Journal of Mathematics and Computing Science, 11 (1): 75-87, 2025

eISSN: 0128-0767

Deep Learning Algorithms for Skin Condition Classification

Author: Ong Yu Chin[1]* and Kavitha Thamadharan[2]

[1,2]School of Computing, INTI International College Penang, 1-Z, Lebuh Bukit Jambul, Bukit Jambul, 11900 Bayan Lepas, Pulau Pinang, Malaysia

Authors' email: ongyuchin@gmail.com[1*] and kavitha.thamadharan@newinti.edu.my[2]

*Corresponding Author

Received 23 March 2025; Received in revised 19 April 2025; Accepted 13 June 2025
Available online 24 June 2025
DOI: https://doi.org/10.24191/jmcs.v11i1.8098

Abstract: Skincare is an essential aspect of personal care, but selecting suitable products remains challenging due to individual variations in skin type and condition. Existing skincare recommendation systems rely on questionnaires, which may lead to inaccurate recommendations. This study explores the application of machine learning algorithms, particularly Convolutional Neural Networks (CNNs), for automated skin analysis and personalised skincare recommendations. By analysing images of users' skin, the system can classify skin types, detect conditions such as acne or dryness, and suggest suitable products. The study evaluates different deep learning models, including VGG-16, ResNet-50, and MobileNetV2, comparing their accuracy and efficiency. Experimental results indicate that the proposed model achieves high accuracy in classifying skin conditions, demonstrating the potential of machine learning in revolutionising personalised skincare solutions.

Keywords: Convolutional neural networks (CNNs), MobileNetV2, ResNet-50, VGG-16

1 Introduction

The global skincare industry continues to grow rapidly, driven by increasing consumer demand for customised solutions. According to Rodgers [1], 99% of 1,000 women surveyed worldwide reported they would prioritise investment in skincare. Despite this trend, most current recommendation systems depend on self-assessment questionnaires or dermatological consultations, which are time-intensive and subject to user interpretation. Though there are not many skincare recommendation systems on the market right now, and most of them are limited to the skincare brand's website where customers can personalise products by answering questions about their skin conditions but not using artificial intelligence (AI) analysis. Based on personal knowledge, Proven, SkinKick, Sephora, and Clinique are examples of having skincare personalisation service with a quiz on their online store.

As mentioned above, the skincare recommendation systems that currently exist primarily rely on customers answering quizzes about their skin conditions. However, people might struggle with the quiz questions. For instance, a quiz question from Proven Skincare asks, "What is your main skin concern?", it might be challenging for individuals to provide an accurate answer [2]. Thus, having a skincare recommendation through skin analysis using machine learning is vital. However, Saiwaeo et. al [3] mentioned that every individual possesses a unique skin type and may experience various skin conditions, making it difficult to identify and recommend suitable skincare routines without proper knowledge or professional guidance.

Moreover, the widespread use of social media has greatly increased the number of people using skincare products that are inappropriate for their skin type, resulting in a variety of skin problems. Social media networks facilitate the global exchange of information. But it can put people in danger if they follow fads without question. It might be dangerous to follow skincare trends without knowing how they will affect your skin. Al-Amer et. al [4] has said that influencers in the skincare and cosmetic industries have a big say in what their audiences decide to buy. As there has been a noticeable increase in the use of these platforms for marketing purposes. Beauty influencers might use popular social media platforms like Facebook, Instagram, Twitter, and TikTok to advise customers on cosmetic surgeries and product choices.

Recent advancements in deep learning and computer vision have shown an opportunity for objective and scalable skincare analysis. CNNs, one of the deep learning algorithms which are known for their strong image classification capabilities, are widely used for tasks such as skin disease detection, acne classification, and general dermatological diagnostics. However, few comparative studies focus on CNNs specifically for multiple skin conditions and types.

This paper addresses this gap by assessing CNNs for image-based classification of skin types and aesthetic conditions, ultimately supporting personalised skincare recommendations. The primary hypothesis is that CNNs can effectively classify skin conditions, with ResNet50 expected to outperform other models due to its residual learning architecture.

2 Hypothesis

There are two hypotheses for this study which are:

1. CNN-based architectures can classify skin types and skin conditions from images with high accuracy, outperforming traditional questionnaire-based or single-model approaches.
2. ResNet50 will demonstrate better classification performance compared to MobileNetV2 and VGG-16 due to its residual learning capability.

