European Journal of Clinical Pharmacy
Print ISSN: 2385-409X **Online ISSN**: 3105-0409
Website: https://farmclin.com/

Original Research Article

Classification of Skin Lesions Images into Distinct Classes Using Deep Learning Techniques

Article History:

Name of Author: Poonam Gaur, Dr. Harsh Sadawarti

Affiliation: *Assistant Professor Department of computer Application Desh Bhagat University, Mandi, Gobindgarh, Punjab*

Corresponding Author:
Poonam Gaur
(bca.poonamgaur@gmail.com)

***Received**:* 10-10-2025
***Revised**:* 28-10-2025
***Accepted**:* 18-11-2025
***Published**:* 30-11-2025

***Abstract**:* In medical image analysis applications, the classification of skin lesion images is challenging due to the complex and diverse nature of skin diseases. Further, the high visual similarity between different lesion types, where benign and malignant lesions usually have not only overlapping colour but also texture and structural patterns, is another major concern. Conventional machine learning approaches depend considerably on manually designed features, which usually fail to represent delicate discriminative patterns in such cases. On the other hand, deep learning models, especially Convolutional Neural Networks (CNNs), have been promising; however, the power of the models is still limited by the computational complexity and generalisation problems. Therefore, it is very important to handle this concern in such a way as to come up with automation of skin lesion classification systems that are not only reliable but also accurate and easy to deploy. In this work, different deep learning techniques, including a pre-trained model and a self-devised CNN model, have been tested on the HAM10000 dataset to provide a suitable solution to this. Results depicted that the proposed CNN model is highly reliable in skin lesion image classifications in terms of computational cost and performance when tested for accuracy, recall, precision and F1 score.

Keywords: Skin Cancer Classification, HAM10000 dataset, pre trained model and proposed CNN model.

INTRODUCTION

Skin diseases constitute a significant global health concern, affecting millions of individuals annually. The increasing incidence of dermatological conditions underscores the need for accurate and timely diagnosis to facilitate effective treatment and management. Conventional diagnostic methods, such as visual examination by dermatologists and dermoscopy, are often subjective and time-consuming, leading to variability in diagnostic accuracy. The integration of artificial intelligence (AI) and deep learning methodologies has emerged as a transformative approach to addressing these challenges by automating skin disease detection and classification [1].

Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized image-based diagnostics by enabling automated feature extraction and classification of medical images. Traditional machine learning approaches often require manual feature selection, which can limit accuracy in complex classification problems. Advanced CNN models have demonstrated superior performance in distinguishing between different skin disease categories, thereby reducing dependency on expert interpretation [2]. Recent studies have explored the potential of CNN-based architectures for automated skin disease detection, with techniques such as Fast R-CNN proving effective in improving classification accuracy and computational efficiency [1].

One of the major challenges in AI-driven dermatological diagnostics is the high similarity among various skin diseases, making it difficult to achieve precise classification. Feature learning techniques have been extensively investigated to enhance the discriminative power of deep learning models, improving their ability to differentiate between visually similar lesions [2]. Additionally, ensemble learning methods have been proposed to combine multiple CNN models, leveraging their complementary strengths to improve overall classification accuracy and robustness [3].

The implementation of image processing techniques and machine learning models has played a critical role in advancing automated skin disease detection. Early methods relied on handcrafted features and conventional classifiers; however, the transition to deep learning has significantly improved the scalability and efficiency of skin disease diagnosis [4]. Hybrid architectures integrating convolutional and recurrent neural networks have further contributed to advancements in classification performance, demonstrating improved accuracy in complex dermatological datasets [5].

In addition to classification, segmentation-based approaches have been developed to refine lesion detection and enhance model interpretability. By applying convolutional neural networks to skin image segmentation, researchers have improved diagnostic precision in both standardized and non-standardized image datasets [7]. Furthermore, the role of external factors such as environmental exposure and oxidative stress in skin disease progression has been explored, emphasizing the importance of preventive measures alongside AI-driven diagnostics [8].

The continuous evolution of AI and deep learning in dermatology presents significant opportunities for enhancing early detection, improving diagnostic accuracy, and facilitating large-scale screening. However, challenges such as dataset bias, ethical considerations, and real-world implementation must be addressed to ensure the reliability and clinical applicability of AI-based diagnostic systems. This research aims to develop a robust CNN model for skin disease classification, leveraging advanced transfer learning techniques and performance optimization strategies to contribute to the field of AI-driven medical diagnostics.

2. Related work

Previously various types of deep models have been implemented in skin cancer classification. The transfer learning model and the self-devised model are most dominant in this. Transfer learning has emerged as a pivotal approach in dermatological diagnostics, enabling improved accuracy in skin disease classification. **(Jain *et al.*, 2024)** explored the potential of AI-powered dermatology, specifically in early skin cancer detection. Their study highlighted how pre-trained convolutional neural networks (CNNs) such as ResNet and InceptionV3 outperform traditional diagnostic methods in identifying malignant lesions. **(Aboulmira *et al.*, 2024)** conducted a comparative analysis of various CNN architectures to determine their efficacy in skin disease classification. Their study systematically evaluated models such as VGG16, DenseNet, and MobileNet, highlighting their strengths and limitations in dermatological applications. **(Babu *et al.*, 2024)** introduced a hybrid approach combining U-Net and CNN architectures to unify segmentation and classification tasks in skin disease diagnosis. Their method utilized U-Net for precise lesion segmentation, followed by CNN-based classification to identify specific skin conditions. The study demonstrated that segmenting the affected regions before classification improved model interpretability and reduced false positives. **(Venkat *et al.*, 2024)** investigated deep learning strategies for multi-class skin disease classification, comparing traditional CNN models with newer architectures such as Vision Transformers (ViTs). Their study found that ViTs outperformed CNNs in capturing long-range dependencies and contextual information within dermatological images. The authors demonstrated that self-attention mechanisms within ViTs provided a more holistic understanding of skin lesions, leading to improved diagnostic accuracy. However, they noted that ViTs require significantly more training data and computational resources than conventional CNNs. **(Vayadande , 2024)** conducted a comprehensive study on innovative machine learning approaches for skin disease identification, emphasizing the role of transfer learning in improving diagnostic accuracy. The study compared standard CNN architectures with advanced transfer learning models such as EfficientNet and NasNet, demonstrating that these models achieve higher accuracy with fewer parameters. The research indicated that transfer learning allows models to leverage knowledge from large-scale medical datasets, enabling better feature extraction and reducing overfitting in small dermatological datasets. **(Jui *et al.*, 2022)** presented a CNN-based approach leveraging transfer learning for skin cancer classification. Their study utilized pre-trained CNN models, including ResNet50 and VGG16, to classify various skin cancer types with high accuracy. The model was trained on a publicly available dataset and fine-tuned using data augmentation techniques to enhance generalization. Experimental results demonstrated that transfer learning significantly improved classification performance, achieving an accuracy exceeding 90%. **(Aboulmira *et al.*, 2022)** conducted a comparative study on multiple CNN models for the classification of 23 distinct skin diseases. They explored the performance of state-of-the-art architectures, including InceptionV3, DenseNet201, and Xception, in distinguishing between various dermatological conditions. The study utilized a large dataset comprising high-resolution images, with models fine-tuned using transfer learning. **(Hridoy *et al.*, 2021)** proposed a skin disorder recognition system using EfficientNet, a transfer learning-based deep learning model. Their study aimed to classify multiple skin conditions by leveraging the advanced feature extraction capabilities of EfficientNet. The researchers trained the model on a diverse dataset, employing augmentation strategies to enhance robustness. Results showed that EfficientNet outperformed conventional CNN architectures in terms of accuracy and computational efficiency. **(Srinivasu *et al.,* 2021)** introduced a hybrid deep learning approach combining MobileNet V2 and LSTM for skin disease classification. Their model leveraged MobileNet V2 for spatial feature extraction, while LSTM captured temporal dependencies within image sequences. This dual-network strategy enabled improved classification accuracy, particularly for complex skin conditions requiring sequential analysis. The study employed transfer learning to fine-tune MobileNet V2, significantly reducing computational costs while maintaining high diagnostic accuracy. Experimental results demonstrated the model's superiority over traditional CNNs, achieving notable improvements in sensitivity and specificity. The authors emphasized the importance of lightweight architectures for mobile health applications, highlighting the potential for real-time, accessible skin disease diagnostics.

Similar to transfer learning model, self-devised model has been proposed i,e **(Singla *et al.,* 2024)** explored the effectiveness of deep learning models in dermatological diagnostics, emphasizing precise classification of skin cancer. The study introduced cutting-edge advancements, leveraging convolutional neural networks (CNNs) trained on extensive datasets to improve accuracy. The authors demonstrated that deep learning architectures significantly outperform traditional diagnostic methods, reducing false positives and negatives. **(Alayba *et al.,* 2024)** focused on improving early disease detection using hybrid models that integrate multi-CNN architectures with handcrafted feature extraction. While primarily centered on Alzheimer's disease, the study's methodology is applicable to dermatological AI models. The authors proposed a feature fusion approach that combines deep learning-based feature extraction with traditional statistical techniques, enhancing model accuracy and robustness. This approach allows the system to capture both high-level patterns from CNNs and low-level texture details from handcrafted features. **(Ansari *et al.,* 2024)** proposed a hybrid CNN-SVM model to enhance the accuracy and efficiency of skin disease classification. The study explored the limitations of deep learning models in handling imbalanced datasets and introduced support vector machines (SVM) as a complementary classification technique. **(Anggriandi *et al.*, 2023**) conducted a comparative analysis of CNN and CNN-SVM models for classifying different types of human skin diseases. The study assessed the strengths and weaknesses of each approach, demonstrating that CNN-SVM models achieved higher accuracy due to their ability to refine classification boundaries. **(Vellela *et al.*, 2023)** explored multi-class skin disease classification using CNNs with color and texture feature extraction techniques. The study emphasized the importance of integrating domain-specific feature engineering to enhance classification accuracy. The authors demonstrated that combining deep learning with handcrafted features improved the model's ability to differentiate between visually similar skin conditions. **(Akshay *et al.,* 2023)** developed a mobile-based automated skin disease detection application called Skin-Vision. The study focused on creating an accessible AI-powered diagnostic tool for dermatological conditions, integrating deep learning models optimized for real-time performance. The authors emphasized the importance of user-friendly interfaces and low-latency processing for practical applications. Their research demonstrated that mobile AI models could achieve high accuracy while operating on limited computational resources. The study evaluated different CNN architectures and found that lightweight models such as MobileNet performed well in mobile environments. The authors also discussed the potential of telemedicine integration to expand access to dermatological diagnostics.

