Network Modeling Analysis in Health Informatics and Bioinformatics (2025) 14:70
https://doi.org/10.1007/s13721-025-00569-3

ORIGINAL ARTICLE

# Improving skin lesion classification using k-means, transfer learning and mobilenet architecture

**Akinbowale Nathaniel Babatunde[1] · Bukola Fatimah Balogun[2] · Ayodele Joseph Babatunde[3] · Salman Olatunji Isiaka[4] · Joseph Bamidele Awotunde[5] · Abidemi Emmanuel Adeniyi[6,7]**

Received: 31 March 2025 / Revised: 9 July 2025 / Accepted: 12 July 2025
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2025

## Abstract

The accurate and early classification of skin lesions remains a critical concern in dermatological diagnostics due to the visual similarity between benign and malignant conditions. This study proposes an enhanced skin lesion classification framework that combines K-Means image segmentation, transfer learning, and the lightweight MobileNet architecture. The HAM10000 dataset was employed, and preprocessing involved augmentation, normalization, and class balancing to improve model generalization. K-Means clustering was utilized to segment the lesion area, improving the model's focus on the region of interest. Transfer learning was applied using a pre-trained MobileNet model, which was fine-tuned on the dataset to enable efficient and accurate classification. The model was evaluated across multiple training epochs, with the best results achieved at epoch 50: 98.20% accuracy, 93.75% precision, 100% recall, and 97.10% F1-score, alongside an AUC of 98.00%. Comparative analysis demonstrates that the proposed model outperforms existing deep learning models in both diagnostic accuracy and computational efficiency. These results indicate that the integration of K-Means segmentation with MobileNet and transfer learning provides a robust solution for real-time and resource-efficient skin lesion classification in clinical applications.

**Keywords** Skin lesion classification · MobileNet · K-Means segmentation · Transfer learning · HAM10000 · Deep learning

## 1 Introduction

Real-time image processing for medical diagnostics has gained significance owing to developments in computer capabilities and machine learning methodologies. Now, various deep learning models have been created and employed in medical diagnostics due to their ability to discern patterns in digital images (Cícero et al. 2016). Convolutional neural networks (CNNs) now constitute the most efficient approach for picture classification. As a result, CNNs have enabled progress in various medical image processing applications, such as illness categorization and detection. CNN models have been utilized to distinguish malignant from benign lesions in dermoscopic pictures of human skin (Nasr-Esfahani et al. 2016). Skin cancer is the most prevalent kind of cancer in the United States. Melanoma is the deadliest form

✉ Abidemi Emmanuel Adeniyi
abidemi.adeniyi@bowen.edu.ng

Akinbowale Nathaniel Babatunde
akinbowale.babatunde@kwasu.edu.ng

Bukola Fatimah Balogun
bukola.balogun@kwasu.edu.ng

Ayodele Joseph Babatunde
linkayodele@gmail.com

Salman Olatunji Isiaka
isiakaosalman2@gmail.com

Joseph Bamidele Awotunde
awotunde.jb@uniloin.edu.ng

[1] Department of Computer Science and Cybersecurity, Kwara State University, Malete, Kwara, Nigeria

[2] Department of Computer Science, Kwara State University, Malete, Kwara, Nigeria

[3] Department of Statistics, University of Ilorin, Kwara, Nigeria

[4] Department of Computer Science, Kwara State Polytechnic, Ilorin, Kwara, Kwara, Nigeria

[5] Department of Computer Science, University of Ilorin, Ilorin, Nigeria

[6] Department of Computer Science, Bowen University, Iwo, Nigeria

[7] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Published online: 21 July 2025

of skin cancer, arising from the abnormal multiplication of pigment-producing cells that provide skin color. The American Cancer Society forecasts that by 2030, around 87,110 new melanoma cases will be identified, with an estimated 9730 deaths expected as a result of melanoma (American Cancer Society, "Key Statistics for Melanoma Skin Cancer, 2030").

The incidence of melanoma has grown by twofold and increased more swiftly than any other cancer kind, making it a significant public health issue in the United States and globally. An accurate diagnosis is crucial for managing this illness, since it enables a wider array of treatment options. Precise and timely identification of melanoma necessitates sophisticated tools and highly proficient workers. Dermoscopy is a specialized technique for obtaining high-resolution magnified images of the skin by light manipulation and the reduction of surface skin reflectance (Codella et al. 2018). However, if practitioners are insufficiently trained in recognizing melanoma via dermoscopy images, many patients will be misdiagnosed. The accuracy of expert dermatologist classification for melanoma is 65–66% (Esteva et al. 2017). As a result, there is a substantial demand for automated image processing systems to identify melanoma from dermoscopy images. However, it continues to pose complications due to the considerable diversity of melanoma in terms of appearance, texture, shape, size, and location in dermoscopy images, along with the visual similarity between melanoma and non-melanoma lesions (Yu et al. 2017).