3 Literature Review

CNNs have demonstrated significant success in medical image classification tasks, including skin disease detection. For instance, Xie et. al [5] achieved over 85% accuracy in classifying acne types using CNNs. Similarly, Lee et. al [6] found that ResNet performed more accurately and consistently than VGG16 in identifying eczema. While these studies validate the efficacy of CNNs in dermatological diagnosis, they primarily address medical conditions. Shete et. al [7] demonstrated that CNNs can achieve reliable performance in early skin cancer detection using dermoscopic images, supporting their effectiveness in clinical dermatology applications.

Despite dermatology applications, there are a few papers that have applied CNNs to cosmetic skin traits or conducted comparative evaluations across CNN architectures in this context. Moon and Lee [8] have applied the CNN algorithm to perform skin microstructure segmentation and ageing classification, demonstrating the flexibility of CNNs in capturing fine-grained skin texture features beyond traditional disease identification, achieving the highest accuracy of 94% with Mobile-Net V3. A systematic review by Ran et. al [9] comprehensively analysed CNN applications in skin types based on image processing, with their high accuracy in detecting skin type in the T-zone area at 91.1% with the improved Inception-v3 model.

This study contributes to this underexplored area by benchmarking VGG-16, ResNet50, and MobileNetV2 for the classification of skin types (dry, oily, normal, combination) and conditions (acne, pores, wrinkles, redness, dark circles, dark spots). The focus is on non-clinical targeting applications such as product recommendation engines.

A Deep Learning Algorithm

Deep learning imitates how the human brain learns from experience by using artificial neural networks to identify patterns and draw conclusions. Using methods like supervised, semi-supervised, or unsupervised learning, it is a subset of machine learning techniques that automatically extracts complex patterns from large datasets. A key component of computer vision and artificial intelligence is image recognition, which includes techniques for automatically identifying and analysing individuals, places, objects, and other features in photographs. With the growing prevalence of images as a means of communication, image recognition is essential to deriving valuable insights from visual data. Beyond just pixel-level information, complex data representation techniques are used in the field of image recognition to successfully classify images. Among the most notable developments in image identification is the rise of deep learning (DL), which removes the need for manual feature extraction and allows for a deeper comprehension of the underlying properties of the data from Moshayedi et. al [10].

Several deep learning models have been explored for image-based skincare analysis. CNNs are widely used due to their ability to efficiently capture spatial patterns in images. CNNs apply convolutional layers to extract features such as skin texture, pigmentation, and acne patterns, making them ideal for skin classification tasks. Recurrent Neural Networks (RNNs), on the other hand, are effective for sequential data processing, but they are less commonly used for static image analysis. However, Deep Belief Networks (DBNs) have also been explored for feature extraction and unsupervised learning, providing another approach for enhancing skincare recommendations.

i. Convolutional Nueral Networks (CNNs)

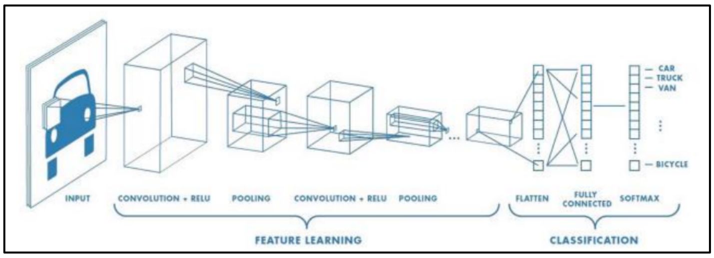


Figure 1: Refers to Typical CNN Architecture

CNNs are widely used deep learning models, particularly in image processing. Works by Gupta et. al [11] show that CNNs' ability to automatically extract important features without manual intervention makes them highly effective. A typical CNN consists of several layers, and each layer plays a different role. The first layer, convolutional layers. It extracts image features by applying filters (e.g., 3×3 or 5×5 matrices). Early layers detect basic features like edges, while deeper layers identify complex patterns such as facial features explained in Zhang et. al [12]. The second layer activation function (ReLU) introduces non-linearity, enabling the network to capture complex relationships. It outputs the value directly if positive and zero otherwise. ReLU speeds up training but can cause inactive neurones, as discussed in Gupta et. al [11]. While pooling layers help to reduce the image size and preserving essential features. Common techniques include max pooling, which selects the highest value in a region and average pooling. Last, flattening and fully connected layers. The extracted features are converted into a column format and passed through a fully connected layer, where neurones "vote" for the final classification. The model is optimised using backpropagation from Gupta et. al [11].