RESEARCH METHODOLOGY

The primary objective of skin disease detection is to improve the classification among various skin lesions, as they come in so many varieties and are so similar to one another. Deep learning techniques have evolved as useful methods for handling such challenging classification issues. Since deep learning is a multiple-layer architecture model that can be trained on large databases semantically at a high level, it has been advocated in a number of medical applications in the past. Firstly the present work has uses transfer learning to classify skin diseases, which include four models i.e. DenseNet121, ResNet50, ResNet18, and VGG16. Various optimisers, epochs, and batch sizes are used for the analysis. Several performance metrics, including accuracy, precision, sensitivity, and F1 score, are used to evaluate the models. Further for enhancing the performance in skin diseases classification of self-devised CNN model is employed. The proposed CNN model is self-designed and has included dense layers, max-pool layers, flatten layers, and different convolutional blocks. The best model is evaluated for its performance on the skin diseases dataset. The proposed methodology is shown in figure 1.-

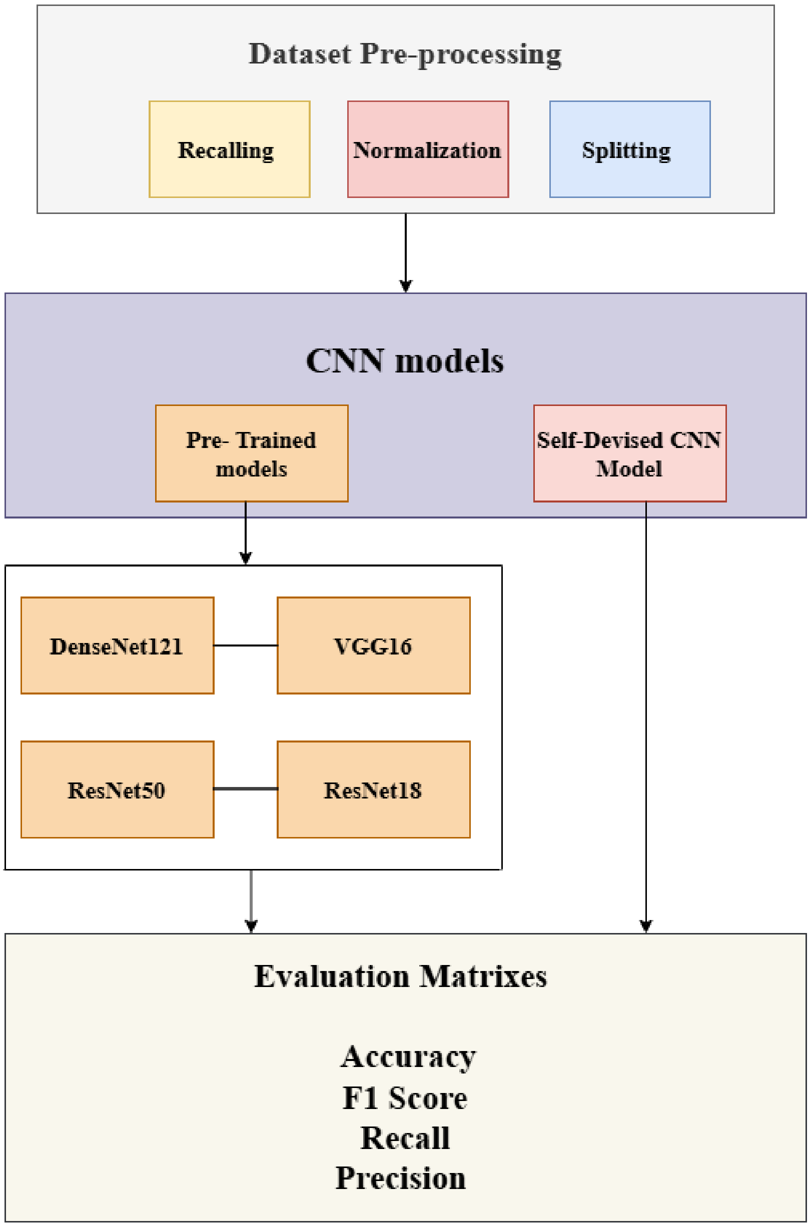

Figure 1: Research Methodology

3.1 Dataset

The HAM10000 dataset, which contains 10,015 dermatoscopy images, is used in this study. Seven different groups of skin diseases are covered by the dataset. There are enough samples in each class for a thorough analysis. The image have a suitable pixel resolution and are of excellent quality. This guarantees comprehensive visual data for training the model. The dataset is extensively utilized in AI applications and dermatology research. It is appropriate for activities involving the classification of skin diseases because to its volume and diversity. The sameple images of dataset is given below.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

Figure 2: Sample images from dataset (a) Actinic Keratoses (b) Benign Keratosis (c) Dermatofibromas (d) Melanocytic Nevus (e) Basal Cell Carcinoma (f) Melanoma (g) Vascular Lesions

3.2 Data pre-processing

3.2.1 Dataset rescaling

In object recognition projects, dataset rescaling is crucial for getting data ready for machine learning algorithms, especially when working with remote sensing images. The images were rescaled to 400*400 from their original dimensions, which varied, such as 600 × 450. The performance of the model is improved and consistency is preserved by standardising the input by rescaling images of different sizes to a predetermined size (for example, 400x400 pixels). Lastly, normalising pixel values—that is, converting them to a range of [0, 1] or [-1, 1]—is a common technique, the specifics of which are provided in the next section. This normalisation is essential to ensure that each pixel value contributes evenly during model training. By preventing any one characteristic from taking centre stage during training, its broader numerical range also aids in the model's convergence.

3.2.2 Normalization

Normalisation is a crucial stage in deep learning, particularly when applying Convolutional Neural Networks (CNNs) to image processing tasks like object detection. The primary objective of normalisation is to ensure that the gradients behave correctly during training and to accelerate convergence, which increases the numerical stability of the model. A critical pre-processing step for training a CNN on image data is normalising pixel values to a standard range, often [0, 1], especially for tasks involving high-dimensional inputs like remote sensing images.

The skin diseases images are usually in RGB (Red, Green, Blue) format. Each pixel in these images has an intensity value that matches its brightness or hue. In RGB images, the Red, Green, and Blue colour channels typically have values between 0 and 255.). Neural networks, on the other hand, typically perform better when input data falls within a smaller numerical range, such as [0, 1] or [-1, 1].

3.2.3 Data augmentation

When working with remote sensing photos, image augmentation techniques can significantly enhance a Convolutional Neural Network's (CNN) performance and generalisation capabilities in tasks such as object detection. With the help of the techniques offered, such as a rotation range of 15, which allows random rotations between -15° and 15°, the model learns to recognise objects from diverse orientations, which is useful in distant sensing applications where objects may appear at different angles.

The existing data was supplemented using a variety of methods, including rotation, shifting, shearing, and zooming. 1./255 normalises pixel values, scaling them between 0 and 1, ensuring faster convergence during training and providing numerical stability and improving the model's efficiency. The shear range: 0.2 adds a shear transformation that slightly distorts images to represent different perspectives, assisting the model in becoming invariant to minor distortions. Zoom_range: 0.2 randomly zooms in or out of images to let the model generalise to objects of varying scales, which is an important feature when identifying objects at different distances in satellite shots. Horizontal_flip adds randomisation by flipping images horizontally, making the model invariant to left-right orientation.

3.3 Deep Learning Techniques

Various deep learning techniques, including transfer learning and self-devised models, have been implemented on the dataset. The detailed description of the model is given below.

3.3.1 DenseNet121

Three transition layers and 4 densely coupled convolution blocks make up the DenseNet121 architecture. It has 80 lakh parameters spread across 121 levels. The convolution layer, which is the initial layer, has 64 filters with a dimension of 7 by 7. The max pool layer is the second layer. The first convolution layer, measuring 1 * 1, and the second convolution layer, measuring 3 * 3, make up the four convolution blocks. As seen in fig. 3.4, a transition layer consisting of an average pool layer and a convolution layer is present after convolution blocks. The final phase involves connecting the completely linked layer and doing the prediction of the skin disease class after all the blocks of convolution and transition layers.

3.3.2 VGG16

OxfordNet, also known as VGG16, is an architectural that bears the name of the Oxford Visual Geometry Group. There are sixteen levels to it. On other datasets outside of ImageNet, VGG also performs better than baselines. Figure 3.5 illustrates the five convolution blocks that make up the VGG16 architecture. There are two convolution layers in the first and second convolution blocks. Three convolution layers make up the third, fourth, and fifth convolution blocks. Each convolution block is followed by a pooling layer. Following the completion of all convolution blocks, the final phase involves connecting a completely linked layer and performing a skin disease class prediction.

3.3.3 ResNet50

A residual neural network with 50 layers is called ResNet50 . Additionally, it can serve as the foundational neural network for several computer vision applications. As seen in fig. 3.6, it is composed of four convolution blocks, referred to as Stages 1, 2, 3, and 4. Three convolutional layers with sizes of (1 * 1), (3 * 3) and (1 * 1) and varying numbers of filters are included in each convolution block. Following the completion of all convolution blocks, the final phase involves connecting a completely linked layer and performing a skin disease class prediction.

3.3.4 ResNet18

The four convolutional blocks that make up the ResNet18 model are referred to as Stages 1, 2, 3, and 4. As seen in fig. 6, each convolution block consists of two convolutional layers, each with a size of (3 * 3) and a distinct number of filters. Following the completion of all convolution blocks, the final phase involves connecting a completely linked layer and performing a skin disease class prediction.

While there are many advantages to using pre-trained models, including time savings and usefulness on small datasets, they may have significant problems when used to solve different problems in the same domain where they were trained. These problems include negative transfer and inaccuracy in identifying decision boundaries among multiple classes in the target domain dataset. Because of this, it is not always be appropriate for real-time applications like skin diseases classification. Therefore, for effective performance in such skin diseases classification applications, it is advised to construct a custom convolution neural network whose learning would be initialized from scratch.

3.3.5 Architecture of CNN model

Architecture of CNN model is presented in figure 1.The input image for this Convolutional Neural Network (CNN) model is 3×128×128, where "3" stands for the RGB colour channels. The model can identify low-level characteristics like edges and corners thanks to a Conv2D layer that processes three input channels and applies sixteen 3x3 filters before a ReLU activation. After that, a 2x2 MaxPooling layer lowers the spatial dimensions, increasing the model's efficiency while preserving crucial data. In order to capture more intricate patterns, the subsequent layer is another Conv2D layer that uses 3×3 filters with ReLU activation to enhance the depth from 16 to 32 channels.

After that, a second Conv2D layer deepens the feature extraction while maintaining a channel size of 32. A second MaxPooling layer with a 2x2 window then further decreases the spatial resolution. In order to allow the model to identify high-level features, the network then applies a third Conv2D layer, increasing the depth from 32 to 64 channels once more using 3×3 filters and ReLU. Before being fed into the dense (completely connected) layers for classification, the 3D output tensor is flattened into a 1D vector.

A ReLU activation occurs after the flattened feature vector is reduced to 64 neurones in the first dense layer (FC1). It is further reduced to 32 neurones by ReLU in the second dense layer (FC2). The final classification logits are then produced by the third dense layer (FC3), which produces a number of neurones equal to the number of classes in the dataset. The model's convolutional and fully connected layers learn spatial hierarchies and abstract characteristics, allowing it to efficiently convert raw image pixels into meaningful class predictions.

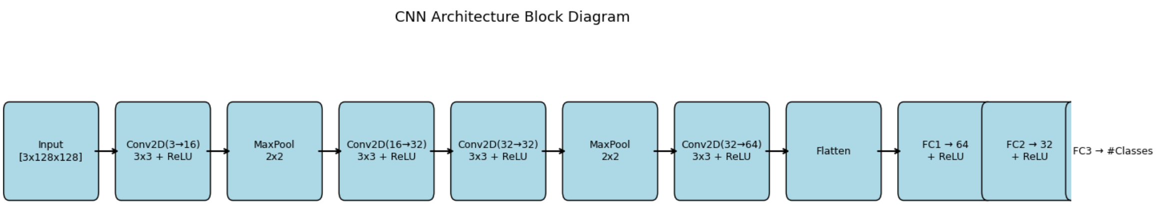

Figure 3: Proposed CNN model

Table 1: Layers' description of CNN model

Layer	Description	Output Shape
Input	RGB Image	(3, 64, 64)
Conv2D 1	16 filters, 3x3, padding=1 + ReLU	(16, 64, 64)
MaxPool 1	2x2 pooling	(16, 32, 32)
Conv2D 2	32 filters, 3x3, valid padding + ReLU	(32, 30, 30)
Conv2D 3	32 filters, 3x3, padding=1 + ReLU	(32, 30, 30)
MaxPool 2	2x2 pooling (default padding=0)	(32, 15, 15)
Conv2D 4	64 filters, 3x3 + ReLU	(64, 13, 13)
Flatten	Flatten all values	64×13×13 = 10816
FC1	Fully connected (10816→64) + ReLU	64
FC2	Fully connected (64→32) + ReLU	32
FC3 (Output)	Fully connected (32→num_classes), no activation	num_classes
Conv2D 1	16 filters, 3x3, padding=1 + ReLU	(16, 64, 64)

As discussed earlier, there are a number of deep learning optimizers like SGD, Adagrad, RMSProp, Adadelta, Adam and Adamax. Out of these deep learning optimizers, the Adam optimizer gives the best performance in classification problems, as per research reports. So, for the current classification problem of identifying the images of eight skin disorders, the researchers have proposed the Adam optimizer. The complete specification of the Adam optimizer in training the proposed CNN model is given below.