To lessen this challenge, some researchers recommend preliminary segmentation. This seeks to dynamically isolate the region of the image that contains the melanoma. Melanomas are then identified based on the segmentation results (Celebi et al. 2007) and (Ganster et al. 2001). (Celebi et al. 2007) presented a systematic methodology for melanoma classification, beginning with automated border detection to distinguish the lesion from the adjacent skin. Subsequently, attributes about shape, color, and texture were gathered and incorporated into a tuning framework, resulting in the determination of the optical feature subset for classification objectives (Celebi et al. 2007). However, some methods demonstrate inadequate performance due to their limited discriminative ability. Due to the shortage of dermatological doctors, various research projects have been initiated to develop automated dermoscopy image processing for the detection of skin diseases. This approach may function as a diagnostic tool for evaluating dermoscopy images (Codella et al. 2018). Numerous computerized image analysis techniques are limited to low-level visual feature representations, segmentation algorithms, and conventional machine learning methods, such as k-nearest neighbors (KNN), support vector machines (SVM), and convolutional neural networks (CNNs) (Codella et al. 2018). Convolutional Neural Networks (CNNs) comprise a sequence of convolutional modules, each typically consisting of three components: convolutional layers, pooling layers, and dense (fully connected) layers. Convolutional Neural Networks (CNNs) employ convolutional filters on raw data to extract both low- and high-level features, including angular lines and face patterns, suitable for classification in various areas. This work presents an effective binary classification technique for skin cancer detection, adept at categorizing the dataset into malignant and non-cancerous classes with performance metrics akin to those of a dermatological specialist. K-means segmentation and a pre-trained MobileNet model are suggested for model building. The primary contributions of this work are outlined thus:

- Enhanced precision and resilience in the diagnosis and forecasting of skin cancer.
- Excellent adaptation skills for cancer identification.
- The investigation seeks to develop an accessible and efficient tool for automated skin cancer diagnosis, facilitating broader implementation in clinical environments and diminishing dependence on expert practitioners.

The rest of the work is organized as follows: Sect. 2 analyzes the existing literature relevant to this subject and specifies the requirements employed for labeling datasets either benign or malignant. Section 3 outlines outcomes and evaluation results. Section 4 provides a full analysis of the outcomes. Section 5 delineates the principal conclusions and proffers recommendations for subsequent research.

## 2 Related works

A wide range of scholars have conducted studies on the detection and diagnosis of many diseases, including cancer. (Inthiyaz et al. 2023) devised an automated methodology of deep learning for the categorization of skin cancers, utilizing the Xiangya-Derm dataset, comprising 150,223 images. An already trained convolutional neural network (CNN) was utilized to classify pictures of skin lesions into four categories: melanoma, eczema, psoriasis, and healthy skin, attaining an AUC of 0.875. This research was performed on a restricted dataset; hence, these results may not be applied to other datasets. Their analysis achieved an AUC of 0.87, suggesting room for improvement; the authors furthermore utilized a deep architecture, ResNet-50, which increases computing costs.

Gajera et al. (2023) developed a computerized system for melanoma diagnosis with pre-trained CNN models. They incorporated several classifiers with eight notable CNN architectures, namely AlexNet, VGG-16, VGG-19,

Inception v-3, ResNet 50, MobileNet, EfficientNet, and DenseNet121. Their examinations employed four skin cancer datasets: ISIC 2016 (Gutman et al. 2016), ISIC 2017 (Codella et al. 2018), PH2 (Mendonca et al. 2016), and HAM10000 (Tschandl et al. 2018). Optimal results were achieved by utilizing DenseNet121 as a feature extractor alongside an MLP classifier, yielding an accuracy of 98.33% and an F1 score of 0.96. The suggested methodologies were assessed using constrained datasets, consisting of 200, 900, and 2000 training images for PH2, ISIC 2016, and ISIC 2017, respectively. This made the DenseNet121 model prone to overfitting. The classification performance on the larger HAM10000 dataset was subpar compared to the other datasets.

Alenezi et al. (2023) recommended using an already trained deep residual network, trained on the ISIC 2017 and HAM1000 datasets, for skin cancer detection and classification. The researchers applied wavelet transform and pooling techniques to eliminate artifacts, such as hair, from skin lesion images. Experiments were conducted with ReLU, PReLU, Sigmoid, and Hardlim to determine the optimal activation function that maximizes classification accuracy on skin lesion datasets. The ReLU activation function achieved the highest classification accuracy of 96.91% and an F1-score of 0.95. This study demonstrates limited generalizability and reduces classification effectiveness on lesion images of diverse sizes, colors, and other variables.

