following test cases have been designed for testing a web app for skin lesion detection (Table 16).

ID	Test Case	Expected Result	Status
1.	Opening web app on any browser	App loaded	✓
2.	Resizing browser window	UI adapts to different browser sizes	✓
3.	Clicking submit button without uploading image	App displays warning message for upload image	✓
4.	Entering invalid Image Link	Display Error message for invalid URL	✓
5.	Uploading image with different lesion types	Correct Class is predicted	✓
6.	Providing Link to image	Backend successfully able load the image	✓

Table 16. Web application test cases

8. Discussion

In this research, we examined the use of deep learning techniques to identify skin lesions. This study accurately classifies the skin lesion using the baseline CNN model and other transfer learning models. The parameter settings of the models were maintained in accordance with one another to guarantee a fair assessment of model performance. The constructed convolutional neural network (CNN) in our experiment showed commendable performance, achieving results that are comparable to previous research work. By using fewer layers and parameters in the proposed CNN model than in earlier work, it was able to achieve excellent accuracy. In addition, we investigated the use of transfer learning models, which also performed well. Notably, our study showed that the transfer learning model with fine-tuning method performed better than the feature selection method.

Among the various transfer learning models tested, MobileNet stood out with its exceptional accuracy, particularly when optimisation techniques were applied. The performance of the improved MobileNet model outperforms prior research in terms of accuracy. For instance, our implementation of the MobileNet model for skin lesion detection on the benchmark dataset yielded an accuracy rate of 86.96%, surpassing the accuracy reported in previous recent studies (Popescu et al., 2022; Castro-Fernández et al., 2020) that also used MobileNet. Additionally, a study by Sae-Lim et al. (2019) achieved an accuracy of 83.23% using a modified MobileNet model with a similar dataset and data augmentation. Also, it outperforms the findings of prior research (Salian et al., 2020), which attained an accuracy rate of 81.52% without data augmentation and 82% with data augmentation. These results underscore the effectiveness of our deep learning approach in the field of skin lesion detection.

However, despite these promising outcomes, our project faced certain limitations and challenges. Notably, we encountered the common issue of data imbalance, which is common in medical datasets. This class imbalance posed challenges for model evaluation and interpretation. As a result, the classification reports for all models showed high performance for the majority-class Melanocytic Nevi lesion, while the performance for minority classes varied. The limitation of computer resources was another significant difficulty. We were unable to build larger models or run many epochs because of the limited computing resources available.

9. Conclusion and Future Work

Our study has shown the potential of deep learning techniques in the diagnosis of skin lesions through the building of a convolutional neural network (CNN) and five transfer learning models. We have designed our own 10-layered CNN architecture. The performance of the CNN model demonstrates that training a well-designed model from scratch also produces good results. Throughout our testing, each model that we deployed performed well. However, the MobileNet model emerged as the outstanding performer, with the best performance in terms of accuracy, recall, F-score, and AUC. We suggested optimisation strategies to further improve MobileNet's performance. One of the optimisation options we suggested was the use of callbacks throughout the model's training phase. By using callbacks in our training process, we added a level of flexibility that made it possible for the MobileNet model to learn more quickly and effectively. We also included a batch normalisation layer as another crucial optimisation technique. The implementation of batch normalisation played a crucial role in improving the MobileNet model's performance, leading to a remarkable accuracy boost. With these optimisation methods, the MobileNet model's accuracy increased to 87%, setting a new benchmark for performance. We implemented a user-friendly web application using Flask to offer easy access to our established model. This platform serves as a gateway to our advanced skin lesion detection model, facilitating access for both healthcare professionals and the user community. By bridging the gap between cutting-edge technology and end-users, online applications enable individuals to take advantage of our study in their pursuit of accurate and rapid skin lesion diagnostics.

The study recommended carrying out testing on a more varied dataset that includes skin lesions from people with varying skin tones from various places throughout the world. This would make sure that the system is reliable and capable of correctly identifying skin lesions in individuals with various skin tones and origins. Through the investigation of methods like adaptive sampling and ensemble learning, we will keep addressing the class imbalance in medical datasets. More advanced models may be created by utilising advanced technologies like TPUs and edge computing devices.