Table 2: Optimizer's parameters

Name of the Optimizers	Specification
Adam	Rate of learning = 0.001 ,beta$_1$ = 0.99,beta$_2$ = 0.999

	,epsilon = 1 x 10-8

The various other parameters of the model has been given in table 3

Table 3: Training parameters of Model

Parameters	Value
Input Shape	128*128
Dataset Division	80:20
Model	Sequential Model
Convolutional Layers	Variable
Activation Function for Hidden Layers	Sigmoid
Activation Function for Convolutional Layers Layer	Relu
Loss Function	Categorical cross entropy
Optimizer	Adam
Metrics	Accuracy
Batch size	32
Epochs	50

RESULTS

The results and detailed analysis of the suggested work are presented in this chapter. The primary objective of this chapter is to evaluate the performance of several models in detecting skin cancer detection through simulation using a dataset of HAM10000 of Dermoscopy images.

4.1 Results of ResNet 50

In this section ResNet 50's results in term of accuracy and loss graph, confusion matrix and classification report has been presented when tested on the given dataset.

4.1.1 Accuracy and loss curve of ResNet 50

The ratio of predicted occurrences or samples to all instance or samples is what defines accuracy. The accuracy is defined in equation (4.1)

Accuracy = TP+TN/TP+TN+FN+FP (4.1)

An accuracy analysis of ResNet 50 is presented below in figure 4

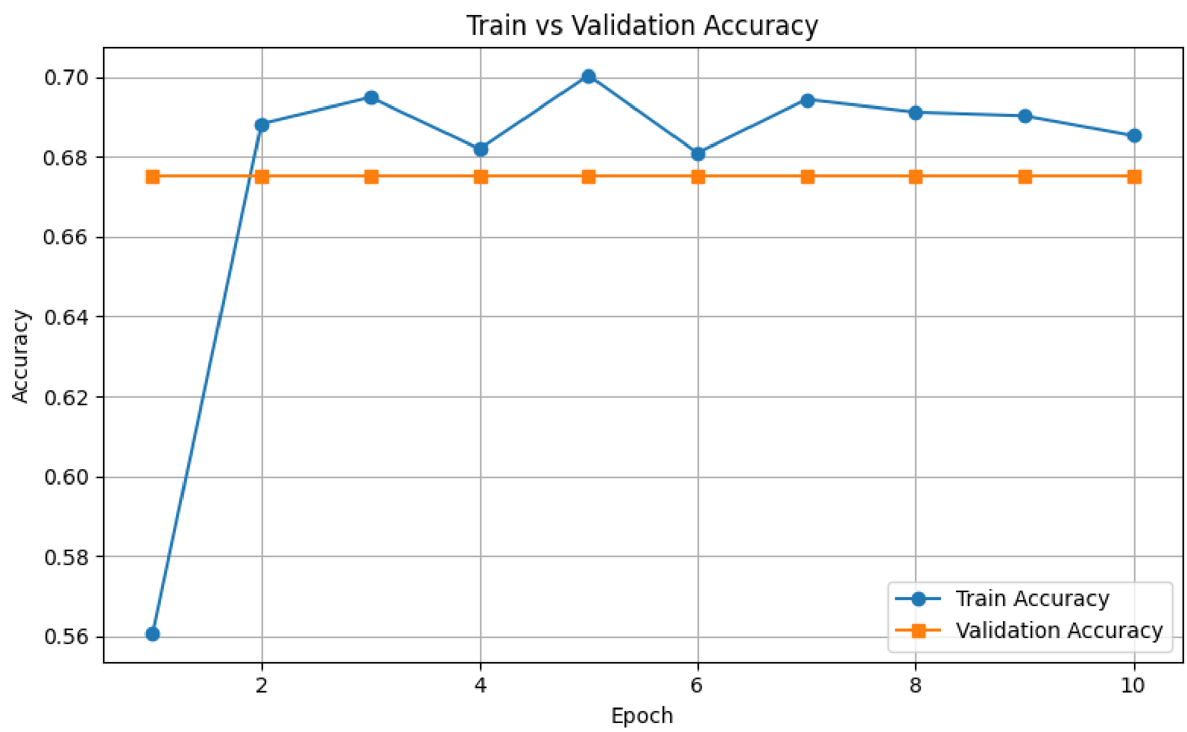

Figure 4: Accuracy graph for ResNet50

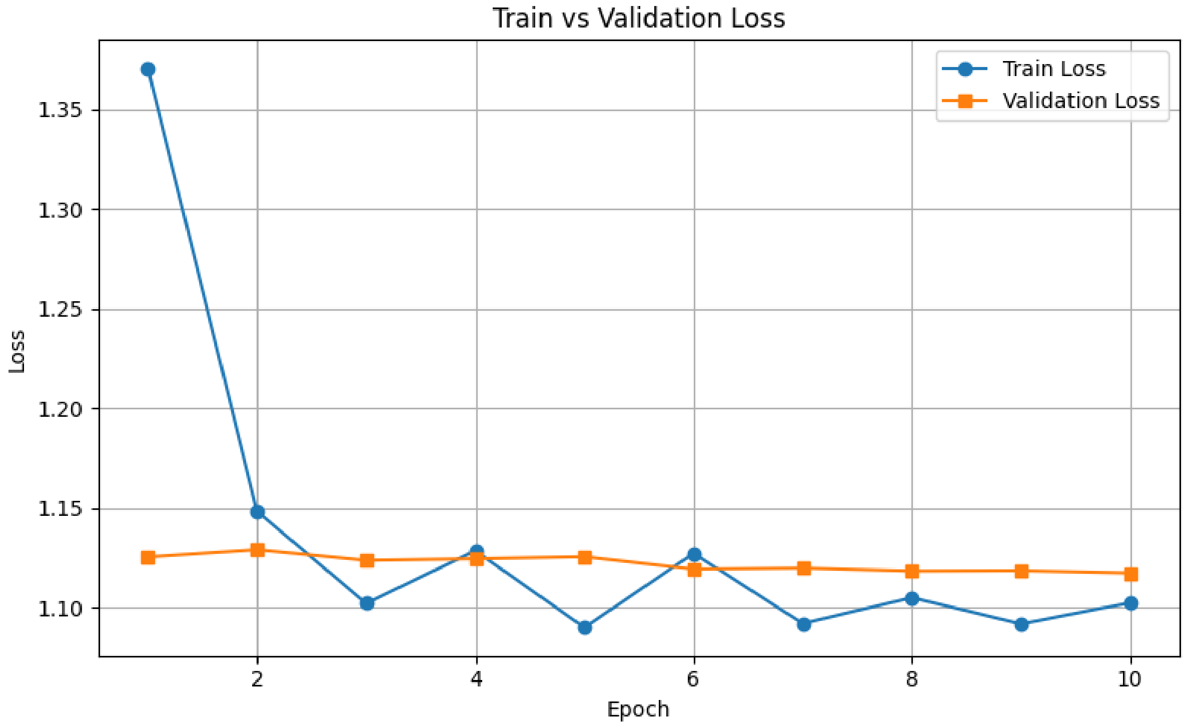

Figure 5: Loss graph for ResNet50

According to the ResNet50 training data, the model's accuracy increased overall from 56.06% in the first epoch to roughly 69% by the third, with very slight variations thereafter. Notwithstanding this early improvement, subsequent epochs show a plateau in training accuracy in the 68–70% range, indicating little room for additional learning. The accuracy gain is largely supported by the training loss, which exhibits a declining trend from 1.37 to roughly 1.10. All epochs show the same validation accuracy of 67.5%. From 1.1253 to 1.1172, the validation loss exhibits a very slight downward trend, but the changes are insignificant, demonstrating the model's poor capacity for generalisation.

4.1.2 Confusion Matrix and classification results of ResNet 50

The HAM10000 dataset's classification report for ResNet50 shows a notable disparity in model performance between classes. With 686 samples, the "nv" (melanocytic nevi) class dominates the dataset, and the model obtains a high recall of 1.00 and an F1-score of 0.81 for this class. The precision, recall, and F1-score for each of the remaining classes—akiec, bcc, bkl, df, mel, and vasc—are all 0.00, though, suggesting that the model was unable to accurately identify any instances of those classes. Because the over-represented "nv" class was correctly predicted, the overall accuracy of 69% is misleading.

While the weighted average (F1-score: 0.56) is skewed by the huge "nv" class, the macro average metrics (precision: 0.10, recall: 0.14, F1-score: 0.12) show the low performance across minority classes.

Table 4: Classification report for ResNet 50

	precision	recall	f1-score	support
akiec	0.00	0.00	0.00	37
bcc	0.00	0.00	0.00	53
bkl	0.00	0.00	0.00	113
df	0.00	0.00	0.00	11
mel	0.00	0.00	0.00	87
nv	0.69	1.00	0.81	686
vasc	0.00	0.00	0.00	13
accuracy			0.69	1000
macro avg	0.10	0.14	0.12	1000
weighted avg	0.47	0.69	0.56	1000

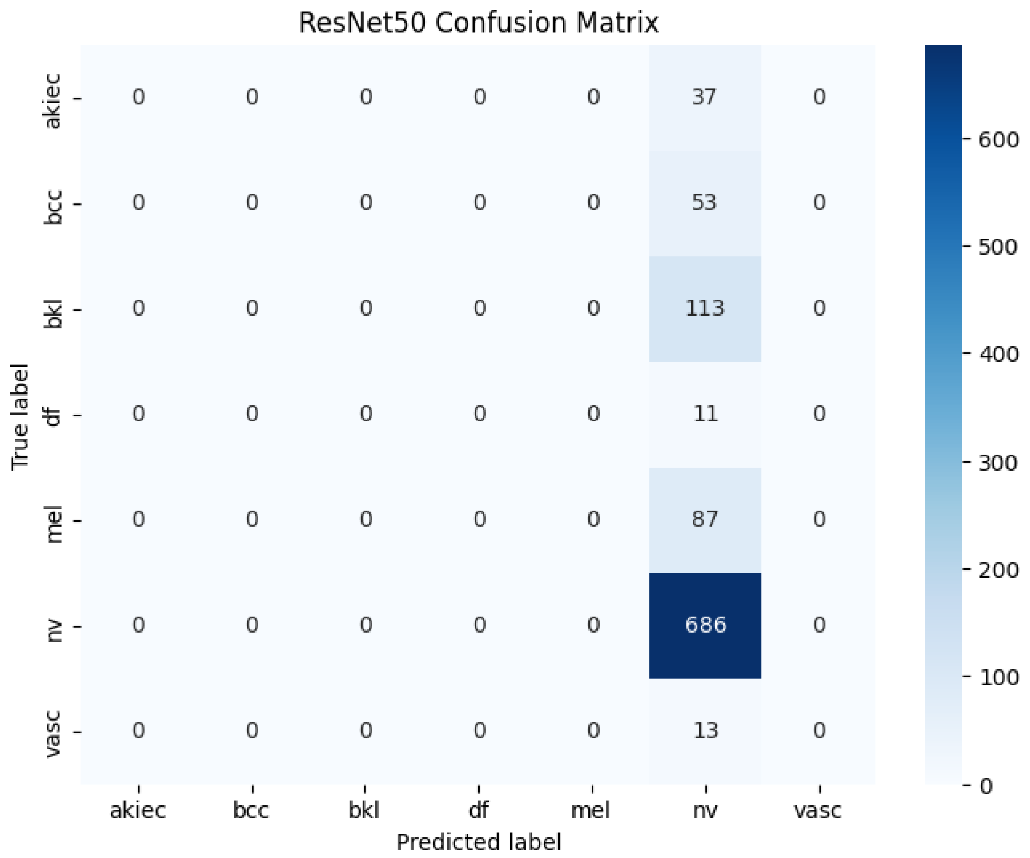

Figure 6: CM for ResNet50

4.2 Results of DenseNet121

In this section DenseNet121's results in term of accuracy and loss graph, confusion matrix and classification report has been presented when tested on the given dataset.