Alenezi et al. (2023) introduced an algorithm combining dilation, normalization, and pooling for removing hairs from skin lesion images. The authors performed relief feature selection to pick features obtained by ResNet101 for training the SVM classifier to characterize melanoma. They also trained SVM on variables collected from AlexNet, DarkNet19, GoogleNet, SqueezeNet (Iandola et al., 2016), Xception, and MobileNetV2, yielding best accuracies of 96.15% and 97.15% on ISIC 2019 and ISIC 2020, respectively, with variables retrieved using ResNet101. The initial data set consisted exclusively of 1168 photos. Deep architectures like ResNet101 were applied for acquiring features, possibly creating overfitting owing to training on a restricted dataset; the suggested approach is confined by the time needed for choosing parameters of the SVM classifier.

Abbas and Gul (2022) proposed a NASNet-based approach for the sorting of melanoma scans inside the ISIC 2020 dataset. The study participants utilized mathematical modifications for data enhancement to improve classification accuracy. The proposed technique achieved an accuracy of 97.7% and an F1-score of 0.97.

Gouda et al. (2022) deployed ESRGAN (Gouda et al. 2022) to generate synthetic pictures, hence enhancing the set of images for training a CNN framework to classify skin lesion photographs. The CNN algorithm trained on the ISIC 2018 dataset achieved an accuracy of 83.2%, equivalent to that of more complex networks such as ResNet-50, InceptionV3, and Inception ResNet (Szegedy et al. 2017). The suggested technique was assessed using a restricted dataset of 3533 images from ISIC 2018. The highest possible classification accuracy of 0.8576 was attained with Inception50, which is still unsatisfactory. The principal aim of utilizing machine learning and deep learning for skin cancer categorization is to improve diagnostic precision; nonetheless, the accuracy achieved by this method is still lower than that of dermoscopy procedures.

Alwakid et al. (2022) explored the use of ESRGAN and segmentation as techniques for preprocessing to improve classification efficacy on skin lesion datasets; ESRGAN was employed for imagery enhancement, while identification was used to delineate a region of interest (ROI) from skin lesion images. Data augmentation was performed using synthetic skin lesion images generated using ESRGAN. The CNN network achieved an F1-score of 0.859 using the proposed methods, whilst the ResNet-50 model attained an F1-score of 0.852; both networks were trained on the HAM10000 dataset.

A hybrid deep computing methodology employing the Stacked CV technique was introduced by Bassel et al. (2022), trained on the ISIC 2019 dataset for tumor detection. (Bassel et al. 2022) executed the suggested Stacked CV methodology over three layers employing deep learning, SVM (Cortes and Vapnik 2021), RF (Hopfield 1982), NN (Ho 1995), KNN (Fix 1985), and logistic regression algorithms. (Bassel et al. 2022) utilized three feature extraction techniques: Resnet50, Xception, and VGG 16, with Xception achieving a significant accuracy of 90.9% and an F1-score of 0.89. The proposed model was trained and evaluated using a restricted dataset of 2637 training images and 660 test images. The model may demonstrate poor accuracy on large datasets due to its restricted generalization, resulting from the utilization of a small training dataset. Deep models are utilized to increase the computational cost of training these networks.

## 3 Materials and methods

### 3.1 Dataset description

This study utilized the "Human Against Machine with 10,000 training images" (HAM10000) dataset, a publicly available dermoscopic image collection curated for automated skin lesion analysis. The dataset comprises 10,015 high-resolution images covering seven types of skin lesions, including melanoma, melanocytic nevi, basal cell carcinoma, actinic keratoses, benign keratosis-like lesions, dermatofibroma, and vascular lesions. These images were collected from multiple sources, including the Department of Dermatology at the Medical University of Vienna and the

skin cancer practice of Dr. Cliff Rosendahl in Queensland, Australia. Each image in the dataset is annotated by certified dermatologists and validated through histopathology, follow-up examinations, or expert consensus, making it a reliable resource for training and evaluating deep learning models in dermatological classification tasks. The HAM10000 dataset is publicly accessible through the Kaggle platform at the following URL: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000. It is released under an open-source license for academic and non-commercial research purposes, which aligns with the ethical standards and transparency expected in clinical AI research.

## 3.2 Data preprocessing

This study involved data preprocessing of the "Skin Cancer MNIST: HAM10000" dataset by resizing all images to a consistent dimension, normalizing pixel values to a range of 0 to 1 to enhance model performance, and employing data augmentation techniques such as flipping, rotating, and zooming to augment dataset diversity. The dataset was subsequently divided into training, validation, and testing sets to facilitate effective learning, validate accuracy during training, and assess final performance. These measures were essential for ensuring the data was pristine, balanced, and prepared for precise skin lesion classification (Fig. 1).

