Asian Journal of Science, Engineering and Technology (AJSET)
eISSN: 3006-6980 (online)
https://ideapublishers.org/index.php/ajset
Vol. 4, No. 1 (January-June), 2025, pp. 87-109
https://doi.org/10.47264/idea.ajset/4.1.6
Original Research Article

AI-based skin cancer detection algorithms: opportunities, challenges and way forward

Muhammad Shaheer[1] | Tayyaba Arooj[1] | Hannan Adeel[1] | Tahir Saleem[1*] | Inamur Rehman Rao[2]

1. Department of Computing, Hamdard University, Islamabad Campus, Islamabad, Pakistan.
2. Department of Information Technology, Hamdard University, Islamabad Campus, Pakistan.

*Corresponding Author Email: tahir.saleem@hamdard.edu.pk

Article History

Received:
17-Mar-2025

Revised:
10-May-2025

Re-revised:
12-Jun-2025

Accepted:
13-Jun-2025

Published:
23-Jun-2025

Abstract:

Skin cancer is a significant global health concern. Early and accurate detection is crucial for enhancing patient outcomes. This study conducts an in-depth literature review to identify commonly used Convolutional Neural Network (CNN) variants, datasets, and key evaluation metrics to assess their performance in classifying benign and malignant skin lesions. Widely used CNN architectures, including ResNet, EfficientNet, DenseNet, AlexNet, VGG, GoogleNet, LeNet-5, Xception, and MobileNet were implemented. A comparative analysis is conducted based on metrics such as accuracy, precision, sensitivity, recall, and F1-score, highlighting the strengths and limitations of each algorithm. The results show that VGG-16 outperforms other models with an accuracy of 97%, followed by VGG-19 and Mobilenet-v2 with 88%. Lastly, this paper highlights the trade-offs between various metrics, providing critical insights for deploying AI-based skin cancer detection algorithms in clinical practice.

Keywords: Artificial intelligence, Deep learning, Convolutional neural network, Melanoma, Skin lesions, Malignant skin lesions, Benign skin lesions, Clinical practice.

How to Cite:
Shaheer, M., Arooj, T., Adeel, H., Saleem, T., & Rao, I. R. (2025). AI-based skin cancer detection algorithms: opportunities, challenges and way forward. *Asian Journal of Science, Engineering and Technology (AJSET), 4*(1), 87-109. https://doi.org/10.47264/idea.ajset/4.1.6

1. Introduction

Skin cancer is a widespread and serious disease that affects many people across the world (Kumar *et al.*, 2013). Timely and accurate diagnosis is important to improve patient outcomes (Rahman *et al.*, 2022). Traditional diagnosis methods often rely on subjective visual assessment, which is time-consuming, expensive and may have variability (Javaid *et al.*, 2021). Moreover, many regions lack access to skilled dermatologists which usually results in delayed or missed diagnoses. This can negatively impact patient diagnosis and increase treatment costs (Das *et al.*, 2021). Cancer emerges when healthy cells undergo abnormal changes that lead to uncontrolled growth and tumour formation (Rahman *et al.*, 2022). Tumours can be classified as benign or malignant tumours. Skin cancer detection refers to techniques which are used to detect cancer using skin lesions.

AI has been used to develop various algorithms and techniques for skin cancer detection that increase accuracy and improve patient diagnostic outcomes (Kavitha *et al.*, 2023). Since many algorithms have been developed for skin cancer detection, each employs different techniques of implementation, so the industry must know the pros and cons of these algorithms so that appropriate algorithms can be chosen as per their requirement. Some popular deep learning algorithms used for skin cancer detection are ResNet, GoogLeNet, VggNet, Xception, InceptionNet, etc. Evaluating the implementation of these algorithms is a big challenge. Therefore, we will also identify metrics for the evaluation of these algorithms. The most used metrics for evaluation in the state-of-the-art include Accuracy, Precision, and Recall (Furriel *et al.*, 2024; Orhan & Yavşan, 2023).

We aim to evaluate and compare AI-based skin cancer detection algorithms based on key performance metrics to identify effective algorithms for clinical use. By assessing these algorithms against standard metrics, we seek to provide recommendations and highlight trade-offs between various metrics. This improves diagnostic accuracy, reduces healthcare costs, and makes skin cancer detection more accessible. This study focuses on AI-based skin cancer detection algorithms. An in-depth literature review has been conducted to explore the state-of-the-art. The commonly used algorithms for AI-based skin cancer detection are selected and implemented. Eventually, these algorithms are evaluated based on their performance in terms of identified performance metrics. By systematic comparison of these algorithms, this study provides recommendations for their practical application in clinical settings. This facilitates the adoption of AI-driven diagnostic tools to improve patient outcomes. Table-1 provides the symbols used and their meanings.

Table-1: Symbols and their meanings

K	Kernel size (e.g., 3 or 5)	C_{in}, C_{out}	Input/Output channels of a layer
K^2	Kernel area (e.g., 3×3 = 9)	C	Generic number of channels
H, W	Height and Width of feature map	M	Input channels in MobileNet blocks
H_l, W_l	Output dimensions at layer l	N	Output channels in MobileNet blocks
T	Expansion factor in MobileNetV2	TP	True Positive
D_k	Depth-wise kernel size (typically 3)	FP	False Positive
L	Number of layers (used in DenseNet)	FN	False Negative
Φ	Compound scaling coefficient (EfficientNet)	TN	True Negative

The rest of the paper is structured as follows: Section 2 presents the literature review. Section 3 presents an overview of 10 representative methods along with their architectural descriptions. Section 4 covers experimental results and discussion based on performance metrics, bar plots, and overall performance. Finally, Section 5 concludes the paper.

2. Review of related literature

This section provides a comprehensive review of related literature on skin cancer detection techniques, data pre-processing techniques, deep learning approaches, and other popular methods utilized for skin cancer detection. Section 2.2 outlines various data sets used in state-of-the-art studies. Section 2.3 discusses the key evaluation metrics to assess the performance of these algorithms.

2.1. State-of-the-art

Several techniques are used for the classification of skin cancer using skin lesions in state of the art, one such study proposed a technique for melanoma skin cancer detection using a hybrid feature extractor (HFF) which combines HOG, LBP, SURF, and VGG-19 based CNN techniques in (Rahman *et al.*, 2022). Furthermore, they combined a Hybrid Feature Extractor (HFE) and a VGG-19-based convolutional neural network (CNN) feature extractor for classification. This study used the HAM10000 dataset for the evaluation of the proposed model. In addition to this, they used accuracy, precision, specificity, and sensitivity as their performance metrics. The proposed model (HFF+CNN) achieved an accuracy of 99.4%.

Another study has done a comparative analysis of different AI-based algorithms (Hasan *et al.*, 2021). It evaluated VGG-16, Support Vector Machine, ResNet50, and self-built sequential models with differing layers. The results show VGG-16 has achieved the highest accuracy at 93.18%. Kumar *et al.* (2023) proposed A hybrid approach for skin lesion detection using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). They used the HAM10000 dataset for evaluation. Results of this study show that they achieved a classification accuracy of 94% for CNN and 97.3% for RNN. Although the combination of CNN and RNN shows good results in enhancing early skin disease detection, there remains a need to compare proposed models with other popular techniques to assess their overall performance and effectiveness.

Orhan and Yavşan (2023) proposed an AI-based detection model for melanoma diagnosis. They used CNN variants including AlexNet, MobileNet, ResNet, VGG-16 and VGG-19 algorithms which were evaluated on the dataset from Kaggle that contains 8598 images. The results show that MobileNet outperforms other models with 84.94% accuracy. Rezaoana *et al.* (2020) proposed a novel parallel CNN Model for skin cancer detection and classification. They used 25,780 images from Kaggle datasets. Their results show the proposed model outperforms VGG-16 and VGG-19 by achieving a precision of 76.17%, recall of 78.15%, and F1-score of 76.92%.

Mazoure *et al.* (2022) presented a novel web application named DUNEScan (Deep Uncertainty Estimation for Skin Cancer). This web application uses six CNN models which include Efficient Net, Inceptionv3, ResNet50, MobileNetv2, BYOL and SwAV. Moreover, the HAM10000 data set was used to evaluate this application. Additionally, this web application allows uploading skin lesions on which it applies different CNN models to detect cancer.

Mazoure *et al.* (2022) proposed a deep learning-based model to diagnose melanoma. Inception-V3 and InceptionResnet-V2 were used for melanoma detection. The study used the HAM10000 dataset for the evaluation of the proposed models. Moreover, the study employs an enhanced super-resolution generative adversarial network trained on around ten thousand images to produce high-quality HAM10000 dataset images. They concluded that their proposed model outperformed the current state-of-the-art.