4.2.1 Accuracy and loss curve of DenseNet121

When assessing a deep learning model's performance during training and validation, the accuracy and loss curves are crucial instruments. These graphs show how well the model learns over time in the example of DenseNet121, a potent convolutional neural network renowned for its dense connectivity and effective feature reuse. The validation curves illustrate how well the model generalises to new data, while the training accuracy and loss curves demonstrate how well the model fits the training data. An accuracy analysis of DenseNet121 is presented below in figure 4.4

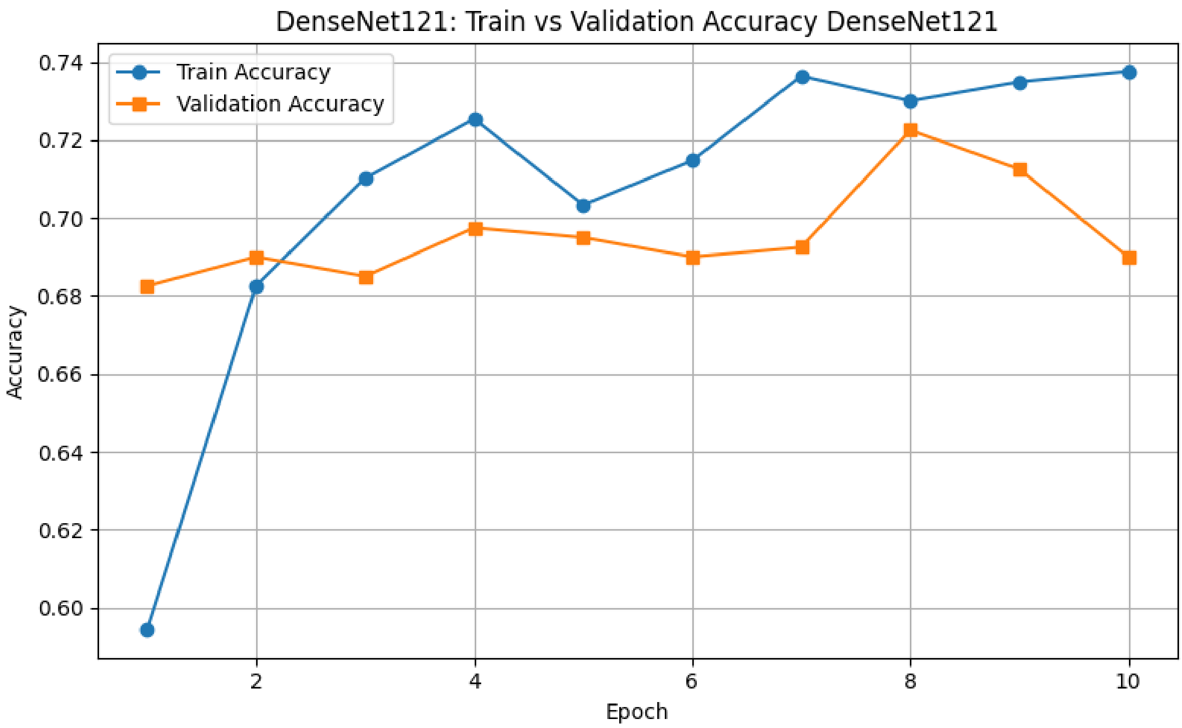

Figure 7: Accuracy graph for DenseNet121

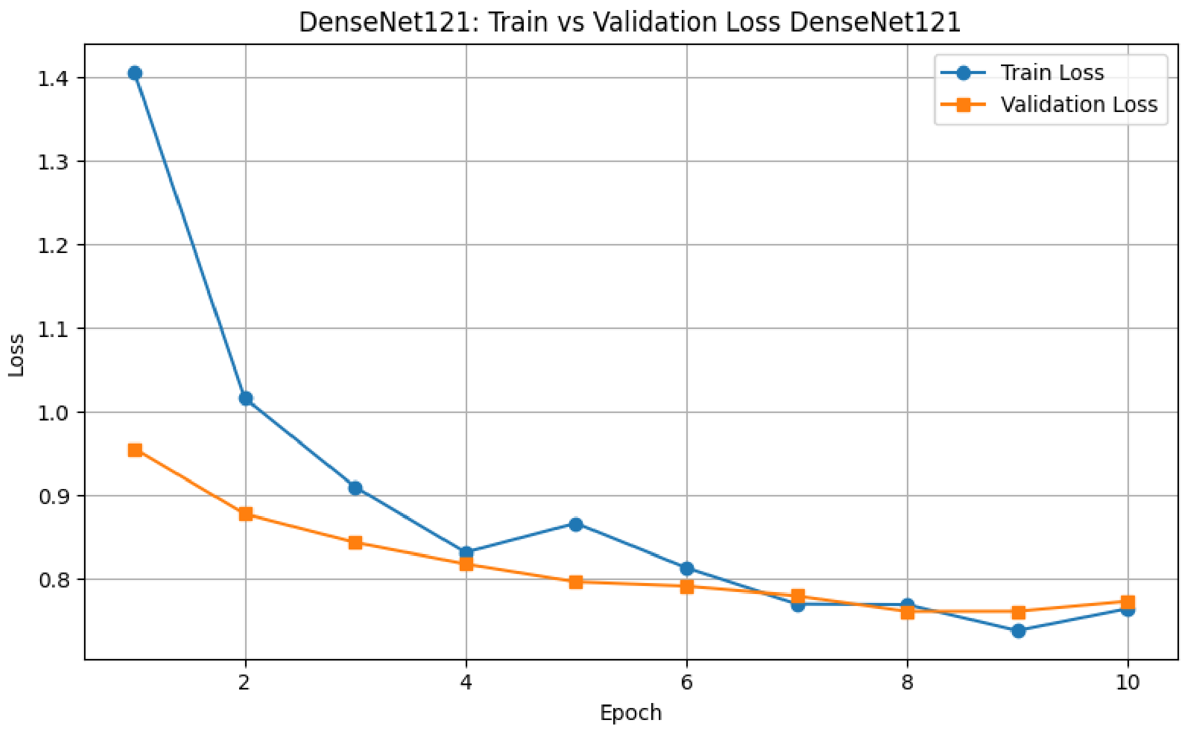

Figure 8: Loss graph for DenseNet121

Effective learning is suggested by the performance study of DenseNet121 over 10 epochs, which shows a steady increase in training accuracy and a corresponding drop in training loss. The model effectively captured significant patterns in the data, as evidenced by the training accuracy, which began at 59.43% in epoch 1 and progressively improved to 73.76% by epoch 10, while the training loss decreased from 1.4061 to 0.7646. However, the validation accuracy fluctuated moderately, starting at 68.25%, reaching a peak of 72.25% in epoch 8, and then declining slightly to 69.00% in the last epoch.

Additionally, the validation loss steadily dropped from 0.9558 to roughly 0.7613.. All things considered, DenseNet121 exhibits a high learning capacity and encouraging outcomes in this training configuration.

4.2.2 Confusion Matrix and classification results of DenseNet121

The model's outstanding performance on the dominating class 'nv', which attained a precision of 0.74, recall of 0.99, and F1-score of 0.85, is the main driver of the DenseNet121 classification performance, which is represented by the metrics and demonstrates a strong overall accuracy of 72%. This suggests that the model predicts this class with a high degree of accuracy and confidence. With F1-scores of 0.07, 0.23, and 0.21 for the minority classes "bcc," "akiec," and "mel," respectively, performance is still much worse. Interestingly, 'df' had zero F1-score, precision, and recall, demonstrating the model's total incapacity to identify this class.

The disparity in predictive skill between classes is demonstrated by the macro average F1-score of 0.31 and recall of 0.27. In contrast, the weighted average F1-score of 0.65 indicates that the model favours the more frequent class and performs better overall as a result of the class imbalance.

Table 5: Classification report for DenseNet121

	precision	recall	f1-score	support
akiec	0.38	0.16	0.23	37
bcc	0.25	0.04	0.07	53
bkl	0.63	0.17	0.27	113
df	0.00	0.00	0.00	11
mel	0.48	0.14	0.21	87
nv	0.74	0.99	0.85	686
vasc	1.00	0.38	0.56	13
accuracy			0.72	1000
macro avg	0.50	0.27	0.31	1000
weighted avg	0.66	0.72	0.65	1000

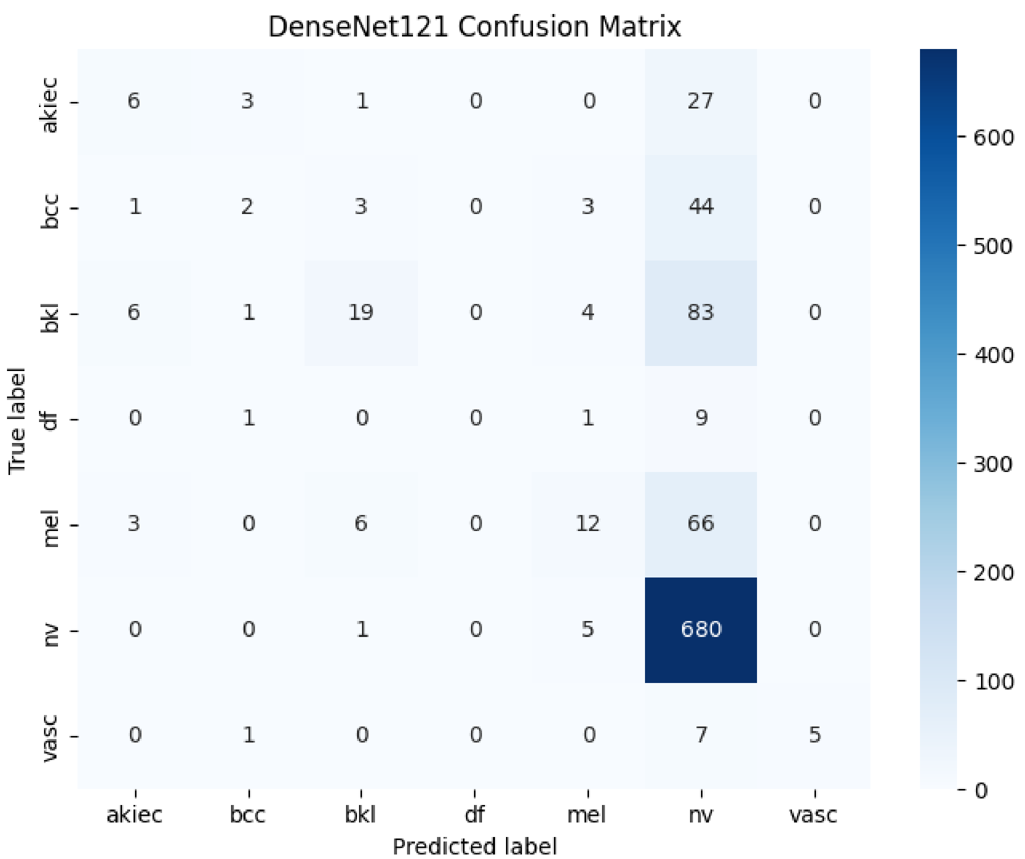

Figure 9: CM for DenseNet121

4.3 Results of VGG16

In this section VGG16's results in term of accuracy and loss graph, confusion matrix and classification report has been presented when tested on the given dataset.