Our findings and recommendations for future studies provide important new perspectives on the potential of deep learning techniques for skin lesion detection. One important finding from our study is the significant influence early skin lesion detection can have on the prompt diagnosis of skin cancer. Skin cancer, including melanoma and non-melanoma varieties, is a rising worldwide health problem that might have serious consequences if not detected and treated immediately. They emphasise the significance of continuing this study in order to improve the precision and usability of our skin lesion detection method in actual healthcare environments.

References

Abhishek, K. and Hamarneh, G., 2021, April. Matthews correlation coefficient loss for deep convolutional networks: Application to skin lesion segmentation. In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 225-229). IEEE.

Ahmed, et al. (2020) 'Skin Lesion Classification With Deep CNN Ensembles', in 2020 28th Signal Processing and Communications Applications Conference (SIU), pp. 1–4. doi: 10.1109/SIU49456.2020.9302125.

Al-Asadi, M. and Altun, A. A. (2022) 'Deep learning with SMOTE techniques for improved skin lesion classification on unbalanced data', Selcuk University Journal of Engineering Sciences, 21(3), pp. 97–104.

Albelwi, S.A., 2022. Deep Architecture based on DenseNet-121 Model for Weather Image Recognition. International Journal of Advanced Computer Science and Applications, 13(10).

Alhadhrami, E. *et al.* (2018) 'Transfer learning with convolutional neural networks for moving target classification with micro-Doppler radar spectrograms', in, pp. 148–154. doi: 10.1109/ICAIBD.2018.8396184.

Ali, L. et al. (2021) 'Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures', Sensors, 21, p. 1688. doi: 10.3390/s21051688.

Al-Saffar, A.A.M., Tao, H. and Talab, M.A., 2017, October. Review of deep convolution neural network in image classification. In 2017 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET) (pp. 26-31). IEEE.

Al-Saqqa, S., Sawalha, S. and AbdelNabi, H. (2020) 'Agile software development: Methodologies and trends', International Journal of Interactive Mobile Technologies, 14(11).

Ansar, W., Shahid, A.R., Raza, B. and Dar, A.H., 2020. Breast cancer detection and localization using mobilenet based transfer learning for mammograms. In Intelligent Computing Systems: Third International Symposium, ISICS 2020, Sharjah, United Arab Emirates, March 18–19, 2020, Proceedings 3 (pp. 11-21). Springer International Publishing.

Ayan, E. and Ünver, H. M. (2018) 'Data augmentation importance for classification of skin lesions via deep learning', in 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), pp. 1–4. doi: 10.1109/EBBT.2018.8391469.

Basheera, S. and Ram, M. S. S. (2019) 'Classification of Brain Tumors Using Deep Features Extracted Using CNN', Journal of Physics: Conference Series. IOP Publishing, 1172(1), p. 012016. doi: 10.1088/1742-6596/1172/1/012016.

Bechelli, S. and Delhommelle, J. (2022) Machine learning and deep learning algorithms for skin cancer classification from Dermoscopic Images, Bioengineering (Basel, Switzerland). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8945332/

Bhattacharjee, A., Murugan, R., Goel, T. and Mirjalili, S., 2023. Pulmonary Nodule Segmentation Framework Based on Fine-Tuned and Pretrained Deep Neural Network Using CT Images. IEEE Transactions on Radiation and Plasma Medical Sciences, 7(4), pp.394-409.

Brower, C. (2023) Python flask: A powerful web framework for building web applications, Medium. Available at: https://medium.com/@chris.brower.96/python-flask-a-powerful-web-framework-for-building-web-applications-

Cancer Research UK (2023) *Melanoma skin cancer statistics*. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer

Castro-Fernández, M. et al. (2022) 'Towards Skin Cancer Self-Monitoring through an Optimized MobileNet with Coordinate Attention', in 2022 25th Euromicro Conference on Digital System Design (DSD), pp. 607–614. doi: 10.1109/DSD57027.2022.00087.