Nawaz *et al.* (2022) proposed a novel method for automated detection of skin lesions combining faster region-based convolutional neural networks (RCNN) with Fuzzy K-means Clustering (FKM) technique. They evaluated the performance of the model on ISBI-2016, ISIC-2017, and PH2 datasets. Moreover, 95.40%, 93.1 %, and 95.6% accuracy have been achieved on the PH2, ISBI-2016, and ISIC-2017 datasets, respectively which concludes that it outperforms the state-of-the-art.

Kavitha *et al.* (2023) proposed a skin cancer segmentation model using a deep learning algorithm called Feature Pyramid Network (FPN). They used CNN architecture like ResNet34, DenseNet121, and MobileNet-v2 for segmentation and DenseNet121 was used for classification. Furthermore, the dataset they used for evaluation is HAM10000 consisting of 10,015 images of seven classes. The result shows that the proposed methodology achieved 80% accuracy with ResNet34, 70% with DenseNet121, 75% with MobileNetv2 in segmentation, and 80% accuracy in classification.

Daghrir *et al.* (2020) proposed a hybrid approach using deep learning and machine learning techniques for skin cancer detection. The dataset consists of 640 skin lesion images from the ISIC archive that were used for evaluation. Their system relied on the prediction of three different methods which include KNN, SVM and CNN trained, and they achieved an accuracy of 57.3%, 71.8%, and 85.5%, respectively. The prediction of these 3 methods was then combined using majority voting and it achieved an accuracy of 88.4%, which shows that the hybrid approach gives the highest accuracy. However, this study could have used more skin lesion images as it requires large data to effectively train the model.

Furriel *et al.* (2024) evaluated different AI-based methods and models used for skin cancer detection and classification. This study analyzed 18 papers related to skin cancer detection. It shows that the popular datasets used in state-of-the-art include HAM10000 and ISIC and the common metrics include accuracy, sensitivity, and specificity. Additionally, results show that CNNs especially ResNet has high performance in skin cancer detection. There is a need to provide a comparative analysis of algorithms based on different metrics.

Civelek and Kfashi (2022) proposed an automatic skin cancer diagnosis using Deep Convolutional Neural Network (DCNN). They used deep learning with Deep CNN and machine learning with Naive Bayes and Random Forest. Moreover, the International Skin Imaging Collaboration (ISIC) dataset which consists of 3297 images of benign and malignant have been used in this study to evaluate the proposed system. Their proposed system outperforms the state-of-the-art and has achieved an accuracy of 99.5%. However, there remains a need to further test the proposed method with different datasets.

Zafar *et al.* (2023) provided a comprehensive review of various AI-based methods for skin cancer detection and classification. This survey covers preprocessing, segmentation, feature extraction, selection, and classification methods used for recognizing skin cancer. Additionally, this survey shows that CNN outperforms traditional methods in classifying image samples and segmentation. However, this survey could only be accurate as an evaluation metric. Ghosh *et al.* (2024) proposed a hybrid deep learning model combining VGG-16 and ResNet50 to classify skin lesions. Moreover, they also employed various other deep learning models and machine learning techniques including Densenet201, InceptionV3, VGG16, and ResNet50. They used 3400 images of nine different classes for evaluation. The proposed hybrid model achieves validation accuracy of 97.50 %, and precision, recall, and F1 scores of 97%, 97%, and 97%, respectively.

Hermosilla *et al.* (2024) explores various techniques used to detect skin cancer. This systematic review shows that widely used datasets are HAM10000, ISIC, and PH2 datasets. Moreover, popular metrics used for evaluation in studies include accuracy, precision, recall (sensitivity) and specificity. This study also shows that CNN particularly ResNet outperforms other algorithms, especially when used in hybrid models. Mukherjee *et al.* (2023) proposed a methodology to develop an Android application that utilizes MobileNet-v2 architecture is proposed. They used a diverse dataset for improved accuracy of skin cancer detection. Their proposed methodology achieved an accuracy of 91.3%.

Rehman *et al.*, (2022) proposed a deep learning-based methodology for the classification of skin cancer lesions. They used transfer learning with CNN variants such as InceptionV3, MobileNetV2, and DenseNet201. Moreover, they used HAM10000 and ISIC 2017 datasets for the evaluation which included a total of 3,297 images of 2 classes i.e., malignant and benign. Additionally, the proposed methodology achieved an accuracy of 95.5%. They also used Grad-CAM visualization which enhances interpretability of the model’s predictions.

Hosny *et al.*, (2018) proposed skin cancer classification method using AlexNet architecture. The proposed methodology replaced the last layer of AlexNet with a SoftMax layer for classification and fine-tuned the network with data augmentation. Moreover, the proposed model is trained using the PH2 dataset which has images of 3 classes. Additionally, the proposed method achieved an accuracy of 98%, sensitivity of 98%, specificity of 98% and precision of 97%. Ly and Verma (2018) proposed a novel CNN model for improving skin cancer detection on mobile platforms. They trained CNN from scratch on a balanced dataset using advanced regularization techniques. The model was evaluated using a composite dataset called the PHDB melanoma dataset which consists of high-resolution skin lesions. Their proposed model achieved an accuracy of 86%.

Naeem and Anees (2024) proposed a novel multidisciplinary framework for skin cancer detection by using a combination of Xception and ResNet101 deep learning models (XR101). This methodology uses the strength of both Xception and ResNet101 to extract features and classify various skin cancer types. Their proposed framework was evaluated using ISIC, PH2, DermPt, and HAM10000 datasets. They also focused on balancing class distribution with the Bordeline-SMOTE technique. The model achieved an accuracy of 98.21%.

Guergueb and Akhloufi (2021) proposed a deep learning approach for melanoma skin cancer detection. They used various CNN architectures including VGG, ResNet, EfficientNet, and DenseNet. They trained models on datasets of around 30,000 images from public sources such as ISIC, Mini-ISIC and applied data augmentation. The results show that EfficientNetB7 outperformed others by achieving high accuracy, sensitivity, and AUC of 99.3%, 98.7%, and 99.01%, respectively.

Barman *et al.* (2022) proposed a transfer learning approach using the GoogleNet architecture for classifying dermoscopic skin lesions. Their methodology includes preprocessing images and using pre-trained models for classification. Moreover, they trained their proposed model on the ISIC dataset. The proposed model achieved an accuracy of 89.93%, precision of 78%, recall of 86%, and F1-score of 73%.

Naeem *et al.* (2020) presented systematic review of deep learning methods for melanoma classification. They identified a total of 512 studies initially and narrowed it down to 25 studies out of 5112 articles. This study shows that the most common datasets are ISIC 2016, 2017, 2018, 2019, 2020, PH2, and DermIS. Furthermore, the algorithms used in these studies are CNN architectures which include AlexNet, VGG-16, ResNet-50, DenseNet, and EfficientNet. They also highlighted that ensemble methods such as pre-trained CNN models outperform. Additionally, the systematic review shows that ensemble methods such as networks like ResNet and DenseNet have achieved high accuracy for melanoma classification.

Ghosh *et al.* (2022) proposed A novel deep learning method called SkinNet, based on CNN architecture. Their proposed methodology includes a pre-processing pipeline that has various pre-processing steps such as digital hair removal, background noise reduction and various filtering techniques such as non-local means of de-noising Gaussian filtering. ISIC and HAM10000 datasets were used for the evaluation of SkinNet-16. Their proposed model has achieved 99.9% accuracy % in skin cancer detection.

2.1. Datasets

Based on the review, the commonly used datasets for skin cancer detection and classification datasets include HAM10000, the dataset from the ISIC archive, and PH2 (Hermosilla *et al.*, 2024). HAM10000 (Human Against Machine) dataset is one of the largest and most used databases for skin lesion classification (Hussein & Abdulazeez, 2024). It contains 11,720 images and is sourced from the International Skin Imaging Collaboration (ISIC) archive. This dataset includes seven different classes of skin lesions (ISIC Archive, 2023). ISIC-2017 was developed for image analysis tools that can automatically diagnose melanoma from skin lesions (ISIC, 2017). ISIC 2017 dataset consists of 2750 skin cancer images with 2000 images for training datasets, 150 images for test datasets, and 600 images for validation datasets. Additionally, the size range for ISIC-2017 is 540 × 722 to 4499 × 6748 pixels (Yilmaz *et al.*, 2021). ISIC-2018 dataset consists of 11527 images with 10,015 for training and 1512 for testing. Moreover, ISIC 2018 has 7 classes of skin lesions. It includes Melanoma, Nevi, Basal cell carcinoma, Actinic Keratosis, Benign Keratosis and Vascular (Cassidy *et al.*, 2022). PH2 dataset is collected from the Hospital Pedro Hispano in Portugal, and it consists of 200 images. Each image is a size of 768 x 560 pixels. It has 3 different types of skin lesions including Atypical Nevus, Common Nevus and Melanoma (Öztürk & Özkaya, 2020).