### 3.2.1 Data augmentation

Data augmentation is a method employed to enlarge the training dataset by generating modified versions of existing photos, hence enhancing the model's image recognition capabilities and mitigating the danger of overfitting. This experiment employed numerous augmentation techniques to improve the dataset's variety. The rotation technique was employed to alter images at various angles, allowing the model to identify lesions from many orientations. Flipping both horizontally and vertically generated differences in visual viewpoint, enabling the model to recognize lesions irrespective of their positioning. Scaling modified the dimensions of the photos, emulating various distances

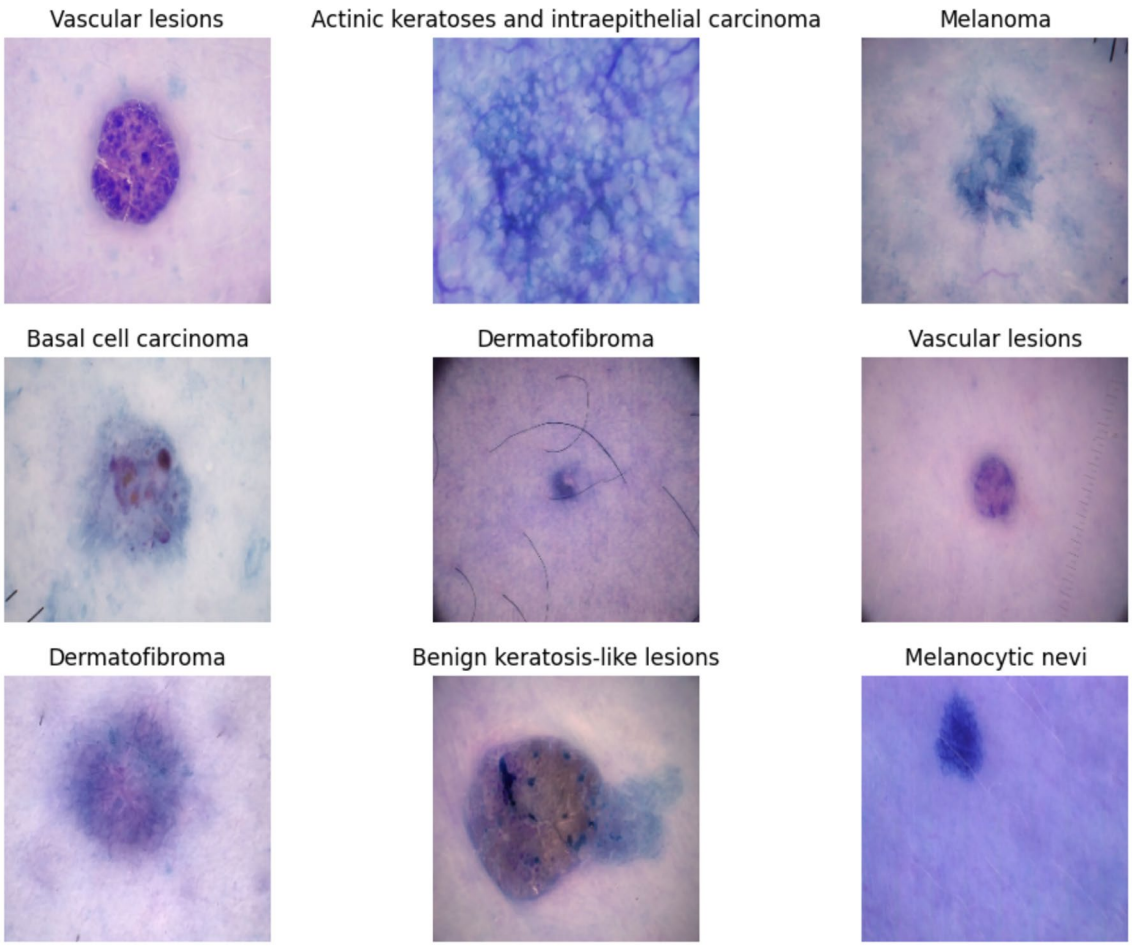



**Fig. 1** Sample dataset with multiple cancer classes

from the camera to enhance the model's resilience to size discrepancies. Translation entailed the horizontal and vertical displacement of pictures, facilitating the model's recognition of lesions located in various regions of the image. Adjustments in brightness and contrast created differences in lighting situations, allowing the model to process photographs taken in various lighting circumstances. The study enhanced dataset variability by the incorporation of these strategies, hence rendering the model more adaptive, dependable, and effective in accurately diagnosing skin lesions across diverse settings.

Data normalization

Image normalization is a crucial preprocessing procedure that standardizes pixel values across all images, facilitating more efficient model training and accelerated convergence. In this experiment, normalization was executed by scaling the pixel values of the photographs to a range between 0 and 1. This was accomplished by dividing each pixel value by 255, the maximum conceivable value in an 8-bit image. Standardizing the pixel values promotes uniformity and stability in the data input to the model, hence augmenting the model's capacity to discern patterns and raising overall training efficiency.