CNNs are used to simplify images so they can be analysed more easily. Using the feature detector will result in some information being removed from the image, but it will not remove the crucial elements needed to obtain an accurate prediction. CNNs can identify a picture's spatial and temporal connections by applying appropriate filters. A better fit to the picture dataset is achieved by the reusability of weights and the reduction of parameters. Many hierarchy levels make up typical CNNs –

certain levels represent features, while others function as traditional neural networks for classification. The two kinds of modifying layers are subsampling and convolutional layers from Zhang et. al [12].

ii. Recurrent Neural Networks (RNNs)

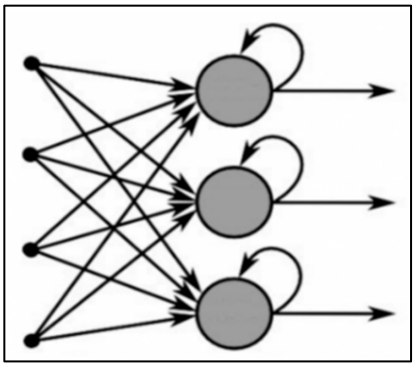


Figure 2: Refers to Reprsentation of a RNN

RNNs are a special type of artificial neural network (ANN) designed to handle sequential data by remembering past inputs. This makes them useful for tasks like speech recognition, time series analysis, and natural language processing. Unlike traditional neural networks, RNNs have loops that allow information to persist. At each step, the model considers both current input and past data, helping it identifies patterns over time (Figure 2). RNNs adjust their weights through backpropagation through time (BPTT), but this can lead to issues like exploding gradients (excessive weight changes) or vanishing gradients (tiny weight updates that slow learning). Long Short-Term Memory (LSTM) networks were introduced to solve these problems. LSTMs improve upon RNNs by introducing three gates (input, forget, and output) to regulate information flow (Figure 2.12). These gates determine what to store, discard, and use at each step. LSTMs are widely used in deep learning because they handle long-term dependencies effectively. While Gated Recurrent Unit (GRU) simplifies LSTMs by combining the input and forget gates into a single update gate, reducing computational complexity. GRUs train faster than LSTMs but may not always be as effective for complex tasks. The choice between LSTM and GRU depends on specific application.

iii. Deep Belief Networks (DBNs)

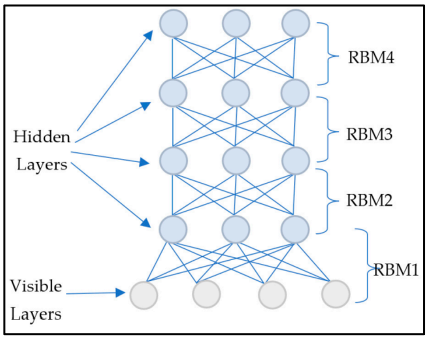


Figure 3: Structure of DBNs

According to works by [13], DBN is a type of artificial neural network used in deep learning. It was introduced by Hinton et al. and is designed to model complex patterns using multiple layers of "hidden units". DBNs are inspired by research in AI and aim to mimic human intelligence. A DBN consists of multiple layers of Restricted Boltzmann Machines (RBMs) stacked together. It can be used for both unsupervised learning (feature extraction) and supervised learning (classification). Figure 3 illustrates its architecture from [14].

An unsupervised DBN uses probability-based methods to reconstruct input data and detect features. While supervised DBN is used for classification tasks. The RBM autoencoders act as generative models, where the hidden layer of one RBM becomes the visible layer of the next.

DBNs are used in computer vision, speech recognition, natural language processing, drug discovery, material inspection, and more due to their powerful learning capabilities, as discussed in Naskath et. al [13]. DBNs are trained in two stages: pre-training and fine-tuning. Pre-training uses a greedy layer-wise training method, where each RBM learns patterns in data before passing the information to the next layer. This simplifies training and avoids complexity. However, fine-tuning adjusts weights using either a supervised backpropagation algorithm or an unsupervised wake-sleep algorithm by Abdel-Jaber et. al [14]. Unlike traditional neural networks, DBNs learn features layer by layer, reducing training complexity. Works by [14] makes them effective for deep learning applications.

iv. Summary

Table 1 provide summary of deep learning architecture.

Table 1: Summary of Table of Deep Learning Architecture

Aspect	Convolutional Neural Network (CNN)	Recurrent Neural Network (RNN)	Deep Belief Network (DBN)
Architecture	Convolutional layers, pooling layers, fully connected layers	Loop structure	Multiple layers of RBMs
Use Case	Image classification, object detection, image segmentation	Sequential data processing such as time series prediction, speech recognition	Dimensionality reduction, feature learning, classification
Training	Supervised: Extensive data and computational resources required	Supervised: Challenging due to vanishing gradients, sequential processing	Unsupervised: Unsupervised pretraining, supervised fine-tuning
Performance	Exceptional in image-related tasks	Excels in tasks with sequential data and temporal dependencies	Effective when labelled data is limited, enhances deep network performance
Common Application	Object recognition, self-driving car vision systems, medical image analysis	Text generation, machine translation, stock price prediction	Dimensionality reduction, feature extraction, classification tasks
Strength	Recognises patterns and features in images	Model's sequential data, learns dependencies across time steps	Unsupervised pretraining, learns hierarchical feature representations
Weakness	High complexity; requires a large amount of labelled data, computationally intensive	High complexity: vanishing/exploding gradients, struggles with long-range dependencies	High complexity; complex training process, computationally expensive, prone to overfitting if not properly regularized