Choi, D. et al. (2021) 'UAV-Driven Structural Crack Detection and Location Determination Using Convolutional Neural Networks', Sensors, 21, p. 2650. doi: 10.3390/s21082650.

Day, R. (2020) Agile in data science: How can scrum work effectively for your team?, Medium. Available at: https://towardsdatascience.com/agile-in-data-science-how-can-scrum-work-effectively-for-your-team-c567208e9d3f

Dertat, A. (2017) Applied deep learning - part 4: Convolutional Neural Networks, Medium. Available at: https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Emara, T. et al. (2019) 'A Modified Inception-v4 for Imbalanced Skin Cancer Classification Dataset', in 2019 14th International Conference on Computer Engineering and Systems (ICCES), pp. 28–33. doi: 10.1109/ICCES48960.2019.9068110.

Făgărășan, C. et al. (2021) 'Agile, waterfall and iterative approach in information technology projects', IOP Conference Series: Materials Science and Engineering, 1169, p. 012025. doi: 10.1088/1757-899X/1169/1/012025.

Fan, X., Dai, M., Liu, C., Wu, F., Yan, X., Feng, Y., Feng, Y. and Su, B., 2019. Effect of image noise on the classification of skin lesions using deep convolutional neural networks. Tsinghua Science and Technology, 25(3), pp.425-434.

Fu'adah, Y. N. et al. (2020) 'Convolutional Neural Network (CNN) for Automatic Skin Cancer Classification System', IOP Conference Series: Materials Science and Engineering. IOP Publishing, 982(1), p. 012005. doi: 10.1088/1757-899X/982/1/012005.

Han, B. et al. (2022) 'A Triple-Structure Network Model Based upon MobileNet V1 and Multi-Loss Function for Facial Expression Recognition', Symmetry, 14(10). doi: 10.3390/sym14102055.

Harangi, B. (2018) 'Skin lesion classification with ensembles of deep convolutional neural networks', Journal of Biomedical Informatics, 86, pp. 25–32. doi: 10.1016/j.jbi.2018.08.006.

Hassan, S.R., Afroge, S. and Mizan, M.B., 2020, June. Skin lesion classification using densely connected convolutional network. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 750-753). IEEE.

Hijazi, S., Kumar, R. and Rowen, C., 2015. Using convolutional neural networks for image recognition. Cadence Design Systems Inc.: San Jose, CA, USA, 9(1).

Hosny, K. M., Kassem, M. A. and Foaud, M. M. (2018) 'Skin Cancer Classification using Deep Learning and Transfer Learning', in 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), pp. 90–93. doi: 10.1109/CIBEC.2018.8641762.

Huang, G. et al. (2017) 'Densely Connected Convolutional Networks', in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

Iqbal, I., Younus, M., Walayat, K., Kakar, M.U. and Ma, J., 2021. Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. Computerized medical imaging and graphics, 88, p.101843.

Jayalakshmi, G. S. and Kumar, V. S. (2019) 'Performance analysis of Convolutional Neural Network (CNN) based Cancerous Skin Lesion Detection System', in 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), pp. 1–6. doi: 10.1109/ICCIDS.2019.8862143.

Junayed, M.S. et al. (2021) A deep CNN model for skin cancer detection and classification, DSpace at University of West Bohemia: Available at: https://otik.uk.zcu.cz/handle/11025/45011

Kassem, M. A. et al. (2021) 'Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review', *Diagnostics*. MDPI, 11(8), p. 1390.

Ker, J., Wang, L., Rao, J. and Lim, T., 2017. Deep learning applications in medical image analysis. Ieee Access, 6, pp.9375-9389.

Khan, M. A. et al. (2019) 'Multi-Model Deep Neural Network based Features Extraction and Optimal Selection Approach for Skin Lesion Classification', in 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1–7. doi: 10.1109/ICCISci.2019.8716400.

Khan, M. A. et al. (2021) 'Multi-Class Skin Lesion Detection and Classification via Teledermatology', IEEE Journal of Biomedical and Health Informatics, 25(12), pp. 4267–4275. doi: 10.1109/JBHI.2021.3067789.