2.2. Performance evaluation metrics

The key performance metrics that are used to evaluate CNN-based architectures are discussed below. These include Accuracy, Precision, Recall, Specificity, F1-Score, False Positive Rate (FPR), False Negative Rate (FNR), False Discovery Rate (FDR), Area Under the Curve (AUC), and Time Complexity. The F1-Score is very useful for uneven class distribution. The FPR is associated with the underlying probability of a Type I error, and it is important to evaluate the model's performance in differentiating benign and malignant cases. FNR represents the proportion of malignant cases that were incorrectly classified as benign, also called the probability of Type II error (Hill, 2004). A high FDR means the model has a high number of false positives. The Area under the curve (AUC) is derived from the Receiver Operating Characteristic (ROC) curve (Sultana & Puhana, 2018). It plots the True Positive Rate (Sensitivity) against the False Positive Rate across different threshold levels (Oumoulylte *et al.*, 2023). AUC value of 1 shows perfect classification while a value of 0.5 shows random guessing. For the rest of the metrics, a higher score represents a better quality. Time complexity is often represented by Big-O notation. Moreover, it is an important factor in determining the efficiency of machine learning models. Time complexity is the amount of computational time that an algorithm takes to complete as the size of the input data grows. The performance metrics widely used in the literature are summarized in Table-2.

Table-2: A systematic review of evaluation metrics for skin cancer detection algorithms

Metric	Formula	Description	When It's Most Useful
Accuracy	TP + TN / (TP + TN + FP + FN)	Measures of overall correctness	When classes are balanced.
Precision	TP / (TP + FP)	Quality of positive predictions	When false positives are costly
Recall (Sensitivity)	TP / (TP + FN)	The ability of the model to capture positives correctly.	When missing positives are risky
Specificity	TN / (TN + FP)	How well the model identifies actual negatives.	When true negatives matter
F1 Score	2 × Precision × Recall / (Precision + Recall)	Harmonic means of precision and recall.	When a balance between precision and recall is needed.
FPR	FP / (FP + TN)	Measures the proportion of benign cases that were incorrectly classified as malignant.	For ROC analysis and threshold tuning.
FNR	FN / (FN + TP)	The proportion of positives is incorrectly labelled as negatives.	When false negatives are critical.
FDR	FP / (FP + TP)	The proportion of predicted positives that are false.	When minimizing false discoveries is key.

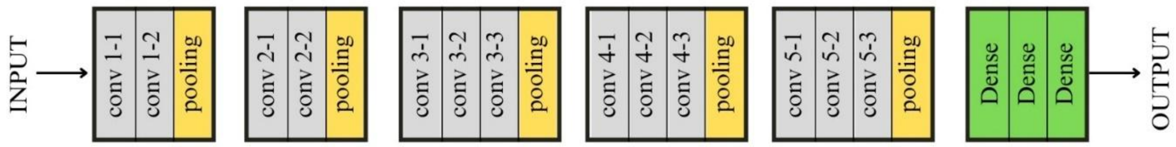
3. Overview of AI-based models

This section presents an overview of representative CNN and deep learning-based methods with their architectural details. This section also highlights their time complexity analysis (Table-3). The main components of deep learning-based models are convolutional layers, pooling layers and fully connected layers (Khashroum *et al.*, 2023). Convolutional layers are building blocks of CNNs where the network applies filters on input data. It takes an image as its input and then applies 3×3 or 5×5 filters to it. The result is called a feature map. Moreover, the padding is applied in each layer to retain the important information. Additionally, there is a stride which is the number of pixels by which the kernel moves. Pooling layers reduce spatial dimensions of feature maps (output of convolutional layers) while keeping the most important information. The activation function introduces non-linearity into the mode filters used after each convolutional layer to add non-linearity into the model. It converts all the negative values to zero. This enables networks to learn more complex patterns. Fully Connected Layers are a feed-forward neural network that takes the flattened output of the last pooling layer and uses it to make predictions (Bhatt *et al.*, 2021). These layers combine all learned features to classify the input. An overview of the representative AI-based models is presented below:

3.1. VGG-16

VGG-16 was introduced by Visual Geometry Group (VGG). VGG-16 consists of 16 weight layers, and its structure consists of thirteen convolutional layers organized in five Blocks having multiple convolutional and max-pooling layers as shown in Figure 1 (Daraghmeh, 2024). Moreover, there are 4096 channels on the first two layers, and 1000 channels in the third layer represent 1000 different label categories. All hidden layers are followed by the ReLU activation function. This process helps determine the probability that an image falls into each of the 1000 categories that VGG-16 can classify (Tao *et al.*, 2021).

Figure 1: Architecture of VGG-16 (Daraghmeh, 2024)



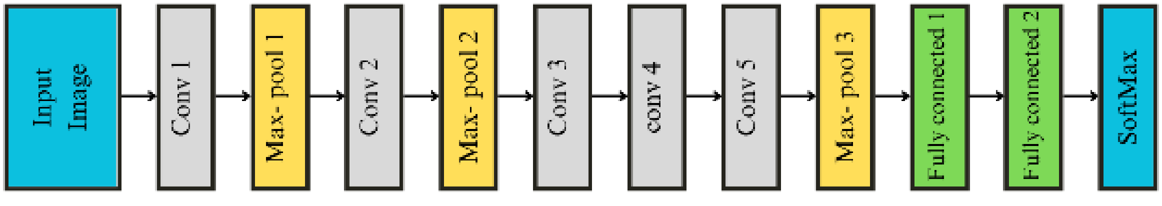
3.2. VGG-19

The VGG 19 architecture has 16 convolutional layers, 3 fully connected layers, a SoftMax layer, and 5 max-pooling layers (Mascarenhas & Agarwal, 2021). VGG-19 has an increased depth, achieved by adding more convolutional layers to the last three blocks. Each convolutional block in VGG-19 is followed by a max-pooling layer, which helps reduce the spatial dimensions, and the network ends with the same fully connected layers as VGG-16. Since VGG-19 has additional layers, it allows the model to capture even finer details (Kurek *et al.*, 2023). However, VGG-19 is computationally more expensive than VGG-16, with approximately 143 million parameters.

3.3. AlexNet

AlexNet also employed dropout in fully connected layers which randomly "drops" neurons while training to encourage the network to learn more robust features. Another innovation was the use of overlapping max-pooling which enhanced the richness of features (Krizhevsky *et al.*, 2012). The architecture of AlexNet (Figure 2) has eight layers with 5 convolutional layers followed by three fully connected layers (Krizhevsky *et al.*, 2012). The first convolutional layer uses 96 filters of size 11×11 with a stride of 4, applied to the input image. The second convolutional layer applies to 256 filters of size 5×5. The third and fourth convolution layers use 384 filters. Fifth convolutional layers use 256 filters. The output is flattened and passed through three fully connected layers.

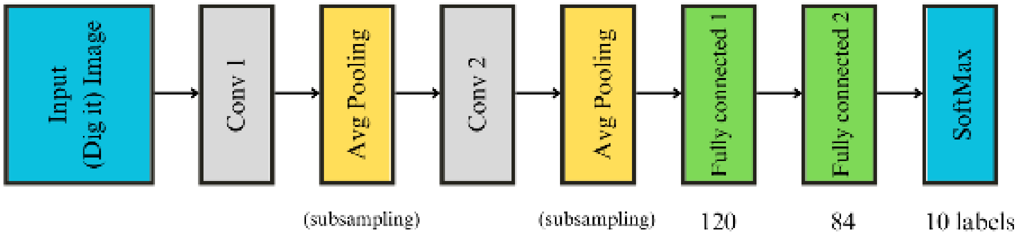
Figure 2: Architecture of AlexNet (Krizhevsky *et al.*, 2012)



3.4. LeNet-5

The architecture of LeNet-5 is relatively simple by modern standards and is very famous as it was the first CNN. LeNet-5 is a feed-forward neural network and consists of seven layers in total. It has two convolutional layers and two pooling layers, followed by three fully connected layers (Li *et al.*, 2021). The input to the network is a 32×32 grayscale image. The Convolutional Layer (C1) which consists of six 5×5 filters that produce six 28×28 feature maps is the first layer of LeNet-5 as shown in Figure 3. This layer captures low-level features like edges and simple shapes.