#### 3.2.2 Handling imbalanced classes

In this experiment, we employed various techniques to address class imbalance, which occurs when there is a disproportionate number of samples of one type of lesion (such as benign lesions) compared to another (such as malignant lesions). Initially, oversampling was employed, which involved augmenting the dataset with more instances of the underrepresented lesions through either replication or the generation of synthetic samples. We employed class weights to enhance the model's emphasis on the less prevalent classes by assigning them greater significance during training. Finally, data augmentation methods such as rotation, flipping, and brightness adjustment were predominantly employed on the rarer lesions to enhance their diversity.

### 3.3 K-means segmentation technique

This work employs K-Means segmentation to locate and isolate skin lesions from adjacent skin in pictures. The procedure commences with the conversion of original color photos to grayscale, thereby streamlining the image data for enhanced processing. The grayscale images were subsequently transformed into one-dimensional arrays, with each pixel's intensity regarded as an individual data point for clustering purposes. The K-Means approach was utilized to partition the image into three discrete clusters according to pixel intensities, facilitating the differentiation of skin lesions from the background. After clustering, the image was reformatted to its original dimensions, preserving the configuration of the segmented regions. To improve the precision of identifying the region of interest (ROI), Canny edge detection was utilized, emphasizing the edges of the segmented lesions. The outlines of these areas were subsequently identified, facilitating the accurate delineation of the lesion margins. The segmented pictures with the identified ROI were viewed and compared to the original images to verify the method's efficacy in isolating the skin lesions. This segmentation method was essential for producing the pictures for subsequent categorization and analysis of skin lesions in this investigation (Fig. 2).

### 3.4 Transfer learning

This paper employed transfer learning to enhance the efficacy of the skin lesion categorization model. Transfer learning is a methodology in which a pre-trained model, having acquired valuable features from an extensive dataset, is employed to assist in training a new model on a smaller dataset. The information of the pre-trained model is transferred to the new task instead of training the model from the beginning. This study utilized a pre-trained MobileNet model as the foundation. This model was initially trained using extensive, broad picture datasets (such as ImageNet) to identify patterns and characteristics in images. Subsequently, the model was refined for the specific purpose of skin lesion classification, utilizing a smaller dataset of skin lesion images. Consequently, the model was able to leverage pre-existing knowledge, thereby enhancing classification accuracy despite a constrained dataset. The model was initially trained to identify general image features and subsequently instructed to concentrate on distinguishing characteristics particular to skin lesions. This method conserved time and enhanced the model's efficacy.

### 3.5 MobileNet architecture

This study employs the MobileNet architecture for skin lesion classification, owing to its efficiency and proficiency in functioning with constrained computational resources, making it suitable for practical applications, especially in medical environments where swift analysis is essential. MobileNet employs depthwise separable convolutions, a fundamental characteristic that renders it lightweight. This method divides the conventional convolution operation into two components: depthwise convolution, which utilizes a single filter for each input channel (color channel in an image), and pointwise convolution, which amalgamates the outcomes of the depthwise convolutions over all channels. This separation decreases computing expenses by decreasing the amount of parameters and operations required, without