Kidoh, M., Shinoda, K., Kitajima, M., Isogawa, K., Nambu, M., Uetani, H., Morita, K., Nakaura, T., Tateishi, M., Yamashita, Y. and Yamashita, Y., 2020. Deep learning based noise reduction for brain MR imaging: tests on phantoms and healthy volunteers. Magnetic resonance in medical sciences, 19(3), pp.195-206.

Lee, J.-G. et al. (2017) Deep Learning in Medical Imaging: General overview, Korean Journal of Radiology. Available at: https://synapse.koreamed.org/articles/1027354

Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R. and Feng, D., 2014, April. Early diagnosis of Alzheimer's disease with deep learning. In 2014 IEEE 11th international symposium on biomedical imaging (ISBI) (pp. 1015-1018). IEEE.

Lopez, V.W.M., Abu, P.A.R. and Estuar, M.R.J.E., 2021, November. Real-time Face Mask Detection Using Deep Learning on Embedded Systems. In 2021 3rd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE) (pp. 1-7). IEEE.

Mall, P.K., Singh, P.K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T. and Ganzha, M., 2023. A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. Healthcare Analytics, p.100216.

Mazhar, T. et al. (2023) 'The Role of Machine Learning and Deep Learning Approaches for the Detection of Skin Cancer', Healthcare, 11(3). doi: 10.3390/healthcare11030415.

Mishra, V., Kumar, A. and Arora, M., 2021, May. Deep convolution neural network based automatic multi-class classification of skin cancer from dermoscopic images. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 800-805). IEEE.

Mohan, R., 2014. Deep deconvolutional networks for scene parsing. arXiv preprint arXiv:1411.4101.

Mueller, V. (2021). Deep transfer learning. [online] Medium. Available at: https://towardsdatascience.com/transfer-learning-3e9bb53549f6.

Mukherjee, S. (2022). The Annotated ResNet-50. [online] Medium. Available at: https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758.

Murphree, D. H. and Ngufor, C. (2017) 'Transfer learning for melanoma detection: Participation in ISIC 2017 skin lesion classification challenge', arXiv preprint arXiv:1703. 05235.

Nahata, H. and Singh, S.P., 2020. Deep learning solutions for skin cancer detection and diagnosis. Machine Learning with Health Care Perspective: Machine Learning and Healthcare, pp.159-182.

Narayanamurthy, V. *et al.* (2018) *Skin cancer detection using non-invasive techniques*, *RSC Advances*. Available at: https://pubs.rsc.org/en/content/articlehtml/2018/ra/c8ra04164d

Pacheco, A.G. and Krohling, R.A., 2019. Recent advances in deep learning applied to skin cancer detection. arXiv preprint arXiv:1912.03280.

Phung and Rhee (10 2019) 'A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets', Applied Sciences, 9, p. 4500. doi: 10.3390/app9214500.

Popescu, D., El-Khatib, M. and Ichim, L., 2022. Skin lesion classification using collective intelligence of multiple neural networks. Sensors, 22(12), p.4399.

Radwan, N. (2019) Leveraging Sparse and Dense Features for Reliable State Estimation in Urban Environments. doi: 10.6094/UNIFR/149856.

Rajakumari, R. and Kalaivani, L., 2022. Breast Cancer Detection and Classification Using Deep CNN Techniques. Intelligent Automation & Soft Computing, 32(2).

Ramprasath, M., Anand, M.V. and Hariharan, S., 2018. Image classification using convolutional neural networks. International Journal of Pure and Applied Mathematics, 119(17), pp.1307-1319.

Romero Lopez, A. et al. (2017) 'Skin lesion classification from dermoscopic images using deep learning techniques', in 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), pp. 49–54. doi: 10.2316/P.2017.852-053.

Saber, A., Sakr, M., Abo-Seida, O.M., Keshk, A. and Chen, H., 2021. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Access, 9, pp.71194-71209.

Sae-Lim, W., Wettayaprasit, W. and Aiyarak, P. (2019) 'Convolutional Neural Networks Using MobileNet for Skin Lesion Classification', in 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 242–247. doi: 10.1109/JCSSE.2019.8864155.