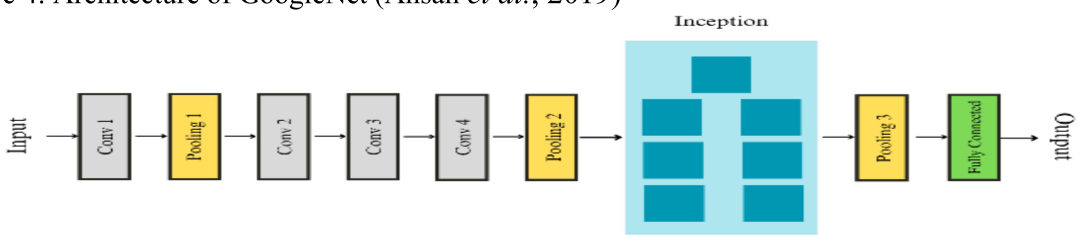
Figure 3: Architecture of LeNet-5



3.5. GoogleNet

The main goal of GoogleNet was to achieve the highest accuracy with less computational cost (Ahsan *et al.*, 2019). The architecture of GoogleNet is 22 layers deep. The inception module used in it allows GoogleNet to detect detailed features with reduced computational cost addressing the issue of choosing the optimal filter size at each layer (Khan *et al.*, 2020). The 1×1 convolution within the module helps reduce the dimensionality and lower computational complexity. They also enable networks to learn more intricate patterns by combining multiple feature maps (Szegedy *et al.*, 2015). The architecture of GoogleNet is shown in Figure 4. To address the vanishing gradient problem, GoogLeNet includes two auxiliary classifiers which are connected to intermediate layers (Ghimire *et al.*, 2021). These auxiliary classifiers act as additional sources of gradient flow during training which can help network converge more effectively. Rather than using fully connected layers as we have seen in earlier CNN architectures, GoogLeNet uses a global average pooling layer before the final output (Qureshi *et al.*, 2022). This layer averages the spatial dimensions of feature maps which produce a compact 1x1 feature vector for each class.

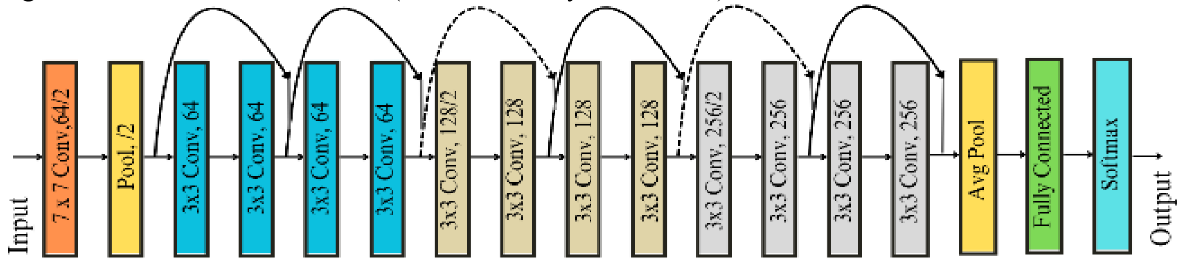
Figure 4: Architecture of GoogleNet (Ahsan *et al.*, 2019)



3.6. ResNet

ResNet employed residual learning that can handle the vanishing gradient problem (Alaeddine & Jihene, 2021). This problem occurs when gradients become too small while back-propagation which makes it difficult to update the weights effectively in very deep networks. In a traditional neural network, each layer learns a function that directly maps input to output. As networks grow deeper, it becomes difficult to optimize this mapping. ResNet addresses this by introducing residual blocks that allow each layer to learn the difference between input and desired output, instead of directly learning output (Kanavos & Mylonas, 2023).

Figure 5: Architecture of ResNet (Kanavos & Mylonas, 2023)

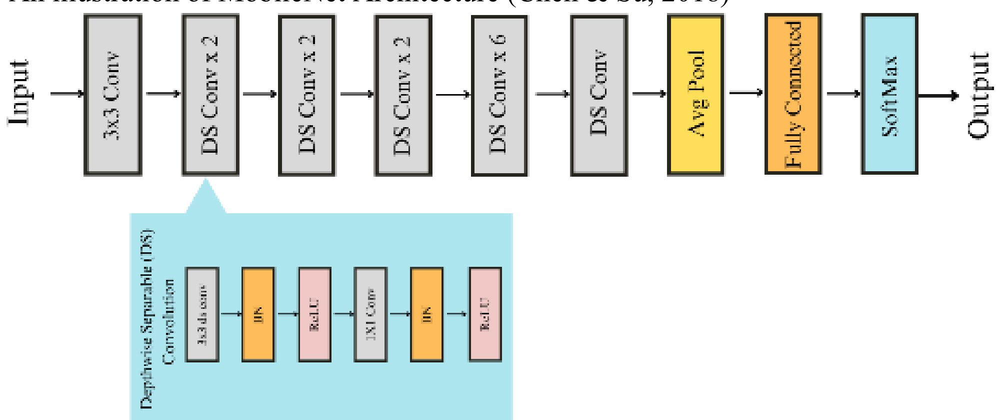


The skip connections help preserve gradients' flow during backpropagation, making it feasible to train deep networks without encountering the vanishing gradient problem. A series of residual blocks grouped into stages where each stage is responsible for learning features at different levels of abstraction. The identity shortcut directly passes input to output, while the convolutional shortcut involves a 1×1 convolution operation. This approach allows ResNet to handle cases where input and output have different dimensions which ensures shortcut connections are valid. ResNet blocks often include batch normalization and ReLU (Rectified Linear Unit) activation functions, which further stabilize training and improve the network's ability to learn complex patterns. ResNet typically uses a global average pooling layer, similar to GoogLeNet, before feeding them into a fully connected layer that produces the output classification. The global average pooling layer helps to minimize the number of parameters and avoid overfitting while also ensuring that the network remains computationally efficient. ResNet architectures have different variants like Resnet-50, 101, and 152 (Khan *et al.*, 2020).

3.7. MobileNet

MobileNet introduces depth-wise separable convolutions, a technique that can reduce the number of parameters and computational complexity which makes it good for environments with restricted computational power (Chen & Su, 2018). The architecture of MobileNet is shown in Figure 6. Following the depth-wise convolution, MobileNet uses pointwise convolution. MobileNet V2 introduced several enhancements, including the inverted residual block with linear bottlenecks. MobileNet V3 refines architecture by incorporating advances like the swish activation function and Squeeze-and-Excitation (SE) modules. These SE modules enhance the representational power of the network. MobileNet V3 uses NAS (Neural Architecture Search) to automatically discover and optimize the model architecture, striking a better balance between latency and accuracy across a range of mobile devices.

Figure 6: An illustration of MobileNet Architecture (Chen & Su, 2018)

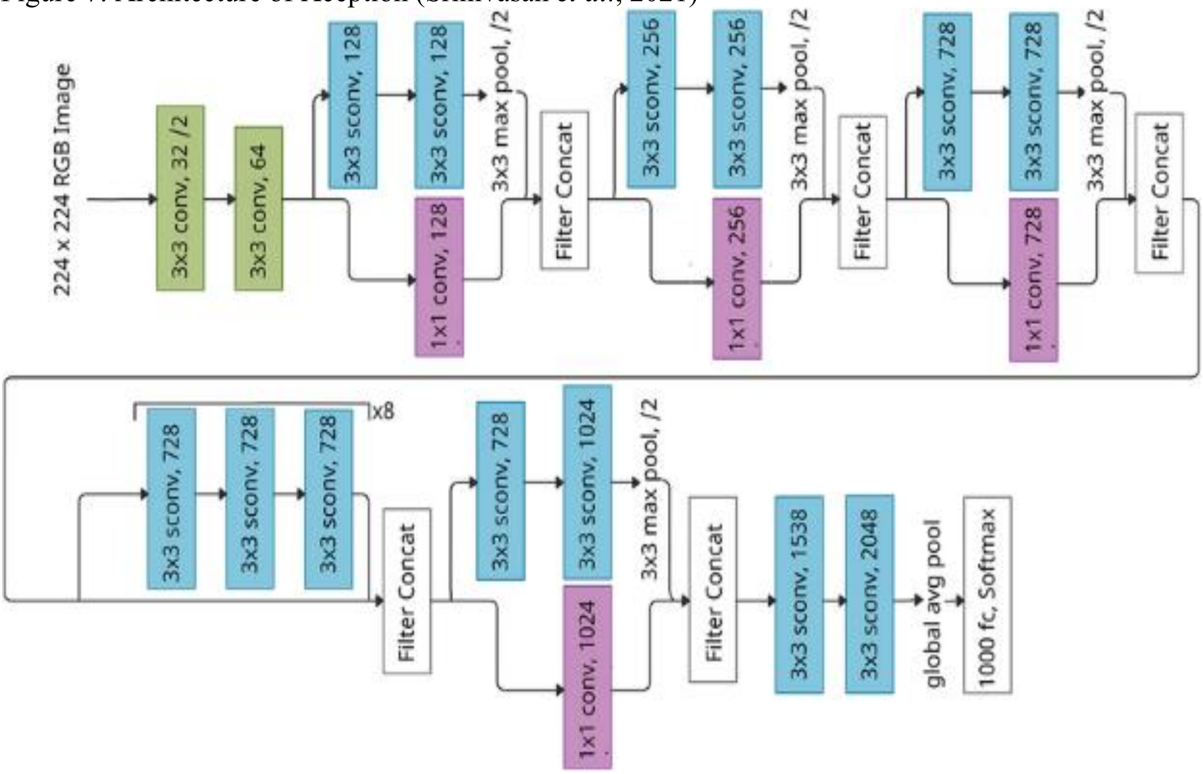


3.8. Xception

The core idea behind Xception is to replace the Inception modules with depth-wise separable convolutions (Khan *et al.*, 2020). The separation of spatial and cross-channel convolutions reduce computational complexity significantly while allowing the network to learn more nuanced and fine-grained features. The Xception architecture (Figure 7) consists of 36 convolutional layers organized into 14 modules, with each module containing one or more depth-wise separable convolution layers (Lo *et al.*, 2019).