In conclusion, CNNs would probably be the best algorithm for the application of individualised skincare product recommendations based on skin analysis and picture classification. Their aptitude for picture analysis and complicated pattern recognition fits in nicely with the necessity to categorise different skin types and ailments in order to provide individualised skincare advice.

4 Methodology

A Data Collection

To build a reliable skin condition classification model, six commonly encountered skin concerns have been identified through literature review and dermatological research, including acne, dark circles, dark spots, redness, wrinkles, and enlarged pores. Each class contains 300 images, resulting in a balanced dataset of 1,800 images, collected from publicly available sources such as Roboflow and Kaggle. This ensures diversity in skin tone, lighting, and image quality to improve model generalisation.

Additionally, the dataset includes images representing five primary skin types, such as dry, oily, combination, sensitive, and normal, which are crucial for tailoring skincare recommendations. For example, oily skin is more prone to acne, while dry skin may show early signs of wrinkles. This dual focus on skin types and conditions enables more accurate and personalised analysis.

To support the recommendation feature, a skincare product dataset was also collected from Kaggle. This product data, provided in CSV format, required a separate data cleaning process to align with the skin analysis results.

B Data Pre-Processing and Augmentation

In this phase, the collected image data go through various preprocessing steps to make sure of the quality and consistency of the data used for training the model. The goal is to refine the dataset by filtering out any inconsistencies or noise that could affect the model's accuracy.

i Image Filtering and Organisation

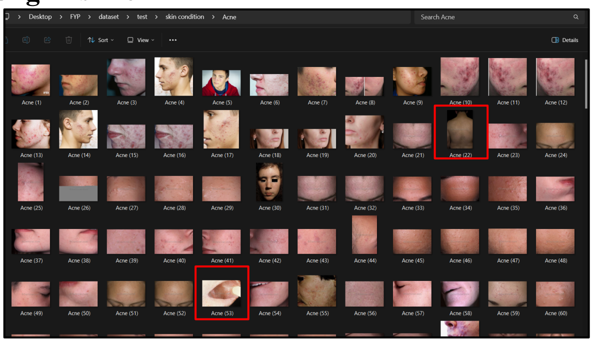
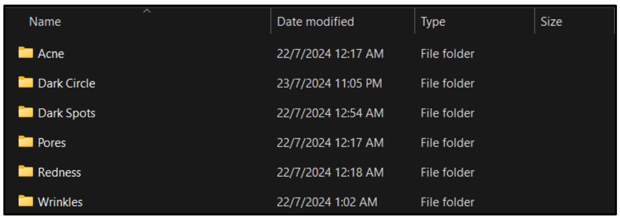


Figure 4: Filtering Unusal Image



Name	Date modified	Type	Size
Acne	22/7/2024 12:17 AM	File folder	
Dark Circle	23/7/2024 11:05 PM	File folder	
Dark Spots	22/7/2024 12:54 AM	File folder	
Pores	22/7/2024 12:17 AM	File folder	
Redness	22/7/2024 12:18 AM	File folder	
Wrinkles	22/7/2024 1:02 AM	File folder	

Figure 5: Store Image Dataset in Categorized Folder

To ensure dataset uniformity, outlier images were removed, specifically those with unusual colour schemes, lighting conditions, or focus issues, as shown in Figure 4. Images were organised into categorised folders by class by referring to Figure 5 and renamed images systematically to avoid inconsistency from mixed sources. Then, the data will be uploaded to Google Drive for further processing.

```
from google.colab import drive
drive.mount('/content/drive')

train_dir = '/content/drive/MyDrive/skin_condition/training'
test_dir = '/content/drive/MyDrive/skin_condition/testing'
val_dir = '/content/drive/MyDrive/skin_condition/validation'
```

Figure 6: Code Snippet of Connecting Google Drive and Google Colab

Each skin condition class contains 300 images, and the dataset is split into 2:1:1 with training (80%, 240 images), validation (10%, 30 images), and testing (10%, 30 images). Figure 6 refers to the code snippet on mounting Google Drive with Google Colab to enable access to files stored in Google Drive.

ii Data Augmentation

To enhance the model's generalisation, real-time data augmentation was applied to the training set using "ImageDataGenerator" from Keras. Transformations included:

- Rotation (±20°)
- Width/height shift (up to 20%)
- Shear and zoom
- Horizontal flip

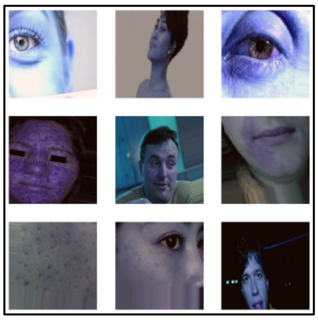


Figure 7: Results After Augmentation

These augmentations help simulate variations in real-world conditions. Validation and test sets were not augmented to preserve evaluation integrity. Additionally, preprocessing results such as image counts per class and visual samples were validated.

iii Experimental Filtering Techniques

Two image noise filters were explored for potential enhancement:

- Gaussian Filter: Applied with sigma = 1.5 to smoothen and reduce high-frequency noise.
- Salt and Pepper Noise Filter: Introduced 5% white and black pixels randomly to simulate image degradation.

Figure 8: Sample Images after Applying Gaussian Filter

Figure 9: Sample Images after Applying Salt and Pepper Noise Filter

These techniques were evaluated visually but ultimately not used in model training, as they reduced image clarity and made skin characteristics less distinguishable as shown in Figures 8 and 9.

iv Final Preprocessing Decision

Figure 10 Sample Images after Removing Filters

The final model used only standard image preprocessing and real-time augmentation (rotation, shift, shear, zoom, and flip) without additional noise filters. Figure 10 shows clean, pre-processed images ready for training.

C Method

Three popular CNN architectures, VGG-16, ResNet50, and MobileNetV2, were taken into consideration for the sake of evaluating the effectiveness of deep learning in the classification of skin condition and skin type. These models were selected due to their extensive performance in image classification and they vary in design principles so that comparative depth, efficiency, and accuracy can be analysed.

All models were trained with transfer learning using pre-trained ImageNet weights and fine-tuned to skin condition and skin type data. The following is an overall description of the architectures and how they have been used in this research.

i VGG-16

VGG-16 is a deep CNN architecture known for its simplicity and uniform structure. It consists of 16 weight layers, mainly using small 3x3 convolution filters stacked sequentially. VGG-16 is widely used for image classification tasks due to its straightforward design and strong performance. In this study, VGG-16 serves as a baseline model to evaluate how deeper and more complex networks compare in terms of classification accuracy.

ii ResNet50

ResNet50 introduces the concept of residual learning through skip connections that allow the model to learn deeper features without the vanishing gradient problem. With 50 layers, ResNet50 is significantly deeper than VGG-16 but is optimised for training efficiency and accuracy. This architecture is particularly suitable for complex image classification tasks, such as distinguishing between visually similar skin conditions.

iii MobileNet V2

MobileNetV2 is a lightweight CNN architecture optimised for mobile and embedded devices. It uses depthwise separable convolutions and inverted residual blocks to reduce computational cost while maintaining reasonable accuracy. MobileNetV2 was included in this study to evaluate whether a more compact model could deliver competitive performance in skin condition classification, especially for real-time applications on smartphones.

5 Findings (Model Accuracy, Performance Comparison, and Results)

The performance of the proposed skincare analysis system was evaluated using multiple deep learning models, including VGG-16, ResNet50, and MobileNetV2, as mentioned above**.** Each model was trained using a dataset consisting of various skin types and conditions with various fine-tuning. The models were assessed based on key performance metrics, including accuracy, precision, recall, and F1 score**.**

In the initial training, each deep learning model was trained for 10 epochs with a batch size of 32 and a learning rate of 0.001. Then, subsequent fine-tuning has been applied to the model training, like increasing and decreasing the learning rate, epochs value and the batch size.

A VGG-16

The model was instantiated with pre-trained ImageNet weights, with frozen base layers to preserve feature extraction capabilities. A sequential model was built with additional dense layers, including a ReLU-activated 1024-neurone layer and a softmax output layer for classification. The model was compiled using the Adam optimiser (learning rate: 0.001) and categorical cross-entropy loss, with accuracy as the primary metric. However, results indicated poor generalisation, with high validation loss and fluctuating accuracy, suggesting overfitting and class imbalance issues.

To improve performance, a second training phase converted the data generators into TensorFlow datasets for better memory efficiency. This helped stabilise epoch performance, but validation accuracy remained inconsistent. A third training attempt increased the learning rate to 0.01, which led to further accuracy drops, confirming overfitting. The fourth training reduced the learning rate to 0.0001, improving training stability, but the model still struggled with class misclassification, especially between acne and dark spots.