4.3.1 Accuracy and loss curve of VGG16

The VGG16 model's accuracy and loss curves offer important information about how well the model performs during training and validation across time. These curves aid in determining how well the model is generalising to new data and learning features from the dataset. Effective learning is typically shown by a lowering loss curve and a continually rising accuracy curve. By keeping an eye on these patterns, we can assess if VGG16 is overfitting, underfitting, or attaining balanced learning. A clear picture of model behaviour and training stability may be obtained by comparing training and validation metrics side by side across these curves. An accuracy analysis of VGG16 is presented below in figure 10

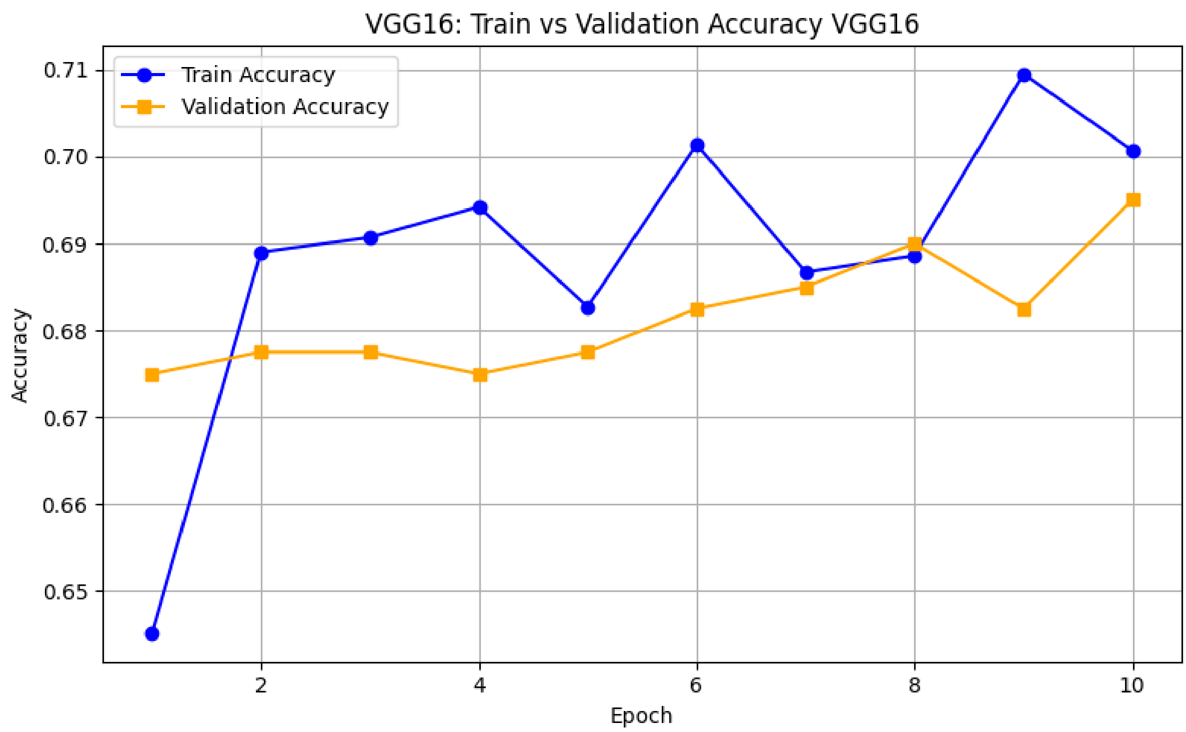

Figure 10 Accuracy graph for VGG16

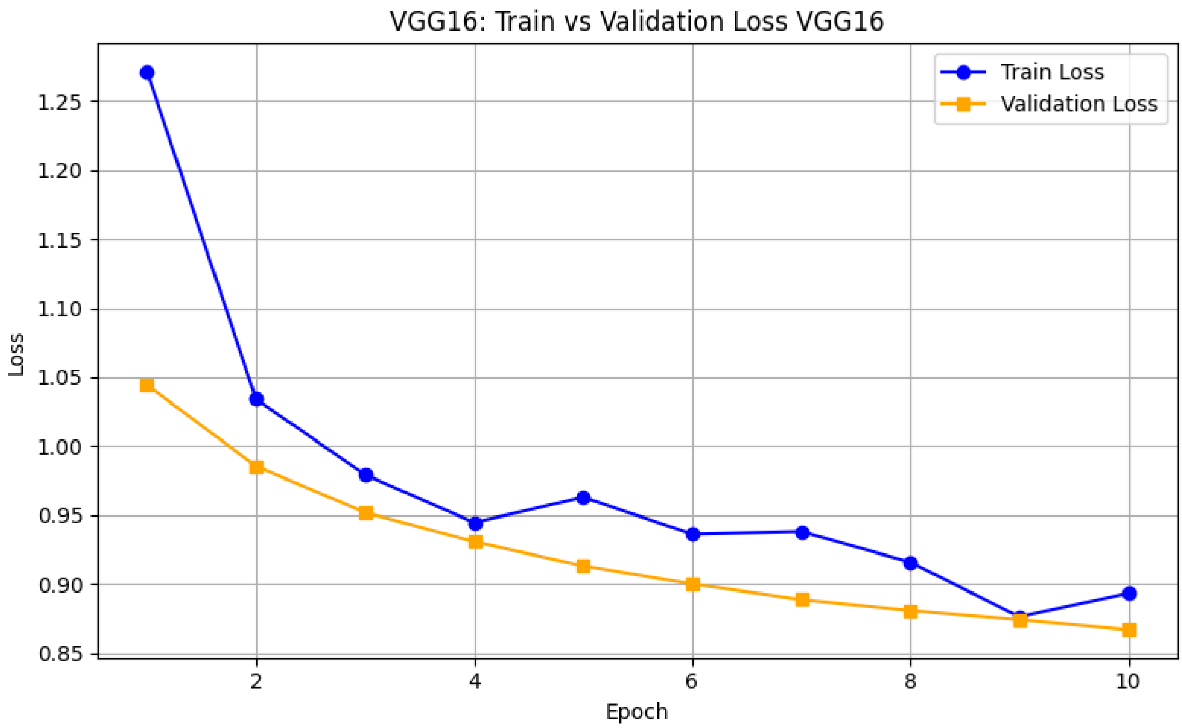

Figure 11: Loss graph for VGG16

The VGG16 model's training and validation results across ten epochs show a steady learning trend. The model is successfully learning from the training data, as evidenced by the training accuracy, which begins at 64.51% and gradually increases to about 70.07%. The training loss correspondingly drops from 1.27 to 0.89, demonstrating the model's growing confidence and decreasing prediction error with time. Accuracy on the validation side exhibits a little but steady increase trend, starting at 67.50% and rising to 69.50% by the tenth epoch. Additionally, validation loss continuously decreases from 1.04 to 0.86.

The results suggest that VGG16 is still useful for this job and retains training stability across epochs, despite being relatively deep. The accuracy and loss curves' smooth convergence provides additional evidence that the learning rate and optimisation configuration are ideal.

4.3.2 Confusion Matrix and classification results of VGG16

The VGG16 model's classification performance shows both its advantages and disadvantages when dealing with unbalanced multi-class data. The model performs well on the majority class 'nv', achieving a precision of 0.73, recall of 0.98, and F1-score of 0.84, demonstrating its excellent effectiveness in accurately recognising this class. Its overall accuracy is 71%. However, precision and recall are much poorer for minority classes like "akiec," "bcc," "bkl," and "mel," with the majority of F1-scores falling below 0.30. For "df" and "vasc," the model entirely fails, scoring 0.00 on all measures.

The macro average F1-score of 0.23 and recall of 0.21 underscore this issue. Nevertheless, the weighted average F1-score of 0.63 reflects reasonable overall performance due to the dominance of correctly classified majority samples.

Table 6: Classification report for VGG16

	precision	recall	f1-score	support
akiec	0.67	0.11	0.19	37
bcc	0.50	0.09	0.16	53
bkl	0.40	0.09	0.14	113
df	0.00	0.00	0.00	11
mel	0.44	0.20	0.27	87
nv	0.73	0.98	0.84	686
vasc	0.00	0.00	0.00	13

accuracy			0.71	1000
macro avg	0.39	0.21	0.23	1000
weighted avg	0.64	0.71	0.63	1000

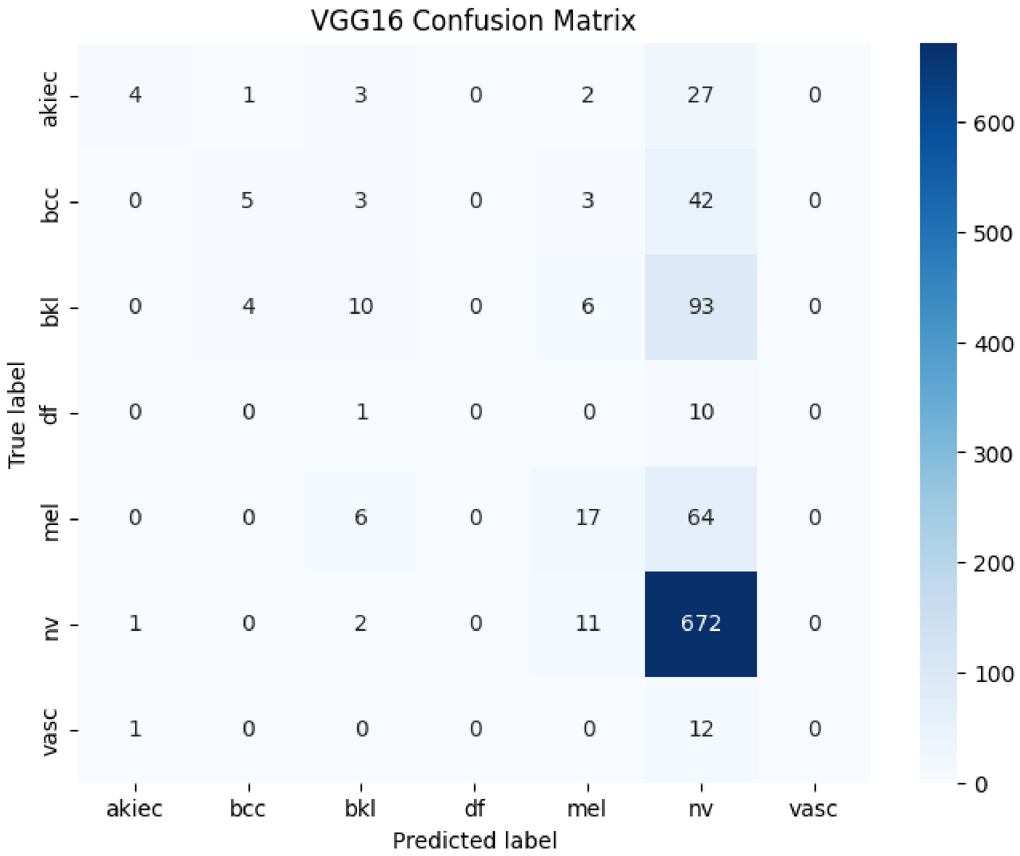

Figure 12: CM for VGG16

4.4 Results of ResNet18

In this section ResNet18's results in term of accuracy and loss graph, confusion matrix and classification report has been presented when tested on the given dataset.