Xception uses residual connections (Qureshi *et al.*, 2022). In Xception, residual connections are employed across most modules to allow for more efficient gradient flow. In the final stages, Xception employs global average pooling (Pathakkan *et al.*, 2022). This combines each feature map into a single average value which drastically reduces the number of parameters and prevents overfitting. The result of this layer is then fed to SoftMax which gives the final predictions.

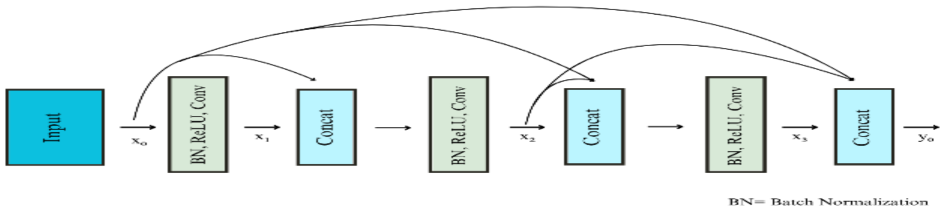
Figure 7: Architecture of Xception (Srinivasan *et al.*, 2021)



3.9. DenseNet

DenseNet has a unique connectivity pattern. In DenseNet, each layer is directly connected to every other layer in a feed-forward manner to improve information flow between layers. In a DenseNet, each layer receives inputs from all preceding layers and passes on its output to all subsequent layers (Huang *et al.*, 2017) (as shown in Figure 8). This is achieved by merging feature maps from earlier layers, rather than summing them up as done in traditional residual networks like ResNet. As a result, the input to any given layer includes not only the raw input data but also considers the feature maps from preceding layers that provide the network with a rich set of features at each stage. This dense connectivity allows a more efficient flow of information and gradients throughout the network.

Figure 8: Architecture of DenseNet (Huang *et al.*, 2017)



3.10. EfficientNet

The key innovation of EfficientNet is the development of a compound scaling method which balances the depth, width, and resolution of the model systematically and efficiently (Tan, 2019; Oza *et al.*, 2022). The architecture of EfficientNet is built upon MobileNetV2's inverted residual blocks (Tan, 2019). These blocks have expansion phases, depth-wise convolution, and projection phases, which help the network learn more complex features without a significant increase in computational cost.

EfficientNet-B0, the smallest model in the EfficientNet family, starts with a 224x224 input image size and uses these inverted residual blocks throughout its architecture. One of its most impressive features is its scalability across a wide range of model sizes (Bhargavi *et al.*, 2023). The architecture of EfficientNet is shown in Figure 9.

Figure 9: Architecture of EfficientNet (Tan, 2019)

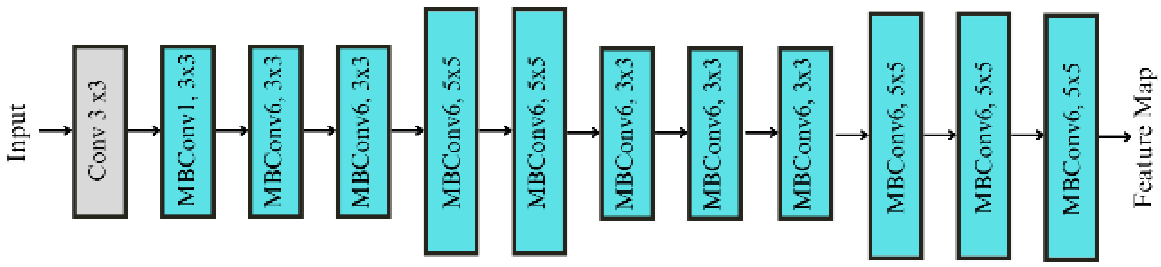


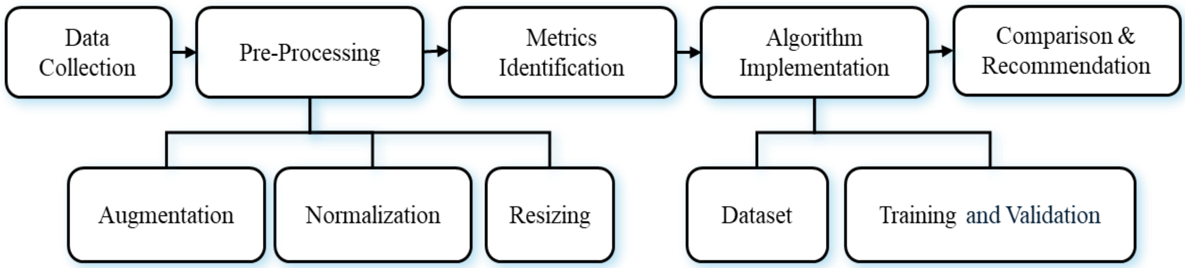
Table-3: Comparison of time complexities for AI-based models for skin cancer detection

Model	Big-O Notation	FLOPs (GFLOPs)	Relative Complexity	Ordinal Complexity
Generic CNN	$O(\sum H_l \cdot W_l \cdot K_l^2 \cdot C_{inl} \cdot C_{outl})$	~0.02–0.1	~10–50×	◇ Very Low
LeNet-5	$O(\sum H \cdot W \cdot C_{in} \cdot K^2 \cdot C_{out})$	~0.002	1×	◇ Very Low
MobileNet-V2	$O(\sum H \cdot W \cdot (tM + D_k^2 \cdot tM + tM \cdot N))$	~0.3	150×	◆ Low
AlexNet	$O(\sum H \cdot W \cdot C_{in} \cdot K^2 \cdot C_{out})$	~0.7	350×	◆ Medium
GoogleNet	$O(\sum H \cdot W \cdot C^2 \cdot K^2)$	~1.5	750×	◆ Medium
ResNet-50	$O(\sum H \cdot W \cdot C^2 \cdot K^2)$	~4.1	2050×	▲ High
DenseNet-121	$O(L^2 \cdot H \cdot W \cdot K^2 \cdot C)$	~2.9	1450×	▲ High
VGG-16	$O(\sum H \cdot W \cdot C^2 \cdot K^2)$	~15.3	7650×	▼ Very High
VGG-19	$O(\sum H \cdot W \cdot C^2 \cdot K^2)$	~19.6	9800×	▼ Very High
Xception	$O(\sum H \cdot W \cdot (C_{in} \cdot D_k^2 + C_{in} \cdot C_{out}))$	~8.4	4200×	▲ High
EfficientNet	$O(\phi \cdot H \cdot W \cdot C_{in} \cdot K^2 \cdot C_{out})$	~0.39	195×	◆ Low

4. Experimental results and discussion

This section presents the experiments performed for skin cancer detection. The workflow for the study at hand is represented in Figure 10.

Figure 10: A block diagram representation of the study workflow



For this study, we used the ISIC (International Skin Imaging Collaboration) dataset (Abdulazeez *et al.*, 2024). It consists of 3297 skin lesion images and two classes i.e. benign and malignant. The benign class has 1440 training images and 360 testing images while the malignant class has 1197 training images and 300 testing images.

The representative algorithms were implemented in Python with TensorFlow and Keras libraries. These algorithms include DenseNet, AlexNet, VGG-16, VGG-19, MobileNet, Xception, EfficientNet, LeNet-5, ResNet and GoogLeNet. After implementing the selected algorithms their performance is evaluated using nine (9) performance metrics. Figure 11 shows a detailed bar plot analysis of the implemented algorithms based on these metrics. A quantitative comparison of the state-of-the-art architectures for skin cancer detection is presented in Table 4.

Figure 11(a) depicts that VGG-16 achieved the highest accuracy of 97%, followed by VGG-19 with 88%, and MobileNet-V2 with 88%. These results show that VGG-16 is the most reliable in terms of making correct predictions. Other CNN architectures such as GoogLeNet and Xception also performed well with an accuracy of 82%. However, models like Lenet-5 and EfficientNet show poor performance as they achieved an accuracy of 51% and 55%, respectively.

As shown in Figure 11(b), MobileNet-V2 achieved the highest precision of 88%, closely followed by VGG-16, VGG-19 and AlexNet. Additionally, VGG-16, VGG-19 and AlexNet performed equally well by achieving a precision of 87%. These models show consistent precision which indicates that they can be trusted to avoid false positives. However, EfficientNet had a precision score of only 27% which shows that it is highly unreliable. Lenet-5 also performed poorly as it achieved a precision of only 50%.

As shown in Figure 11(c), VGG-16 again had the best recall score of 87%, followed closely by MobileNetv2 and VGG-19 with a recall of 88% and 87% respectively. GoogLeNet and Xception also performed well with 82% recall each. However, Lenet-5 and EfficientNet performed poorly as they achieved only 50% recall which indicates that they have failed to detect a huge portion of malignant cases which may result in missed diagnoses in real-world applications.