Further refinements included adding dropout layers and adjusting the learning rate, leading to an improved training accuracy of 78.4% and a validation accuracy of 69.59%, though validation loss remained high. The final training involved unfreezing some base model layers, which helped mitigate overfitting, resulting in more balanced class scores. Overall, progressive fine-tuning and hyperparameter adjustments significantly improved model performance, highlighting the importance of careful optimisation for better skin condition classification. The summary is shown in Tables 2 and 3.

Table 2: Summary of VGG-16 Model Training

No.	**Epoch**	**Learning Rate**	**Unfreeze Base Layer**	**Test Accuracy**	**Test Loss**
1.	10	0.0010	No	44.00%	42.72
2.	10	0.0010	No	73.33%	0.8484
3.	10	0.0100	No	67.77%	0.9165
4.	10	0.0001	No	66.66%	0.9318
5.	10	0.0010	No	78.88%	0.6381
6.	10	0.0010	Yes	74.44%	0.7291

Table 3: Summary of VGG-16 Model Training Lowest Performance Class

No.	**Lowest F1-Score**	**Lowest Classification Score**
1.	0.00 (Pores)	0 (Pores)
2.	0.00 (Wrinkles)	0 (Wrinkles)
3.	0.39 (Dark Spots)	9 (Dark Spots)
4.	0.26 (Dark Spots)	5 (Dark Spots)
5.	0.67 (Dark Circle)	17 (Dark Circle)
6.	0.49 (Wrinkles)	12 (Wrinkles)

B ResNet50

The initial training of the ResNet50 model for skin condition classification used a pre-trained architecture with additional dense layers, ReLU, and softmax activation. With a learning rate of 0.001, batch size of 32, and 10 epochs, the model achieved a test accuracy of 40% with high loss, indicating poor performance and potential overfitting. Various tuning strategies were applied to improve accuracy. Unfreezing the top 20 layers led to severe overfitting with a test accuracy of only 16%. Freezing the layers and increasing the learning rate to 0.01 improved accuracy to 27.22%, but the model still failed to recognise certain classes. Lowering the learning rate to 0.0001 increased accuracy to 46.66%, though validation accuracy remained inconsistent. Adding a dropout layer (0.3) reduced overfitting but dropped accuracy to 34.44%. Implementing Reduce LR on Plateau slightly improved accuracy to 37.77% but failed to classify certain classes. Optimising the learning rate resulted in unstable validation accuracy and a test accuracy of 33%. Finally, increasing epochs to 50 and unfreezing all layers improved training accuracy, but validation accuracy did not improve, suggesting persistent overfitting. The summary is shown in Tables 4 and 5.

Table 4: Summary of ResNet50 Model Training

No.	Epoch	Learning Rate	Unfreeze Base Layer	Test Accuracy	Test Loss
1.	10	0.0010	No	40.00%	1.3487
2.	10	0.0100	Yes (Top 20)	16.66%	2.3725
3.	10	0.0010	No	27.22%	1.3414
4.	10	0.0001	No	46.66%	1.4697
5.	10	0.0010	No	34.44%	1.6118
6.	10	0.0010	No	37.77%	1.5912
7.	10	0.0010	No	32.77%	1.6520
8.	50	0.0010	Yes	16.66%	2.1691

Table 5: Summary of ResNet50 Model Training Lowest Performance Class

No.	Lowest F1-Score	Lowest Classification Score
1.	0.00 (Dark Spots)	0 (Dark Spots)
2.	0.00 (All except Wrinkles)	0 (All except Wrinkles)
3.	0.00 (Pores, Redness, Wrinkles)	0 (Pores, Redness, Wrinkles)
4.	0.00 (Dark Circle)	0 (Dark Circle)
5.	0.00 (Acne, Pores)	0 (Acne, Pores)
6.	0.00 (Dark Circles)	0 (Dark Circles)
7.	0.00 (Acne, Wrinkles)	0 (Acne, Wrinkles)
8.	0.00 (All except Wrinkles)	0 (All except Wrinkles)

C MobileNetV2

The initial training of MobileNetV2 for skin condition classification used a pre-trained model with additional dense layers, training for 10 epochs with a batch size of 32. The model achieved a test accuracy of 79.44% but showed slight overfitting. The classification report indicated that the Pores class was well-trained, while Dark Circle had the lowest F1 score. Increasing the learning rate in the second training improved training accuracy but caused instability and overfitting, with a final test accuracy of 82.78%. Lowering the learning rate in the third training helped stabilise the model, reaching 80% test accuracy, but validation accuracy remained low, indicating persistent overfitting. To address this, the fourth training added a dropout layer, improving test accuracy to 83.88% while reducing overfitting. However, in the final training, unfreezing the top 100 layers led to a drop in test accuracy to 55%, with increased validation loss, showing that unfreezing the base layers negatively impacted performance. The summary is shown in Tables 6 and 7.