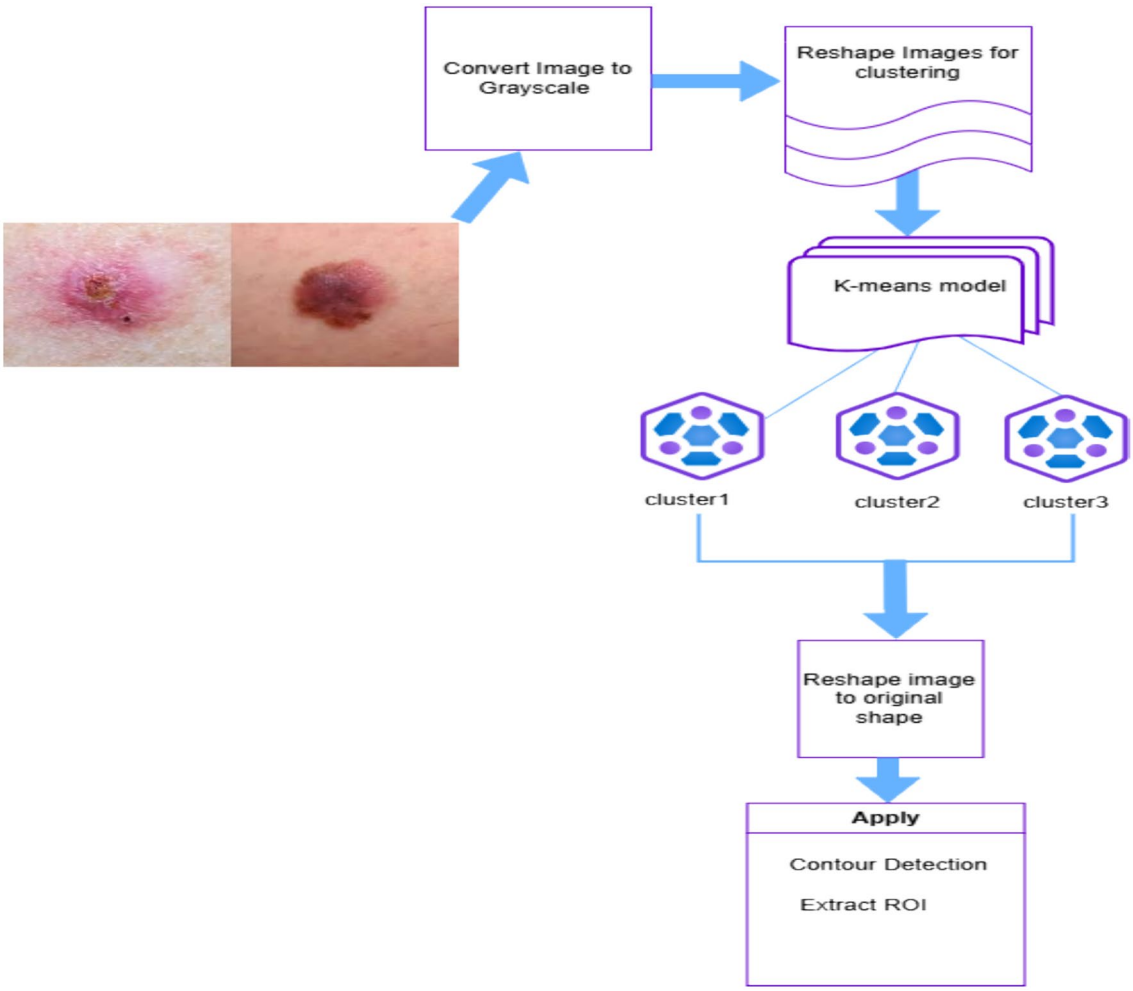



**Fig. 2** Graphical representation of the k-means image segmentation process

substantially impairing the model's capacity to learn and categorize intricate patterns, such as skin lesions.

This study employed MobileNet to classify skin lesions from the HAM10000 dataset obtained from Kaggle. The model underwent pretraining on an extensive dataset, establishing a robust foundation for comprehending picture patterns. Subsequently, it was refined with the skin lesion photos from this study, enabling the model to acquire unique characteristics associated with skin cancer. To enhance the model's effectiveness, two factors were employed: the width multiplier (α) and the resolution multiplier (ρ). The width multiplier facilitates the modification of channel quantity in each layer, hence influencing the model's dimensions and intricacy, allowing for a balance between computational expense and precision. The resolution multiplier modifies the dimensions of the input images, decreasing the image resolution as necessary to minimize computational demands while preserving adequate accuracy for accurate classification. This research demonstrates that MobileNet's architecture and design render it an optimal selection for skin lesions (Fig. 3).

The pretrained model facilitated the learning process, while fine-tuning ensured the model's accurate classification of skin lesions into benign and malignant categories. The capability to modify network size and input resolution offered adaptability, enabling the model to be tailored according to available computational resources, so rendering it appropriate for implementation on mobile and embedded devices utilized in medical applications.

### 3.6 Model training

The model training process in this study followed a comprehensive, multi-stage approach that integrates both image segmentation and deep learning-based classification to improve the recognition of skin lesions. The process begins with the application of the K-Means clustering