Furthermore, MobileNet-V2 stands out with the highest F1-score of 88% as you can see in Figure 11(d). It shows MobileNet-V2's balanced performance in both precision and recall. VGG-16 and VGG-19 follow closely, both have achieved an F1 score of 87% each.

EfficientNet with its F1 score of just 35% shows poor performance as compared to other architectures. It also shows that it does not perform well in both precision and recall. Lenet-5 also performed poorly here with only 50%.

It can be seen in Figure 11(e), DenseNet has achieved a specificity of 90%, followed by VGG-19 at 88%, and VGG-16 at 87%. This shows that these algorithms excel in correctly identifying benign cases. Lenet-5 performed the worst, with a specific score of 63% which means that it frequently misclassified benign cases as malignant cases.

DenseNet performed well by achieving an FPR score of 10% as shown in Figure 11(f). Followed by VGG-19 and VGG-16 with a False Positive Rate of 12% and 13%, respectively. These models performed well in minimizing false positives. Additionally, EfficientNet had the lowest FPR of 0%, but we have to be careful when considering this result since the model's overall performance was not good in other metrics. On the other hand, Lenet-5 had the highest FPR at 37% which means a very high chance of misclassifying benign cases and could lead to overdiagnosis.

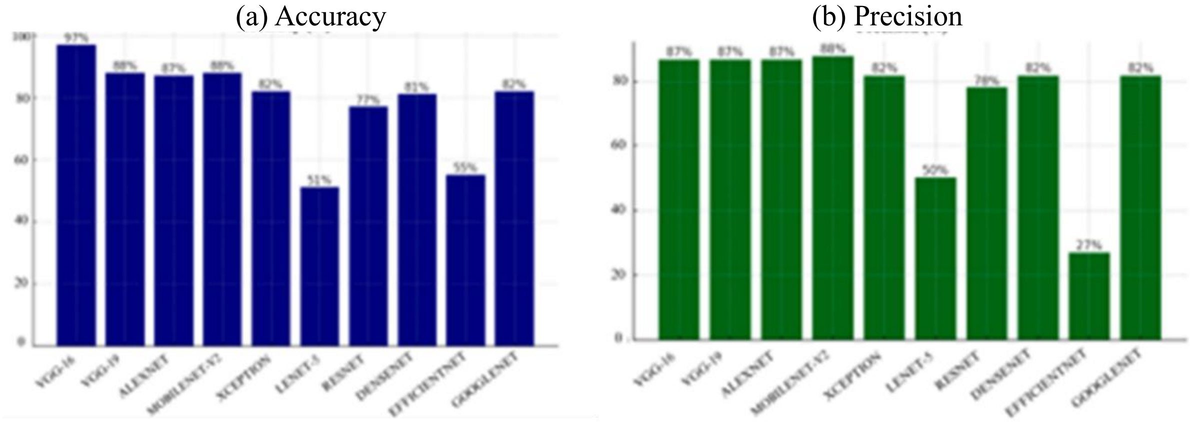
Figure 11(g) shows that AlexNet and MobileNetv2 both performed very well and achieved a low FNR of 10% each. Closely followed by VGG-16, VGG-19 and ResNet with an FNR of 13% each. This shows they can be a good choice for effectively identifying malignant cases. On the other hand, EfficientNet and Lenet-5 achieved the highest FNR of 100% and 64%, respectively. This highest percentage makes EfficientNet and Lenet-5 a poor choice for skin cancer detection.

As shown in Figure 11(h), MobileNetv2 again performed well with an FDR of only 10%, closely followed by VGG-19 which achieved a False Discovery Rate (FDR) of 14%. These models show robustness in minimizing false discoveries. Additionally, EfficientNet had the lowest FDR of 0%, however, care should be taken when considering this result, since the model's overall performance was not good in other metrics. Moreover, Lenet-5 had a high FDR of 55% which further shows its unreliability in this study.

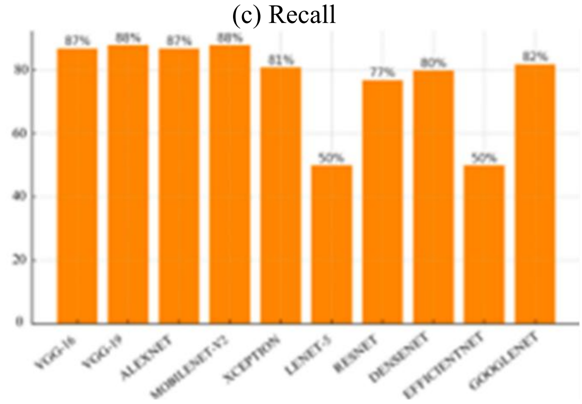
As shown in Figure 11(i), both VGG-19 and MobileNet-v2 achieved the highest AUC scores of 88%. Followed by VGG-16 and AlexNet with an AUC of 87%. These models demonstrate high effectiveness in differentiating between malignant and benign cases which makes them highly reliable. On the other hand, EfficientNet and Lenet-5 had the lowest AUC of 50% which further shows their unreliability.

Figure 11: A bar-plot analysis for state-of-the-art AI-based skin cancer detection algorithms