Sakib, S. et al. (2020) 'DL-CRC: Deep Learning-Based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach', IEEE Access, 8, pp. 171575–171589. doi: 10.1109/ACCESS.2020.3025010.

Salian, A. C. et al. (2020) 'Skin Lesion Classification using Deep Learning Architectures', in 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), pp. 168–173. doi: 10.1109/CSCITA47329.2020.9137810.

Shetty, B., Fernandes, R., Rodrigues, A.P., Chengoden, R., Bhattacharya, S. and Lakshmanna, K., 2022. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Scientific Reports, 12(1), p.18134.

Srivastava, A., Bhardwaj, S. and Saraswat, S. (2017) 'SCRUM model for agile methodology', in 2017 International Conference on Computing, Communication and Automation (ICCCA), pp. 864–869. doi: 10.1109/CCAA.2017.8229928.

Tschandl, P. (2018) "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." Harvard Dataverse.

Vidya, M. and Karki, M. V. (2020) 'Skin Cancer Detection using Machine Learning Techniques', in 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp. 1–5. doi: 10.1109/CONECCT50063.2020.9198489.

Vyshnavi, V.R. and Malik, A., 2019. Efficient Way of Web Development Using Python and Flask. Int. J. Recent Res. Asp, 6(2), pp.16-19.

Waweru, A. K. et al. (2020) 'Deep Learning in Skin Lesion Analysis Towards Cancer Detection', in 2020 24th International Conference Information Visualisation (IV), pp. 740–745. doi: 10.1109/IV51561.2020.00130.

Wei, Y., Gao, M., Xiao, J., Liu, C., Tian, Y. and He, Y., 2023. Research and Implementation of Traffic Sign Recognition Algorithm Model Based on Machine Learning. Journal of Software Engineering and Applications, 16(6), pp.193-210.

Yani, M., S, M. T. B. I., Si and S.t., M. T. C. S. (2019) 'Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry's Nail', Journal of Physics: Conference Series. IOP Publishing, 1201(1), p. 012052. doi: 10.1088/1742-6596/1201/1/012052.

Zafar, K. et al. (2020) 'Skin Lesion Segmentation from Dermoscopic Images Using Convolutional Neural Network', Sensors, 20(6). doi: 10.3390/s20061601.

Zhang, N. et al. (2020) 'Skin cancer diagnosis based on optimized convolutional neural network', Artificial Intelligence in Medicine, 102, p. 101756. doi: 10.1016/j.artmed.2019.101756.

Zhao, W. and Du, S., 2016. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing, 54(8), pp.4544-4554.

Appendices

- **Exploratory Data Analysis Code**

  https://colab.research.google.com/drive/1xGKkr7PNW-aW-GqBhBLM3RGhNJhkS-COv?usp=drive_link

- **Baseline CNN Model Implementation Code**

  https://colab.research.google.com/drive/1HhK4iBcHdqxFg9JKQDEKy1KGI-jTQBa64?usp=drive_link

- **MobileNet Model and Optimisation Implementation Code**

  https://colab.research.google.com/drive/1kl0MMlpEHguWVTXmdXgsNYqt8vt-iE9T?usp=drive_link

- **VGG16 Model Implementation Code**

  https://colab.research.google.com/drive/1UsR_8q4b0tGER3f_vIcixDBD-foAnyxx5?usp=drive_link

- **DenseNet Model Implementation Code**

  https://colab.research.google.com/drive/1LXmPPm01UDkBvR-4p9m3pcc-JOsAsb8OZ?usp=drive_link

- **InceptionV3 Model Implementation Code**

  https://colab.research.google.com/drive/1ZhpUTQS1pO5HIUYqGgeFsXmXIwBg-zEdO?usp=drive_link

- **ResNet50 Model Implementation Code**

  https://colab.research.google.com/drive/1KRWDfWzNQPY4cfwXaXt6qcbLz0wLp92v?usp=sharing