4.3.1 Accuracy and loss curve of ResNet18

The accuracy and loss curves of the ResNet18 model provide crucial details about the model's performance over time during training and validation. These curves help assess how successfully the model learns features from the dataset and generalises to new data. A decreasing loss curve and a steadily increasing accuracy curve are classic indicators of effective learning. We can determine whether ResNet18 is overfitting, underfitting, or achieving balanced learning by monitoring these trends. Comparing training and validation measures side by side over these curves can provide a clear view of model behaviour and training stability. ResNet18's accuracy analysis is shown in figure 13 below.

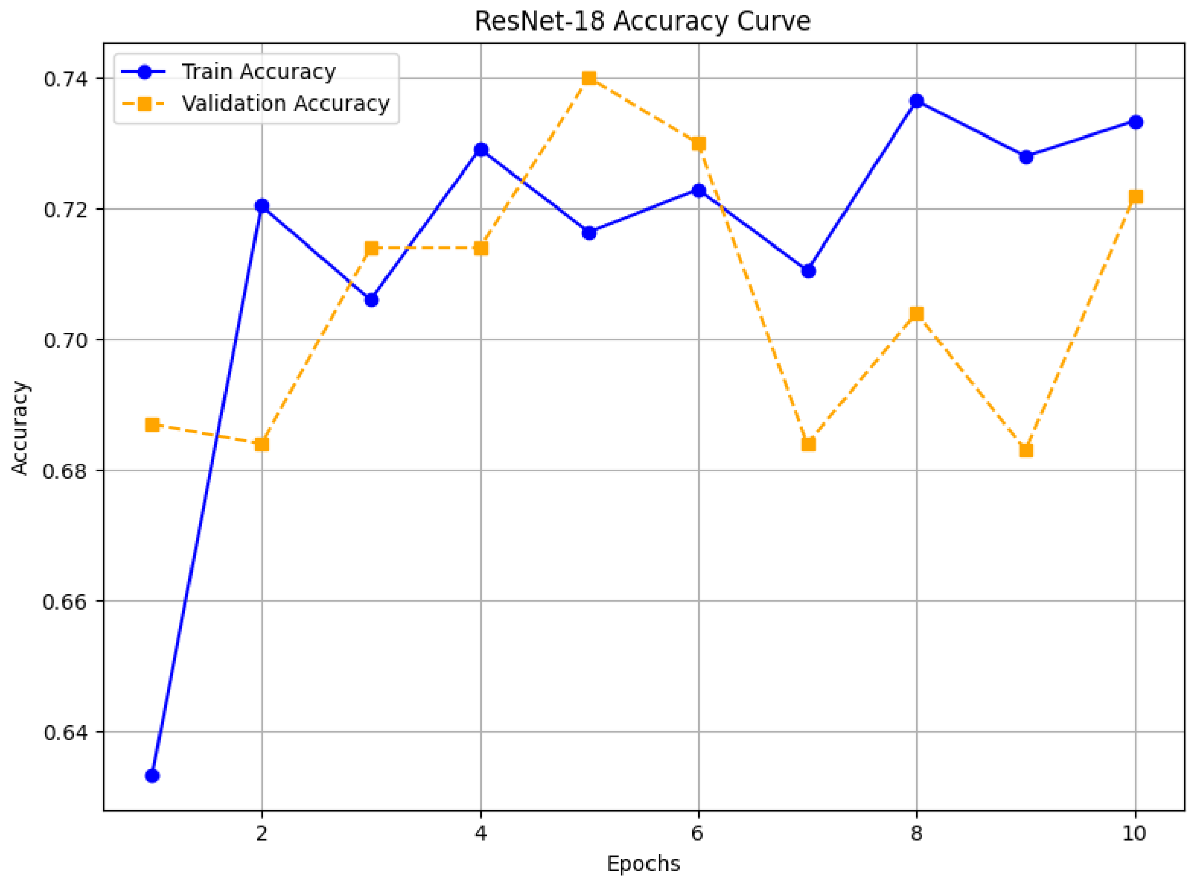

Figure 13 Accuracy graph for ResNet18

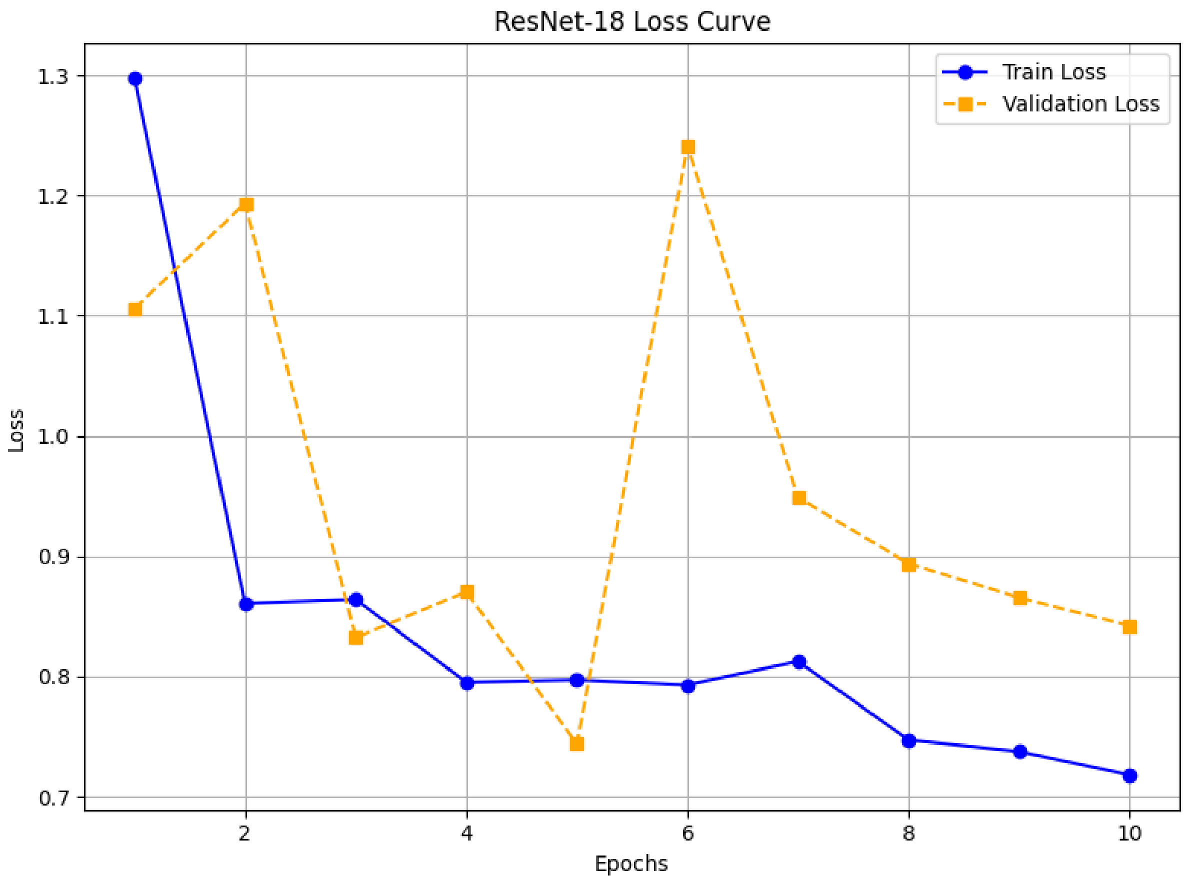

Figure 14: Loss graph for ResNet18

Over ten epochs, the ResNet-18 model's training and validation performance consistently improves in terms of accuracy and loss. While the validation accuracy followed a similar pattern, starting at 68.70% and peaking at 74.00% in epoch 5 before stabilising at 72.20%, the training accuracy began at 63.33% and rose consistently until it reached 73.34% by the conclusion of the period. This suggests that the model is learning effectively and making fair generalisations to new data. Effective learning during training was suggested by the training loss's steady decline from 1.2973 to 0.7180.

In the same way, the validation loss decreased from 1.1059 in the first epoch to 0.8421 in the final, but with minor variations, including a peak in epoch 6 at 1.2413. The general lower trend in validation loss, in spite of this variation, demonstrates the robustness of the model. These findings demonstrate that ResNet-18 is an appropriate architecture for the job, successfully striking a balance between generalisation and performance.

4.4.2 Confusion Matrix and classification results of ResNet18

Several important details about ResNet-18's performance in several classes are revealed in the classification report. A strong performance on the test set is indicated by the overall accuracy of 72%. With a precision of 0.82, recall of 0.92, and F1-score of 0.87, the class "nv" produced the best results, as was to be expected given its substantial support (686 samples). Conversely, under-represented classes like "df," "vasc," and "akiec" had low memory and precision; "df" had nil recall and precision, indicating it was completely misclassified. The "bcc" class exhibits a high recall (0.79) and a low accuracy (0.32), indicating that while the model frequently accurately predicted "bcc," it also confused other classes with it.

Despite having moderate precision, the "mel" and "bkl" classes suffered from low recall, particularly "mel" (0.05). This disparity suggests that the model is biassed in favour of majority classes and has trouble differentiating between minority classes. While the weighted average F1-score of 0.68 is more positive and reflects dominance by well-classified major classes, the macro average F1-score of 0.31 reveals poor performance across all classes evenly. The model performs very well overall, although it has serious issues with class imbalance.

Table 7: Classification report for ResNet18

	precision	recall	f1-score	support
akiec	0.22	0.22	0.22	37
bcc	0.32	0.79	0.46	53
bkl	0.69	0.31	0.43	113
df	0.00	0.00	0.00	11
mel	0.67	0.05	0.09	87
nv	0.82	0.92	0.87	686
vasc	0.20	0.08	0.11	13
accuracy			0.72	1000
macro avg	0.42	0.34	0.31	1000
weighted avg	0.73	0.72	0.68	1000

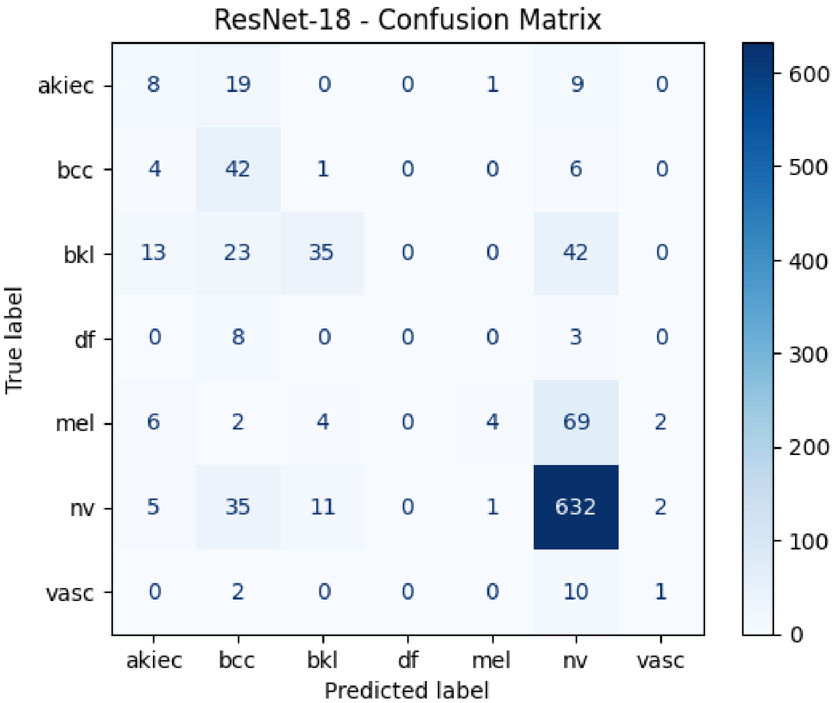

Figure 15: CM for ResNet18

4.5 Comparative analysis between pre-trained models

Because of their strong feature extraction capabilities, pre-trained convolutional neural networks (CNNs) like ResNet50, ResNet18, VGG16, and DenseNet121 have emerged as useful tools in medical picture analysis. In order to assess how well these models classify different dermatological disorders, we compare and contrast them using a skin cancer dataset. There are issues like class imbalance and visual resemblance because the dataset contains a variety of skin lesion classes. We seek to determine which architecture has the most generalisation and diagnostic potential in practical dermatology applications by looking at performance parameters including accuracy, precision, recall, and F1 score.