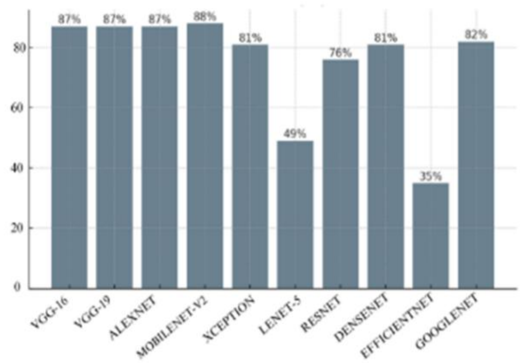
(a) Accuracy (b) Precision



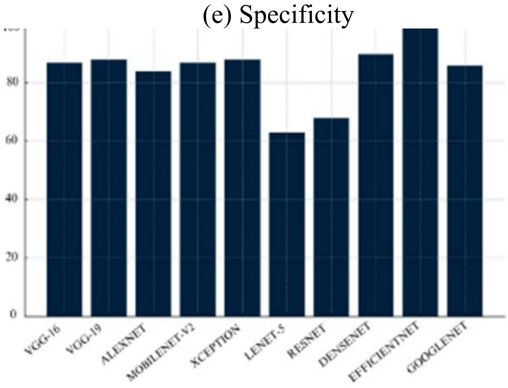
(c) Recall



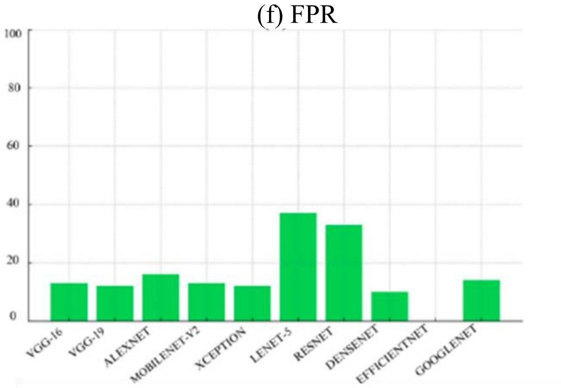
(d) F1 score



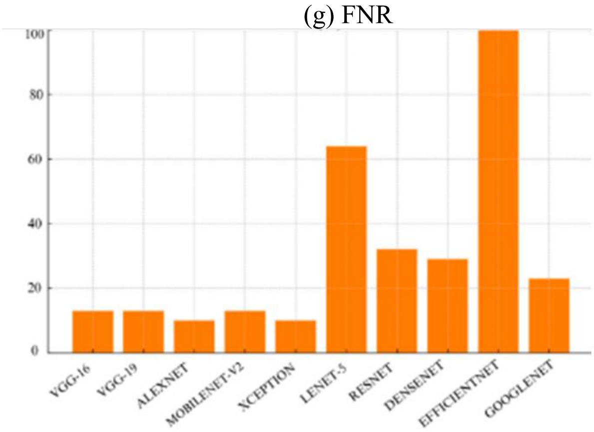
(e) Specificity



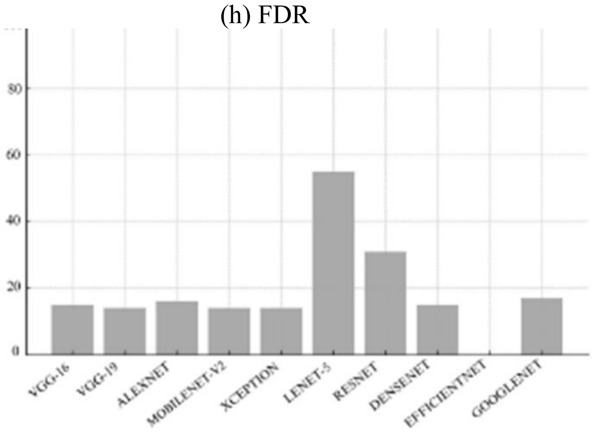
(f) FPR



(g) FNR



(h) FDR



(i) AUC

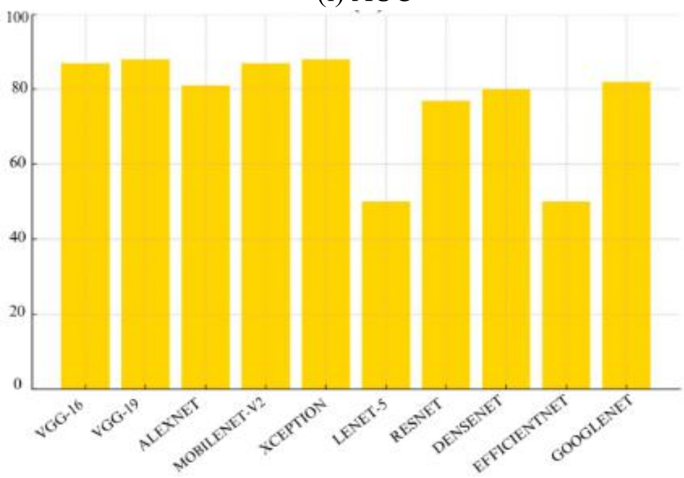


Table-4 lists the quantitative results to consider the overall performance. It can be observed that certain CNN architectures perform well in specific metrics. For instance, VGG-16 performed well across various metrics including accuracy, Precision, recall, and F1-score. VGG-19 also has shown good performance in both precision and AUC. This means it's good at distinguishing benign and malignant cases without too many false positives.

MobileNet-V2 also performed well across many metrics which include Precision, Recall, F1-score, FNR, FDR, and AUC. This makes MobileNet-V2 a reliable choice for skin cancer detection and classification. Additionally, DenseNet achieved the lowest False Positive Rate (FPR) and False Discovery Rate (FDR), which makes it highly reliable for avoiding unnecessary treatments.

Meanwhile, Lenet-5 consistently struggled across almost all metrics which shows that it is not suitable for skin cancer detection and classification. Overall, the choice of algorithm depends on specific context and priorities in a medical setting. It depends on whether you prioritize avoiding missed diagnoses (recall) or reducing false positives (specificity).

Table-4: Performance Comparison of CNN architectures

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)	FPR (%)	FNR (%)	FDR (%)	AUC (%)
VGG-16	97	87	87	87	87	13	13	15	87
VGG-19	88	87	88	87	88	12	13	14	88
AlexNet	87	87	87	87	84	16	10	18	87
MobileNet	88	88	88	88	87	13	10	10	88
Xception	82	82	81	81	88	12	26	16	81
LeNet-5	51	50	50	49	63	37	64	55	50
ResNet	77	78	77	76	68	33	13	31	77
DenseNet	81	82	80	81	90	10	29	15	80
EfficientNet	55	27	50	35	100	0	100	0	50
GoogleNet	82	82	82	82	86	14	23	17	82

5. Conclusion

Based on our evaluation, we can conclude that MobileNetv2, VGG-16 and VGG-19 were top-performing CNN architectures in this study. MobileNetv2 performed well across many metrics which include Precision, Recall, F1- score, FNR, FDR, and AUC. It achieved the highest Precision, Recall, F1- score and AUC scores of 88% each. VGG-16 achieved the highest accuracy of 97%. VGG-19 also performed well with 88% accuracy and the highest recall and AUC of 88% which shows its strong ability to differentiate between malignant and benign cases. Furthermore, DenseNet also demonstrated strong performance, especially in False Positive Rate (FPR) and False Discovery Rate (FDR). It achieved a specificity of 90% and the lowest FPR of 10%. However, models like LeNet-5 and EfficientNet performed poorly across most metrics. Therefore, they are not recommended for practical skin cancer detection.

In a practical clinical setting, the choice of model would depend on the priorities of the healthcare provider. For instance, MobileNetv2, VGG-16 and VGG-19 with a recall of 88%,88% and 87%, respectively, would be an excellent choice if detecting malignant cases (recall) is prioritized. However, if avoiding false positives (specificity) is more critical than DenseNet with 90% specificity or VGG-19 with 88% specificity would be more appropriate choices.

Overall, the performance of these models shows promising results for improving skin cancer detection and diagnosis using AI-based algorithms. However, careful selection of algorithms is important to optimize diagnostic accuracy. Furthermore, the choice of algorithm may depend on various factors such as specific clinical requirements or computational resources, etc. Additionally, challenges such as data privacy, model interpretability, ensuring robust performance under varied conditions, and managing computational costs are important in deploying deep learning models for skin cancer detection. To address these challenges, robust countermeasures like data augmentation techniques, model regularization and interpretability methods can be implemented. Overall, AI-based skin cancer detection algorithms have huge potential to support clinical decision-making and improve patient outcomes.

Declaration of conflict of interest

The author(s) declared no potential conflicts of interest(s) with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD:

Muhammad Shaheer https://orcid.org/0000-0003-1928-5926

Tayyaba Arooj https://orcid.org/0009-0000-6273-7938

Hannan Adeel https://orcid.org/0000-0002-5348-1237

Tahir Saleem https://orcid.org/0000-0001-7828-5382

Inamur Rehman Rao https://orcid.org/0000-0002-9180-9758

Publisher's Note

IDEA Publishers Group (AJSET IDEA-PG) stands neutral with regard to the jurisdictional claims in the published maps and the institutional affiliations.

References

Ahsan, A. S. M. M. I., Alif, S. M., Kibria, J. B., & Gomes, P. E. (2019). *Detection of skin cancer using convolutional neural network* [Master's thesis, Brac University].

Alaeddine, H., & Jihene, M. (2021). Deep residual network in network. *Computational Intelligence and Neuroscience, 2021*, 6659083. https://doi.org/10.1155/2021/6659083

Alwakid, G., Gouda, W., Humayun, M., & Jhanjhi, N. (2023). Diagnosing melanomas in dermoscopy images using deep learning. *Diagnostics, 13*(10), 1815. https://doi.org/10.3390/diagnostics13101815

Barman, S., Biswas, M. R., Marjan, S., Nahar, N., Hossain, M. S., & Andersson, K. (2022). Transfer learning based skin cancer classification using GoogleNet. In *International Conference on Machine Intelligence and Emerging Technologies* (pp. 238–252). https://doi.org/10.1007/978-3-031-24801-6_18

Bhargavi, M., Renugadevi, R., Sivabalan, S., Phani, P., Ganesh, J., & Bhanu, K. (2023). Ensemble learning for skin lesion classification: A robust approach for improved diagnostic accuracy (ELSLC). In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 390–395). https://doi.org/10.1109/ICIMIA60377.2023.10425943

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics, 10*(20), 2470. https://doi.org/10.3390/electronics10202470

Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J., & Yap, M. H. (2022). Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Medical Image Analysis*, 75, 102305. https://doi.org/10.1016/j.media.2021.102305

Chen, H. Y., & Su, C. Y. (2018). An enhanced hybrid MobileNet. In *2018 9th International Conference on Awareness Science and Technology (ICAST)* (pp. 308–312). https://doi.org/10.1109/ICAwST.2018.8517216

Civelek, Z., & Kfashi, M. (2022). An improved deep CNN for an early and accurate skin cancer detection and diagnosis system. *International Journal of Engineering Research and Development, 14*(1), 721–734.

Daghrir, J., Tlig, L., Bouchouicha, M., & Sayadi, M. (2020). Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. In *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1–5). https://doi.org/10.1109/ATSIP49321.2020.9231837

Das, K., Cockerell, C. J., Patil, A., Pietkiewicz, P., Giulini, M., Grabbe, S., & Goldust, M. (2021). Machine learning and its application in skin cancer. *International Journal of Environmental Research and Public Health, 18*(24), 13409. https://doi.org/10.3390/ijerph182413409

Daraghmeh, O. F. S. (2024). *A hybrid artificial intelligence approach for early detection of breast cancer and classification from mammogram images in Palestine* [Doctoral dissertation, Al-Quds University].

Furriel, B. C., Oliveira, B. D., Prôa, R., Paiva, J. Q., Loureiro, R. M., Calixto, W. P., Reis, M. R., & Giavina-Bianchi, M. (2024). Artificial intelligence for skin cancer detection and classification for clinical environment: A systematic review. *Frontiers in Medicine, 10*, 1305954. https://doi.org/10.3389/fmed.2023.1305954

Ghosh, P., Azam, S., Quadir, R., Karim, A., Shamrat, F. J. M., Bhowmik, S. K., Jonkman, M., Hasib, K. M., & Ahmed, K. (2022). SkinNet-16: A deep learning approach to identify benign and malignant skin lesions. *Frontiers in Oncology, 12*, 931141. https://doi.org/10.3389/fonc.2022.931141

Ghosh, H., Rahat, I. S., Mohanty, S. N., Ravindra, J., & Sobur, A. (2024). A study on the application of machine learning and deep learning techniques for skin cancer detection. *International Journal of Computer and Systems Engineering, 18*(1), 51–59.