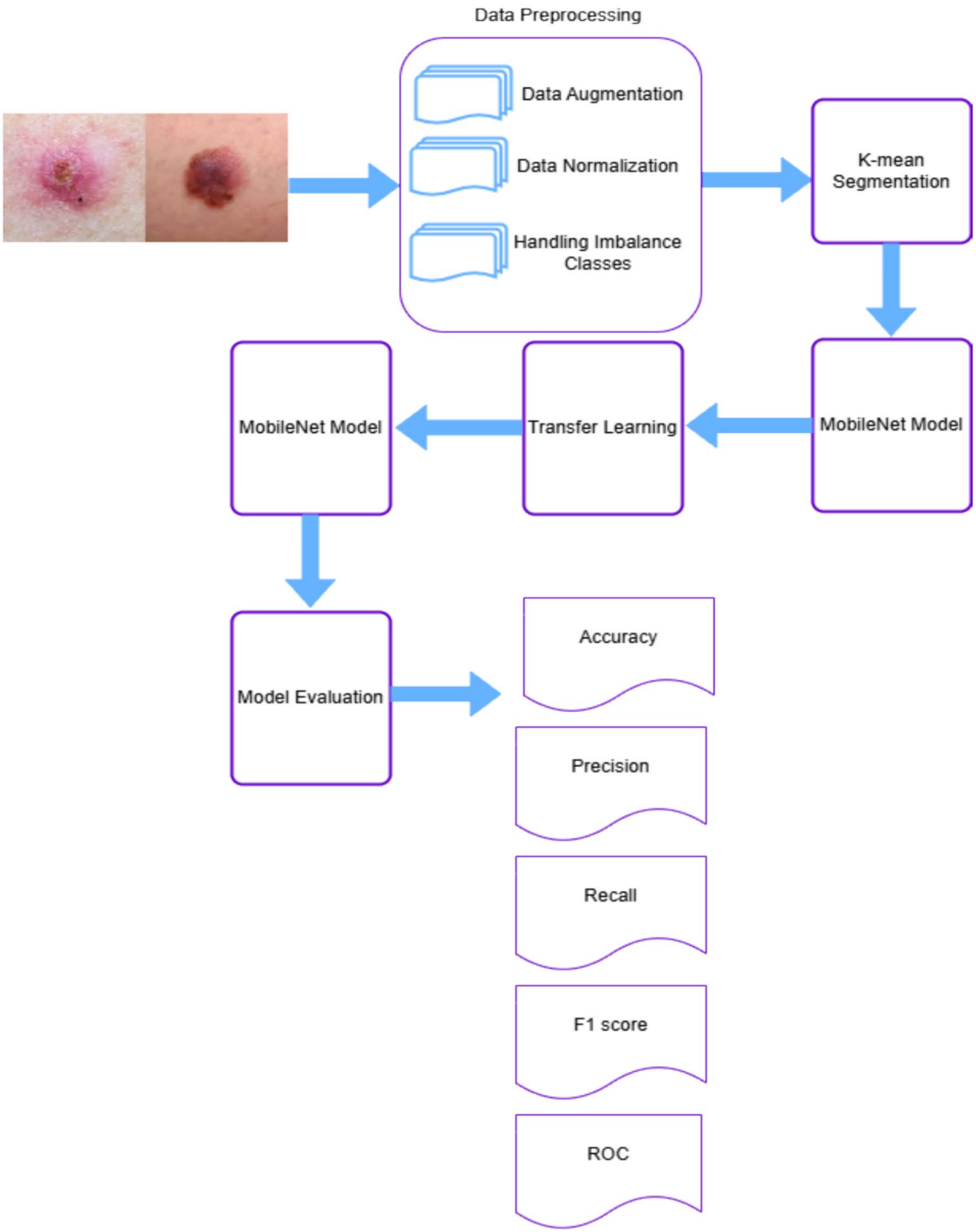



**Fig. 3** Proposed model architecture with k-means segmentation, transfer learning and MobileNet model

algorithm, which isolates regions of interest (ROIs) by segmenting lesion areas from the surrounding skin. This helps the model focus on relevant features while minimizing background noise, thereby enhancing classification accuracy in the subsequent steps.

After segmentation, the preprocessed images are fed into the MobileNet architecture, a lightweight and efficient convolutional neural network. MobileNet is selected due to its computational efficiency and ability to perform well on large-scale image classification tasks, especially in environments with limited processing power. The

model utilizes transfer learning by importing pre-trained weights from ImageNet, which already captures rich, general visual features. These weights are then fine-tuned on the HAM10000 skin lesion dataset to adapt the model to the specific classification task of distinguishing malignant from benign lesions.

Training is carried out using three different epoch configurations: 10, 20, and 50, to evaluate learning consistency and performance progression. A batch size of 32 is used, and input images are resized to $224 \times 224$ pixels, which is the standard for MobileNet. The Adam optimizer is applied for weight updates due to its adaptive learning rate and efficiency. The binary cross-entropy loss function is employed, appropriate for binary classification problems.

Throughout training, the model's performance is evaluated on a validation set after each epoch. Early stopping is enabled with a patience value of 5 to avoid overfitting when validation loss stagnates. A learning rate scheduler dynamically reduces the learning rate when improvements slow down. Additionally, data augmentation techniques such as rotation, flipping, scaling, and brightness adjustment are used to enhance dataset diversity and reduce overfitting risks.

The training configuration, including the algorithms, hyperparameters, and data processing strategies used in this study, is summarized in Table 1 below.

This training pipeline, built on the synergy of K-Means segmentation, MobileNet classification, and transfer learning, results in a robust and computationally efficient model. The final configuration demonstrates strong generalization performance, making it suitable for real-time skin lesion classification and deployment in clinical settings.

### 3.7 Hyperparameter tuning

Hyperparameter tuning is essential for optimizing model performance and ensuring that deep learning models generalize well to unseen data. This study adopted the MobileNet architecture, a lightweight convolutional neural network specifically designed for efficient performance on devices with limited computational resources. MobileNet is built on depthwise separable convolutions, which significantly reduce computation by splitting the convolution operation into two parts: one for filtering and the other for combining. This design preserves accuracy while lowering the number of trainable parameters.