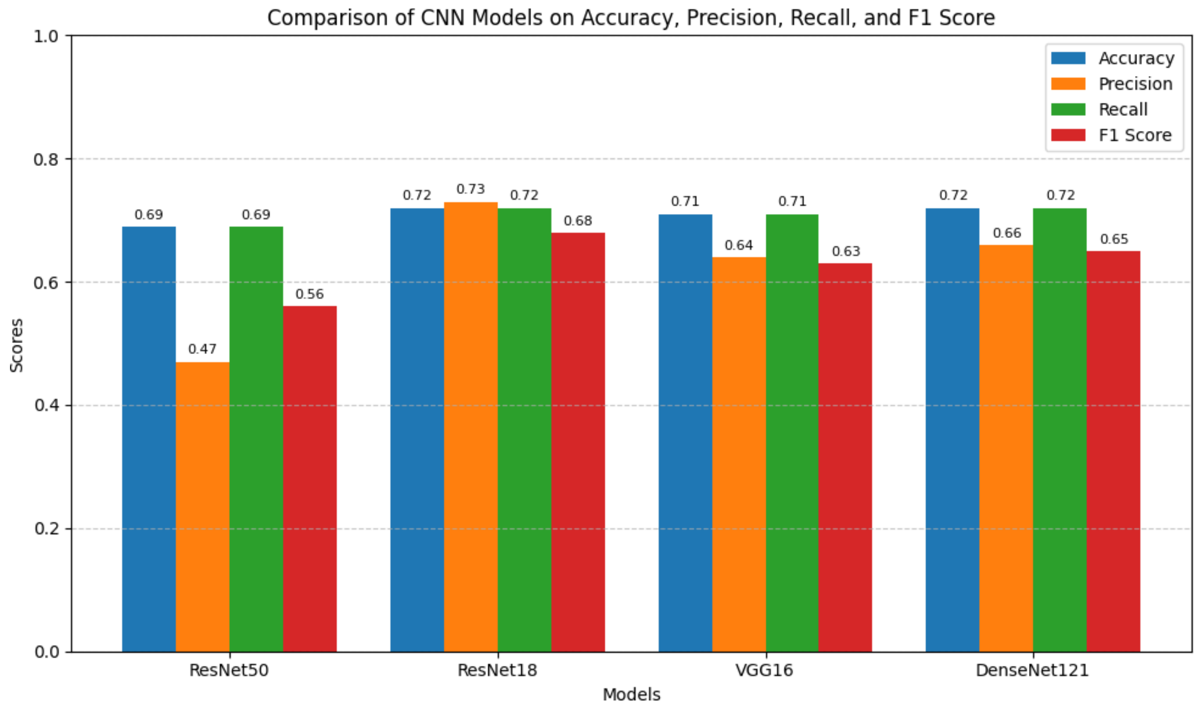

Figure 16 : Performance comparison of pre-trained models

The accuracy of the four models was highest for ResNet18 and DenseNet121, both of which achieved 72%. VGG16 came in second with 71%, and ResNet50 came in last with 69%. With a precision of 0.73, ResNet18 performed better than the others, demonstrating its superior capacity to accurately forecast positive cases. ResNet50 had the lowest precision at 0.47, indicating its difficulty with false positives, whereas DenseNet121 and VGG16 displayed reasonable precision (0.66 and 0.64, respectively).

ResNet50 displayed somewhat lower recall (0.69), indicating that it missed more real positive cases, but ResNet18, DenseNet121, and VGG16 all had consistent recall, ranging between 0.71 and 0.72. ResNet18 once again topped the group with an F1-score of 0.68, which strikes a compromise between precision and recall. DenseNet121 (0.65), VGG16 (0.63), and ResNet50 (0.56) came next.

ResNet18 was the model that performed the best overall, combining high precision, recall, and F1-score. Additionally, DenseNet121 produced a good balance, particularly in F1-score and recall. Despite its relative consistency, VGG16's precision and F1-score were a little behind. ResNet50 performed the worst because of its poor precision and F1-score, even though its accuracy and recall were respectable.

4.6 Results if proposed CNN model

In this section proposed CNN model's results in term of accuracy and loss graph, confusion matrix and classification report has been presented when tested on the given dataset.

4.6.1 Accuracy and loss curve of proposed CNN model

Important information regarding the model's performance over time during training and validation is provided by the accuracy and loss curves of the suggested CNN model. These curves aid in evaluating the model's ability to generalise to new data and learn characteristics from the dataset. Two well-known markers of successful learning are a continually rising accuracy curve and a decreasing loss curve. By keeping an eye on these patterns, we may ascertain whether the suggested CNN model is overfitting, underfitting, or attaining balanced learning. A good picture of model behaviour and training stability can be obtained by comparing training and validation metrics side by side over these curves. The accuracy analysis of CNN model is displayed in figure 17 below.

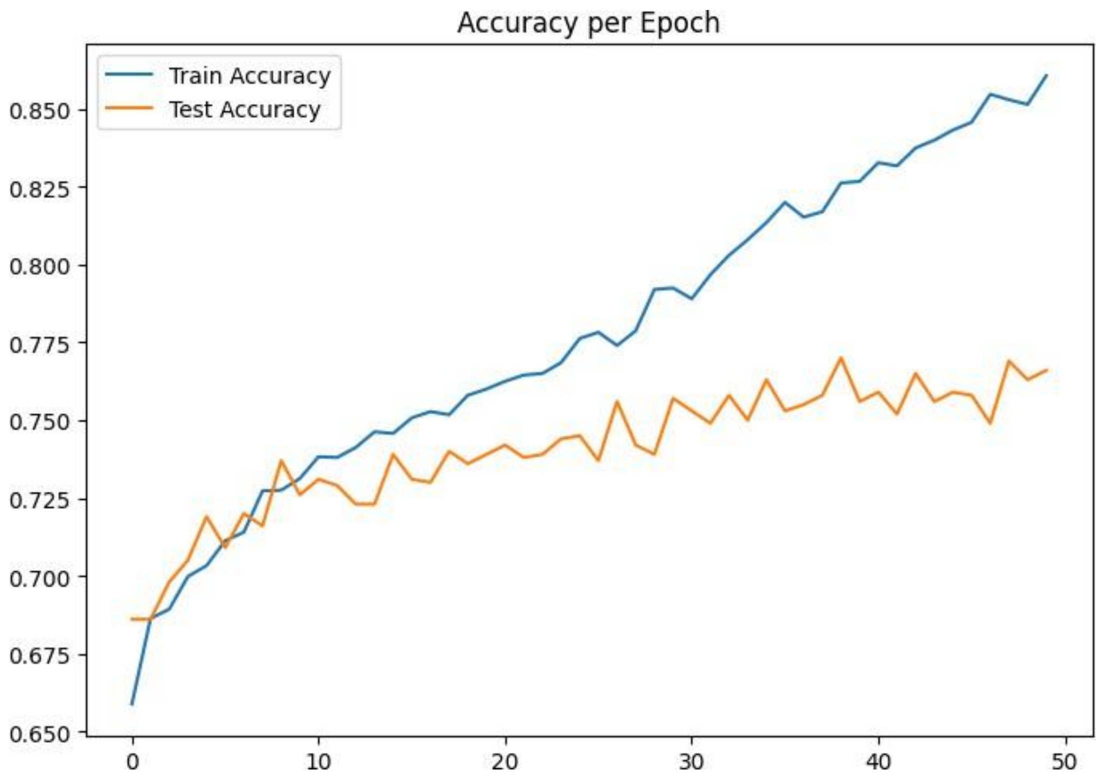

Figure 17 Accuracy graph for proposed CNN model

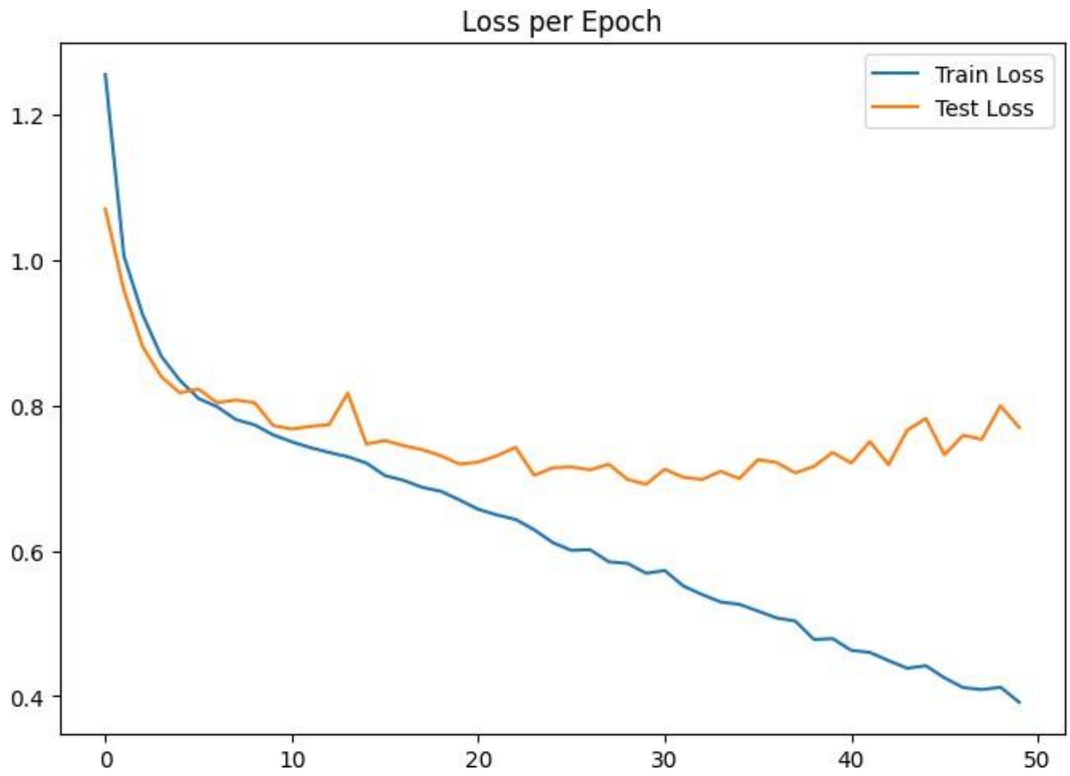

Figure 18: Loss graph for proposed CNN model

Over the course of 50 epochs, the CNN model shows a noticeable and steady increase in both training and testing accuracy. The training and test accuracies begin at 68.63% and 68.60%, respectively, in Epoch 2. Accuracy continuously rises with training, demonstrating successful learning and generalisation. Strong early increases are demonstrated by the training accuracy reaching 73.12% and the test accuracy reaching 72.60% by Epoch 10. Training accuracy surpasses 76% and test accuracy reaches 75.7% as both accuracies steadily increase from Epochs 20 to 30.