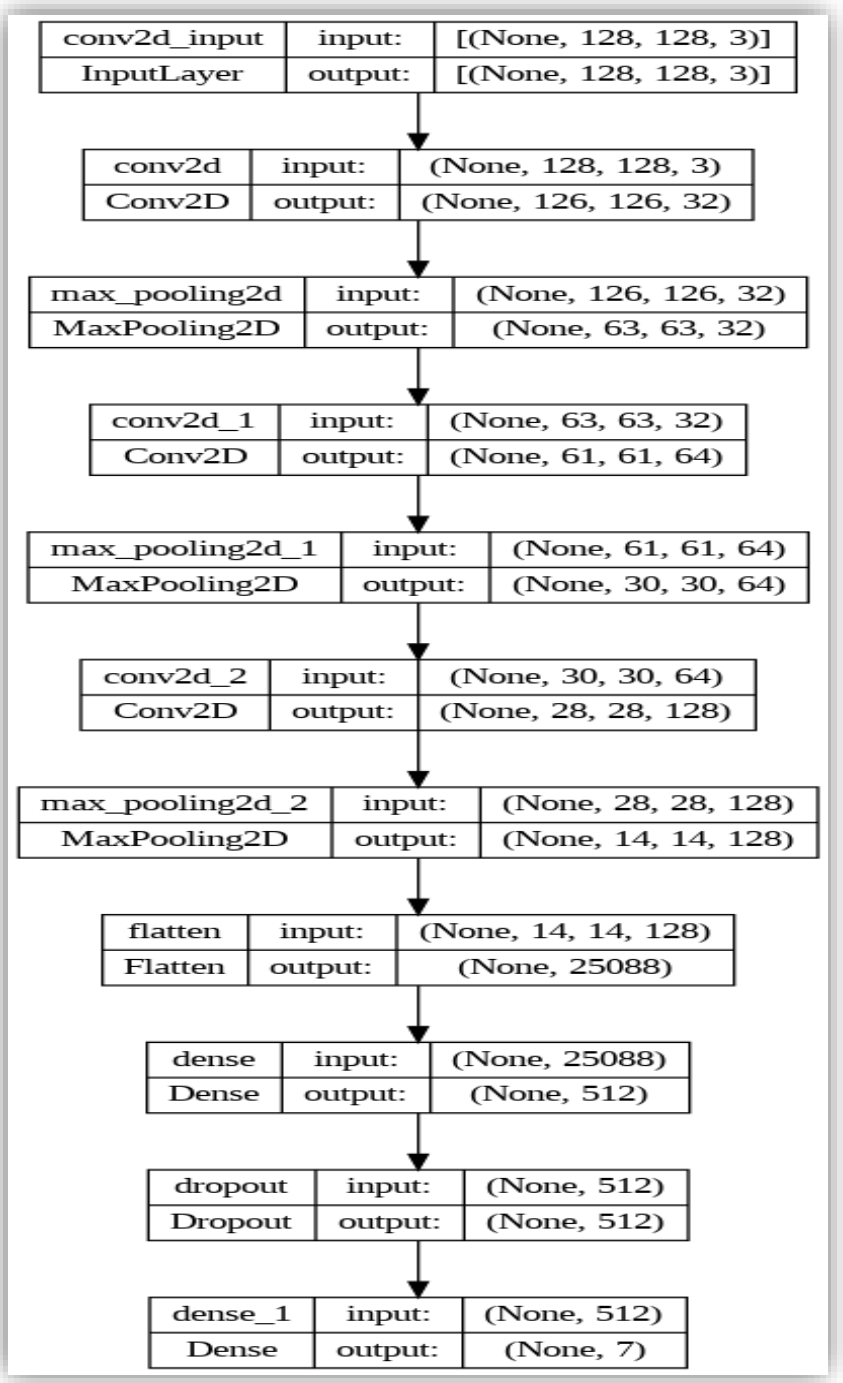
- **Web Application Code**

  https://drive.google.com/drive/folders/1OnjPn68_IvJP8mnDwXkRKv1IXHSYR-unH?usp=drive_link

- **Full Demonstration of Web Application**

  https://drive.google.com/file/d/150yuNO_8FyGs1vsQ2Wo-0hQM-3_3F207/view?usp=drive_link

- **Baseline Model Architecture**



Baseline CNN Architecture