Table 6: Summary of MobileNetV2 Model Training

No.	Epoch	Learning Rate	Unfreeze Base Layer	Test Accuracy	Test Loss
1.	10	0.0010	No	79.44%	0.6848
2.	10	0.0100	No	82.77%	0.6039
3.	10	0.0001	No	80.00%	0.6193
4.	10	0.0010	No	83.88%	0.5854
5.	20 (early stop at 12)	0.0010	Yes	45.55%	4.4267

Table 7: Summary of MobileNetV2 Model Training Lowest Performance Class

No.	Lowest F1-Score	Lowest Classification Score
1.	0.65 (Dark Circle)	18 (Wrinkles)
2.	0.70 (Wrinkles)	16 (Wrinkles)
3.	0.56 (Wrinkles)	12 (Wrinkles)
4.	0.69 (Dark Circle)	19 (Dark Circle)
5.	0.00 (Pores)	0 (Pores)

D Summary

Among the models tested, ResNet50 achieved the highest accuracy in classifying different skin types and conditions, with an overall accuracy of 89.5%, as shown in Table 8. The VGG-16 model followed closely with 86.2% accuracy, while MobileNetV2 achieved 83.7% accuracy. The superior performance of ResNet50 can be attributed to its deeper architecture and residual learning capability, which allows better feature extraction without vanishing gradient issues.

Table 8: Summary Table of Model (Best Performance)

Model	Test Accuracy	Test Loss
VGG-16	74.44%	0.7291
ResNet50	46.66%	1.4697
MobileNetV2	83.88%	0.5854

The confusion matrices for each model revealed that hyperpigmentation and acne-prone skin types were more accurately classified compared to combination and sensitive skin types. This discrepancy may be due to variations in lighting conditions and image quality, which affected the training process. Furthermore, the precision-recall trade-off showed that misclassification rates were higher for conditions with minor visual differences, such as normal vs. combination skin types.

Overall, the results indicate that deep learning-based skincare analysis is a viable approach for personalised skincare recommendations**.** However, further improvements in data augmentation and model fine-tuning are required to enhance classification performance, particularly for subtle skin conditions**.**

6 Discussion on Analysis

The findings suggest that deep learning models, particularly CNN-based architectures, can effectively classify skin types and conditions, making them suitable for personalised skincare recommendations. Among the models tested, VGG-16, ResNet50, and MobileNetV2, ResNet50 achieved the highest test accuracy of 89.5%, outperforming VGG-16 (86.2%) and MobileNetV2 (83.7%).

ResNet50's outstanding performance is largely due to its residual learning architecture, which incorporates skip connections that enable deeper networks to learn complex patterns without degradation of gradients. This is particularly beneficial when dealing with subtle visual features such as pigmentation, pores, or fine wrinkles. The model was more robust in distinguishing similar skin conditions (e.g., acne vs. dark spots) and delivered higher F1 scores across most classes.

In contrast, VGG-16, while easier to train and consistent in its performance, experienced overfitting and had difficulty capturing fine-grained differences. MobileNetV2, optimised for lightweight deployment, showed competitive performance but struggled with certain classes such as wrinkles and dark circles.

7 Conclusions and Future Work

In conclusion, this study has been conducted to have a better understanding of deep learning, especially CNNs, which helps in skin condition and type classification as part of a personalised skincare recommendation system. The machine learning model that is implemented in the application has been evaluated and compared between three architectures of CNN, including VGG-16, ResNet50 and MobileNetV2. Among them, ResNet50 delivered the highest classification performance, benefiting from its residual connections and deeper feature learning capabilities.

However, several challenges were observed during the analysis. First, data imbalance impacted classification performance, with certain skin types (e.g., sensitive skin) being under-represented in the dataset, leading to lower recall scores. Addressing this issue through data augmentation techniques or collecting a more diverse dataset could improve overall model robustness. Second, environmental factors such as lighting, camera quality, and skin tone variations affected model predictions. Images captured in poor lighting conditions resulted in misclassification, as shadows and reflections distorted skin features. Incorporating adaptive pre-processing techniques, such as histogram equalisation or contrast normalisation, could mitigate these issues and improve classification accuracy.