Ghimire, P., Piya, S., & Gurung, A. M. (2021). Comparative study of face mask recognition using deep learning and machine learning classifiers. In *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)* (pp. 1–9). https://doi.org/10.1109/ICSES52305.2021.9633877

Guergueb, T., & Akhloufi, M. A. (2021). Melanoma skin cancer detection using recent deep learning models. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 3074–3077). https://doi.org/10.1109/EMBC46164.2021.9630391

Hussein, A. A., & Abdulazeez, A. M., (2024). A review on utilizing machine learning classification algorithms for skin cancer. *Journal of Applied Science and Technology Trends, 5*(2), 60–71. https://jastt.org/index.php/jasttpath/article/view/191

Hasan, M. R., Fatemi, M. I., Khan, M. M., Kaur, M., & Zaguia, A. (2021). Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. *Journal of Healthcare Engineering, 2021*, 5895156. https://doi.org/10.1155/2021/5895156

Hermosilla, P., Soto, R., Vega, E., Suazo, C., & Ponce, J. (2024). Skin cancer detection and classification using neural network algorithms: A systematic review. *Diagnostics, 14*(3), 454. https://doi.org/10.3390/diagnostics14030454

Hill, D. E. (2004). Errors of judgment and reporting in a law merchant system. *Theory and Decision, 56*(3), 239–267.

Hosny, K. M., Kassem, M. A., & Foaud, M. M. (2018). Skin cancer classification using deep learning and transfer learning. In *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)* (pp. 90–93). https://doi.org/10.1109/CIBEC.2018.8641792

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708). https://doi.org/10.1109/CVPR.2017.243

ISIC. (2017). Skin lesion analysis towards melanoma detection challenge. https://challenge.isic-archive.com/landing/2017/

Javaid, A., Sadiq, M., & Akram, F. (2021). Skin cancer classification using image processing and machine learning. In *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)* (pp. 439–444). https://doi.org/10.1109/IBCAST51254.2021.9393272

Kanavos, A., & Mylonas, P. (2023). Deep learning analysis of histopathology images for breast cancer detection: A comparative study of ResNet and VGG architectures. In *2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)* (pp. 1–6). https://doi.org/10.1109/SMAP59823.2023.10294321

Kavitha, S., Shalini, R., Sree, N. H., & Akash, J. (2023). Intelligent segmentation and classification for skin cancer prediction. In *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* (pp. 1–6). https://doi.org/10.1109/ICAECA56562.2023.10199876

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review, 53*(8), 5455–5516. https://doi.org/10.1007/s10462-020-09825-6

Khashroum, Z., Rahimighazvini, H., & Bahrami, M. (2023). Applications of machine learning in power electronics: A specialization on convolutional neural networks. *ENG Transactions, 4*(1), 1–5.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097–1105.

Kumar, R. R., Varun, R., Sreeshwan, J., Kumar, K. A., Rana, U., & Rajyalakshmi, A. (2023). Feasible skin lesion detection using CNN and RNN. In *Proceedings of the E3S Web of Conferences (Vol. 430, 01050). EDP Sciences*.

Kurek, J., Szymanowski, K., Chmielewski, L., & Orłowski, A. (2023). Advancing chipboard milling process monitoring through spectrogram-based time series analysis with convolutional neural network using pretrained networks. *Machine Graphics and Vision, 32*(1), 89–108.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems, 33*(12), 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Lo, W. W., Yang, X., & Wang, Y. (2019). An Xception convolutional neural network for malware classification with transfer learning. In *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (pp. 1–5). https://doi.org/10.1109/NTMS.2019.8763836

Ly, P., Bein, D., & Verma, A. (2018). New compact deep learning model for skin cancer recognition. In *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 255–261). https://doi.org/10.1109/UEMCON.2018.8796659

Mascarenhas, S., & Agarwal, M. (2021). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification. In *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)* (Vol. 1, pp. 96–99).

Mazoure, B., Mazoure, A., Bédard, J., & Makarenkov, V. (2022). DUNEScan: A web server for uncertainty estimation in skin cancer detection with deep neural networks. *Scientific Reports, 12*(1), 179. https://doi.org/10.1038/s41598-021-03898-1

Mukherjee, N., Dolzake, N., Ubhare, H., Sahu, S., Sharma, S., & Shirdhankar, S. (2023). Melanoma espial employing deep learning applied to MobileNet. *Journal of Artificial Intelligence in Medicine, 15*(2), 123–135.

Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R. A., Rehman, A., Iqbal, M., & Saba, T. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. *Microscopy Research and Technique, 85*(1), 339–351. https://doi.org/10.1002/jemt.23908

Naeem, A., Farooq, M. S., Khelifi, A., & Abid, A. (2020). Malignant melanoma classification using deep learning: Datasets, performance measurements, challenges and opportunities. *IEEE Access, 8*, 110575–110597. https://doi.org/10.1109/ACCESS.2020.2997831

Naeem, A., & Anees, T. (2024). A multiclassification framework for skin cancer detection by the concatenation of Xception and ResNet101. *Journal of Computing & Biomedical Informatics, 6*(1), 205–227.

Orhan, H., & Yavşan, E. (2023). Artificial intelligence-assisted detection model for melanoma diagnosis using deep learning techniques. *Mathematical Modelling and Numerical Simulation with Applications, 3*(2), 159–169.

Oumoulylte, M., Alaoui, A. O., Farhaoui, Y., El Allaoui, A., & Bahri, A. (2023). Convolutional neural network-based approach for skin lesion classification. *Data & Metadata, 2*, 171.

Ozturk, Ş., & Özkaya, U. (2020). Skin lesion segmentation with improved convolutional neural network. *Journal of Digital Imaging, 33*(4), 958–970. https://doi.org/10.1007/s10278-020-00343-z

Oza, P. R., Sharma, P., & Patel, S. (2022). A transfer representation learning approach for breast cancer diagnosis from mammograms using EfficientNet models. *Scalable Computing:*

Practice and Experience, 23(1), 51–58.

Panthakkan, A., Anzar, S., Al Mansoori, S., Mansoor, W., & Al Ahmad, H. (2022). A systematic comparison of transfer learning models for COVID-19 prediction. *Intelligent Decision Technologies, 16*(3), 557–574. https://doi.org/10.3233/IDT-220038

Qureshi, M. N., Umar, M. S., & Shahab, S. (2022). A transfer-learning-based novel convolution neural network for melanoma classification. *Computers, 11*(5), 64. https://doi.org/10.3390/computers11050064

Rahman, M. M., Nasir, M. K., Nur, A., Khan, S. I., Band, S., Dehzangi, I., Beheshti, A., & Rokny, H. A. (2022). Hybrid feature fusion and machine learning approaches for melanoma skin cancer detection. *Journal of Healthcare Engineering, 2022*, 123456.

Rehman, M. Z. U., Ahmed, F., Alsuhibany, S. A., Jamal, S. S., Ali, M. Z., & Ahmad, J. (2022). Classification of skin cancer lesions using explainable deep learning. *Sensors, 22*(18), 6915. https://doi.org/10.3390/s22186915

Rezaoana, N., Hossain, M. S., & Andersson, K. (2020). Detection and classification of skin cancer by using a parallel CNN model. In *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)* (pp.380–386). IEEE.

Srinivasan, K., Garg, L., Datta, D., Alaboudi, A. A., Jhanjhi, N., Agarwal, R., & Thomas, A. G. (2021). Performance comparison of deep CNN models for detecting driver's distraction. Sensors, 21(12), 4130. https://doi.org/10.3390/s21124130

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–9). IEEE.

Sultana, N. N., & Puhan, N. B. (2018). Recent deep learning methods for melanoma detection: A review. In *Mathematics and Computing: 4th International Conference, ICMC* (pp. 118–132). Springer. https://doi.org/10.1007/978-981-13-0023-3_12

Tan, M. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.

Tao, J., Gu, Y., Sun, J., Bie, Y., & Wang, H. (2021). Research on VGG16 convolutional neural network feature classification algorithm based on transfer learning. In *2021 2nd China International SAR Symposium (CISS)* (pp. 1–3). IEEE.

Yilmaz, A., Kalebasi, M., Samoylenko, Y., Guvenilir, M. E., & Uvet, H. (2021). Benchmarking of lightweight deep learning architectures for skin cancer classification using ISIC 2017 dataset. *arXiv preprint arXiv:2110.12270*.

Zafar, M., Sharif, M. I., Sharif, M. I., Kadry, S., Bukhari, S. A. C., & Rauf, H. T. (2023). Skin lesion analysis and cancer detection based on machine/deep learning techniques: A comprehensive survey. *Life, 13*(1), 146. https://doi.org/10.3390/life13010146