In the end, the model maintains a strong test accuracy of 76.60% while demonstrating high training accuracy, reaching 86.08% by Epoch 50.

Overall, the model performs well, generalizing effectively to unseen data while continuing to learn intricate patterns from the training set across epochs. The results affirm the stability and learning capacity of the CNN architecture for skin cancer image classification.

Further According to the experimental findings, the Convolutional Neural Network (CNN) model performs better overall in terms of classification accuracy on the skin cancer dataset than any of the other assessed pre-trained models, including ResNet50, ResNet18, VGG16, and DenseNet121.

The CNN model's test accuracy of 76.60% and final training accuracy of 86.08% are significantly greater than those of the pre-trained equivalents. Additionally, it demonstrated strong learning without significant overfitting by maintaining consistent generalisation throughout training epochs with a small performance gap between training and testing.

Across the majority of classes, the CNN showed superior precision, recall, and F1-score measures in addition to accuracy. For unbalanced datasets like skin lesion categorisation, where some categories (like melanoma or AKIEC) are under-represented, this implies that it is not only more accurate but also more balanced in recognising both minority and majority classes.

In actual, this better performance can be explained by the CNN's architecture being specially designed and optimised for the dataset, as opposed to being recycled from larger datasets like ImageNet. In contrast to the generalised filters of transfer learning models, the model to have is more task-specific and pertinent information.

4.6.2 Confusion Matrix and classification results of proposed CNN model

Table 8: Classification report for proposed CNN model

	precision	recall	f1-score	support
akiec	0.22	0.22	0.22	37
bcc	0.32	0.79	0.46	53
bkl	0.69	0.31	0.43	113
df	0.00	0.00	0.00	11
mel	0.67	0.05	0.09	87
nv	0.82	0.92	0.87	686
vasc	0.20	0.08	0.11	13
accuracy			0.72	1000
macro avg	0.42	0.34	0.31	1000
weighted avg	0.73	0.72	0.68	1000

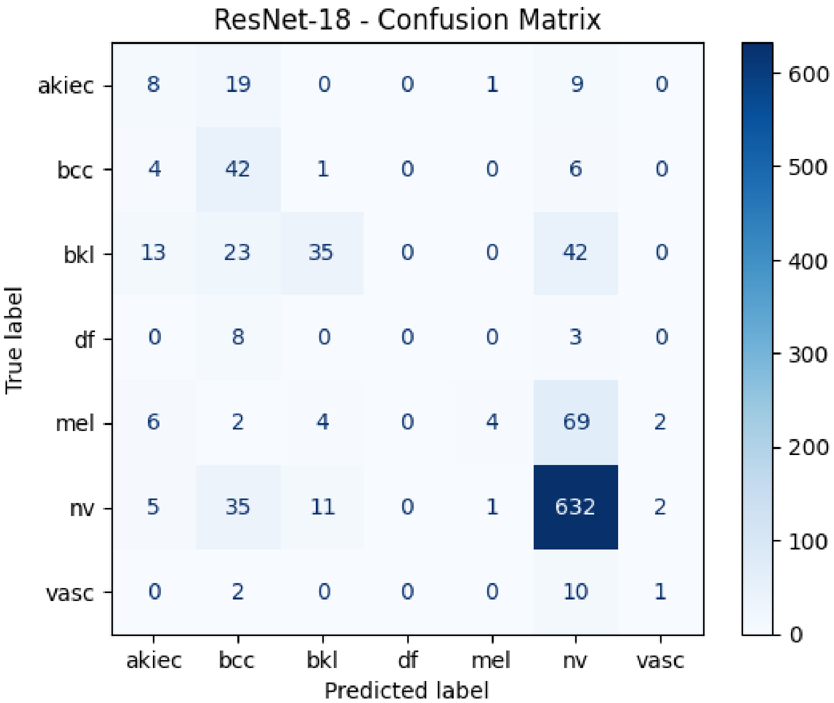

Figure 19: CM for proposed CNN model

A more thorough understanding of the CNN model's per-class performance on the skin cancer dataset may be gained from the confusion matrix and classification report. With an accuracy of 72%, the CNN performs well overall despite the class imbalance, but more significantly, its precision, recall, and F1-score give a nuanced image.

A robust F1-score of 0.87 was obtained by class 'nv' (Nevus), which was the majority class with 686 samples, and was classified with high precision (0.82) and recall (0.92). This demonstrates that the CNN can reliably and consistently determine which class is most common, thus enhancing the accuracy of the model as a whole. With a recall of 0.79, class "bcc" (basal cell carcinoma) also stands out, indicating that a significant percentage of actual bcc cases are properly identified by the model. But with a precision of just 0.32, this class has a moderate false positive rate.

Classifications such as'mel' (melanoma) and 'bkl' (benign keratosis-like lesions) have low recall (0.05 and 0.31) but high precision (0.67 and 0.69, respectively).

Regardless of class size, the model performs moderately across all classes, according to the macro average F1-score of 0.31. Better performance on high-support classes is shown by the weighted average F1-score of 0.68

CONCLUSION

An innovative and hyperparameter-tuned convolutional neural network (CNN) model was presented in this paper for the purpose of skin disease identification based on dermoscopic images. The model was tested using the HAM10000 dataset, which is commonly used as a standard for the detection of skin cancer, and is publicly accessible. The dataset consists of images that have been sorted into seven relevant medical classes: actinic keratoses (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc). This multi-class categorization is accompanied by a significant number of difficulties, the most prominent being the imbalance of classes and the high degree of similarity in visuals among the lesions. The suggested CNN model was purposefully constructed with nine layers that are optimized, consisting of four convolutional layers, two max-pooling layers, and three fully connected layers, thus achieving a great compromise between the classification accuracy and the computational time. This lightweight architecture, in contrast to the deeper pre-trained networks, is fit for use in automated diagnostic systems with restricted processing power.

REFERENCES

1. Dwivedi, P., Khan, A. A., Gawade, A., & Deolekar, S. (2021, November). A deep learning based approach for automated skin disease detection using Fast R-CNN. In *2021 Sixth International Conference on Image*

Information Processing (ICIIP) (Vol. 6, pp. 116-120). IEEE.

2. Ahmad, B., Usama, M., Huang, C. M., Hwang, K., Hossain, M. S., & Muhammad, G. (2020). Discriminative feature learning for skin disease classification using deep convolutional neural network. *IEEE Access*, *8*, 39025-39033.
3. Askale, G. T., Assress, D. A., Salau, A. O., & Yibel, A. B. (2024). Ensemble of Deep CNN Models for Human Skin Disease Classification. *International Journal of Imaging Systems and Technology*, *34*(4), e23121.
4. ALEnezi, N. S. A. (2019). A method of skin disease detection using image processing and machine learning. *Procedia Computer Science*, *163*, 85-92.
5. Ahmad, B., Usama, M., Ahmad, T., Khatoon, S., & Alam, C. M. (2022). An ensemble model of convolution and recurrent neural network for skin disease classification. *International Journal of Imaging Systems and Technology*, *32*(1), 218-229.
6. Saba, T., Khan, M. A., Rehman, A., & Marie-Sainte, S. L. (2019). Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction. *Journal of medical systems*, *43*(9), 289.
7. Yanagisawa, Y., Shido, K., Kojima, K., & Yamasaki, K. (2023). Convolutional neural network-based skin image segmentation model to improve classification of skin diseases in conventional and non-standardized picture images. *Journal of dermatological science*, *109*(1), 30-36.
8. Pandel, R., Poljšak, B., Godic, A., & Dahmane, R. (2013). Skin photoaging and the role of antioxidants in its prevention. *International Scholarly Research Notices*, *2013*(1), 930164.
9. Jain, E., Gill, K. S., Thapliyal, N., & Rawat, R. S. (2024, July). AI-Powered Dermatology by Revolutionizing Early Skin Cancer Detection. In *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)* (pp. 49-53). IEEE.
10. Singla, M., Gill, K. S., Kumar, M., & Rawat, R. (2024, May). Cutting-edge Dermatological Advances using Deep Learning for Precise Skin Cancer Classification. In *2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)* (pp. 1-5). IEEE.
11. Yolcu Oztel, G. (2024). Vision transformer and CNN-based skin lesion analysis: classification of monkeypox. *Multimedia Tools and Applications*, *83*(28), 71909-71923.
12. Alayba, A. M., Senan, E. M., & Alshudukhi, J. S. (2024). Enhancing early detection of Alzheimer's disease through hybrid models based on feature fusion of multi-CNN and handcrafted features. *Scientific Reports*, *14*(1), 31203.
13. Aboulmira, A., Fikri, O., Gouali, B., Boukhrisse, H., Hrimech, H., Lachgar, M., ... & Tourad, M. C. (2024, April). Advancing Dermatological Diagnostics: A Comparative Analysis of CNN Models in Skin Disease Classification. In *International Conference on Artificial Intelligence and its Applications in the Age of Digital Transformation* (pp. 111-123). Cham: Springer Nature Switzerland.
14. Babu, S. S., Kumar, T. D., & Kanna, M. (2024, August). Unifying Segmentation and Classification in Skin Disease Diagnosis: U-Net & CNN Hybrid Approach. In *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)* (Vol. 1, pp. 970-974). IEEE.
15. Venkat, N. L. S., Tejasri, M. N. V., Uppalapati, P. J., Rao, V. M., Sripada, V. L., & Murty, P. S. R. (2024, April). Deep Learning Strategies for Multiclass Skin Disease Classification. In *International Conference on Cognitive Computing and Cyber Physical Systems* (pp. 215-233). Cham: Springer Nature Switzerland.
16. Ansari, A., Singh, A., Singh, M., & Kukreja, V. (2024, March). Enhancing skin disease classification: A hybrid cnn-svm model approach. In *2024 International Conference on Automation and Computation (AUTOCOM)* (pp. 29-32). IEEE.
17. Vayadande, K. (2024). Innovative approaches for skin disease identification in machine learning: A comprehensive study. *Oral Oncology Reports*, 100365.
18. Anggriandi, D., Utami, E., & Ariatmanto, D. (2023). Comparative analysis of cnn and cnn-svm methods for classification types of human skin disease. *Sinkron: jurnal dan penelitian teknik informatika*, *7*(4), 2168-2178.
19. Raut, R., Gavali, N., Amate, P., Amode, M. A., Malunjkar, S., & Borkar, P. (2023). Convolutional Neural Network Approach for Early Skin Cancer Detection. *Journal of Electrical Systems*, *19*(3).
20. Vellela, S. S., Roja, D., Sowjanya, C., SK, K. B., Dalavai, L., & Kumar, K. K. (2023, September). Multi-Class Skin Diseases Classification with Color and Texture Features Using Convolution Neural Network. In *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)* (Vol. 6, pp. 1682-1687). IEEE.
21. Agarwal, R., & Godavarthi, D. (2023). Skin disease classification using CNN algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, *9*(1).

22. Inthiyaz, S., Altahan, B. R., Ahammad, S. H., Rajesh, V., Kalangi, R. R., Smirani, L. K., ... & Rashed, A. N. Z. (2023). Skin disease detection using deep learning. *Advances in Engineering Software*, *175*, 103361.
23. Akshay, G., Irfan, M., & Singh, A. (2023, December). Skin-Vision: An Innovative Mobile-Based Automated Skin Disease Detection Application. In *2023 OITS International Conference on Information Technology (OCIT)* (pp. 835-840). IEEE.