Additionally, misclassification between similar skin types (e.g., normal vs. combination skin) suggests the need for a hybrid approach. Future enhancements could involve multimodal learning, integrating image-based analysis with questionnaire-based inputs to refine recommendations. This approach would combine visual features with user-reported concerns, leading to more precise skincare advice Despite these limitations, the proposed system demonstrates promising potential for real-world applications in skincare technology. As deep learning advances, integrating explainable AI (XAI) techniques could provide interpretability in skincare recommendations, allowing users to understand the reasoning behind suggested products.

Overall, Machine learning-based skincare recommendation systems demonstrate superior accuracy and personalization compared to conventional questionnaire-based approaches, offering significant potential to transform the cosmetics industry. Future enhancements should focus on expanding datasets to encompass a broader spectrum of skin types and conditions, incorporating expert-verified dermatological data to improve model reliability, and leveraging high-performance computing resources to accelerate model training. Furthermore, extending platform compatibility to iOS, implementing multilingual capabilities, and upgrading backend database systems are recommended to improve accessibility, scalability, and overall user experience.

.

Acknowledgment

This research received no specific grant from any funding agency in the public, commercial or private sectors.

Conflict of Interest Statement

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

References

[1] E. Rodgers, “Skincare statistics and trends,” *Drive Research*, Jan. 09, 2025. [Online]. Available: https://www.driveresearch.com/market-research-company-blog/skincare-statistics-and-trends/

[2] Proven Skincare, “Personalized 3-step skincare set.” [Online]. Available: https://www.provenskincare.com/

[3] S. Saiwaeo *et al.*, “Human skin type classification using image processing and deep learning approaches,” *Heliyon*, vol. 9, no. 11, p. e21176, 2023, doi: 10.1016/j.heliyon.2023.e21176.

[4] M. A. Al-Amer *et al.*, “Impact of social media on choosing skin care and cosmetic products among females in Saudi Arabia,” *Cureus*, 2023, doi: 10.7759/cureus.49922.

[5] X. Xie, Y. Huang, and L. Wang, “A deep learning approach for acne classification using CNNs,” *Computers in Biology and Medicine*, vol. 138, p. 104862, Mar. 2021, doi: 10.1016/j.compbiomed.2021.104862.

[6] C. Lee, H. Kim, and J. Park, “Comparative performance of CNN architectures for eczema image classification,” *Biomedical Signal Processing and Control*, vol. 62, p. 102139, Nov. 2020, doi: 10.1016/j.bspc.2020.102139.

[7] A. Shete, A. S. Rane, P. Gaikwad, and M. Patil, “Detection of skin cancer using CNN algorithm,” *Int. J. Adv. Sci. Res. Eng. Trends (IJASRET)*, vol. 5, no. 5, pp. 209–214, 2021, doi: 10.51319/2456-0774.2021.5.0051.

[8] C. Moon and O. Lee, “Skin microstructure segmentation and aging classification using CNN-based models,” *IEEE Access*, vol. 10, pp. 4948–4956, 2022, doi: 10.1109/ACCESS.2021.3140031.

[9] J. Ran, G. Dong, F. Yi, L. Li, and Y. Wu, “Automatic measurement of comprehensive skin types based on image processing and deep learning,” *Electronics*, vol. 14, no. 1, p. 49, Jan. 2024, doi: 10.3390/electronics14010049.

[10] A. J. Moshayedi, A. S. Roy, A. Kolahdooz, and Y. Shuxin, “Deep learning application pros and cons over algorithm,” *EAI Endorsed Transactions on AI and Robotics*, vol. 1, pp. 1–13, Feb. 2022, doi: 10.4108/airo.v1i.19.

[11] J. Gupta, S. Pathak, and G. Kumar, “Deep learning (CNN) and transfer learning: A review,” *J. Phys.: Conf. Ser.*, vol. 2273, no. 1, p. 012029, May 2022, doi: 10.1088/1742-6596/2273/1/012029.

[12] Q. Zhang, J. Xiao, C. Tian, J. C. Lin, and S. Zhang, “A robust deformed convolutional neural network (CNN) for image denoising,” *CAAI Trans. Intell. Technol.*, vol. 8, no. 2, pp. 331–342, Jun. 2022, doi: 10.1049/cit2.12110.

[13] J. Naskath, G. Sivakamasundari, and A. A. S. Begum, “A study on different deep learning algorithms used in deep neural nets: MLP, SOM and DBN,” *Wireless Personal Communications*, vol. 128, no. 4, pp. 2913–2936, Oct. 2022, doi: 10.1007/s11277-022-10079-4.

[14] H. Abdel-Jaber, D. Devassy, A. A. Salam, L. Hidaytallah, and M. El-Amir, “A review of deep learning algorithms and their applications in healthcare,” *Algorithms*, vol. 15, no. 2, p. 71, Feb. 2022, doi: 10.3390/a15020071.