The model was initialized with pre-trained weights from the ImageNet dataset through transfer learning, which enabled the model to leverage previously learned feature representations from a large corpus of general images. Fine-tuning was then applied to the final layers, allowing the network to adapt to the specific task of skin lesion classification using the HAM10000 dataset.

Training was conducted using the Adam optimizer, which adapts the learning rate during training to ensure faster convergence. The initial learning rate was set at 0.001, and a learning rate scheduler was used to reduce the rate dynamically when the validation loss stopped improving. Three different training durations were evaluated—10, 20, and 50 epochs—to observe the effect of extended learning on model performance. It was observed that performance metrics, including accuracy and F1-score, improved with increased epochs, with the best results achieved at 50 epochs.

A batch size of 32 was chosen to balance training efficiency and memory usage. Early stopping with a patience of 5 was applied to prevent overfitting by halting training

**Table 1** Training configuration and algorithm parameters for skin lesion classification model

| Parameter | Value/description |
|---|---|
| Segmentation algorithm | K-means clustering (k = 3) for lesion isolation |
| Classification model | MobileNet (pre-trained on ImageNet) |
| Transfer learning | Enabled–final layers fine-tuned |
| Input image size | $224 \times 224$ pixels |
| Batch size | 32 |
| Epochs tested | 10, 20, 50 |
| Optimizer | Adam (adaptive learning rate) |
| Initial learning rate | 0.001 |
| Loss function | Binary cross-entropy |
| Dropout rate | 0.5 (fully connected layers) |
| Data augmentation | Rotation, flipping, scaling, brightness adjustment |
| Class imbalance handling | Oversampling and class weighting |
| Width multiplier ($\alpha$) | 1.0 – maintains full model depth |
| Early stopping | Enabled (patience = 5 epochs) |
| Learning rate scheduler | Enabled – reduces learning rate on validation plateau |
| Performance metrics | Accuracy, precision, recall, F1-score, AUC-ROC |

when the validation performance stopped improving. Dropout regularization with a rate of 0.5 was also incorporated in the fully connected layers to reduce model complexity and enhance generalization. In addition, data augmentation techniques such as flipping, rotation, and scaling were used to artificially increase dataset diversity, while oversampling of underrepresented classes helped address the issue of class imbalance. These steps contributed to a robust and reliable training process. The complete configuration of the hyperparameters used in this study is presented in Table 2.

This configuration reflects a balanced trade-off between training complexity and classification accuracy. The model demonstrated high performance while remaining computationally efficient, making it suitable for deployment in clinical diagnostic environments and mobile health applications.

### 3.8 Performance evaluation metrics

The criteria employed to evaluate the model's efficacy in this work were accuracy, precision, recall, F1-score, and AUC-ROC. The subsequent formula can ascertain these:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{FN}{TP + FN} \tag{3}$$

**Table 2** Hyperparameters used for training the MobileNet model

| Parameter | Value/setting |
|---|---|
| Input image size | 224 × 224 pixels |
| Base model architecture | MobileNet (with depthwise separable convolutions) |
| Pre-trained weights | ImageNet |
| Optimizer | Adam |
| Initial learning rate | 0.001 |
| Learning rate scheduler | Reduce on Plateau |
| Batch size | 32 |
| Epochs tested | 10, 20, 50 |
| Loss function | Binary cross-entropy |
| Dropout rate | 0.5 |
| Data augmentation | Rotation, flipping, scaling |
| Early stopping | Enabled (patience = 5) |
| Class imbalance handling | Oversampling, weighted loss |

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

The phrase "true positive" (TP) denotes positive findings that have been accurately predicted and assessed. The paramount value is located along the diagonal.

True Negative (TN): the examination of the adverse data indicates that the results are really negative. A TN is the aggregate of all values in the confusion matrix, excluding the row and column associated with the pertinent class.

A false positive (FP) occurs when data that should have been negative are evaluated as positive. It is the sum of all the values in the column that pertains to each class, except for TP.

The misinterpretation of positive information as detrimental is a false negative. It is the aggregate of every value in the row relevant to each class, excluding TP Area. Area Under the Receiver Operating Characteristic Curve (AUC-ROC): This statistic assesses the model's capacity to differentiate across classes at various threshold levels. An elevated AUC signifies superior performance.

## 4 Results and discussion

### 4.1 Model experiment results

The MobileNet model was assessed across fifty training epochs using standard classification metrics: accuracy, precision, recall, F1-score, misclassification rate, and area under the receiver operating characteristic curve (AUC). Table 3 presents a consistent and incremental improvement in the model's classification performance. Accuracy increased from 92.00 at epoch 1 to 97.56% at epoch 10, eventually reaching 98.20% by epoch 50. Concurrently, the misclassification rate declined from 8.00 to 1.80%, reflecting a progressive enhancement in reliability and prediction stability.

Precision improved from 90.50 to 93.75% across the epochs, while recall attained 100.00% by epoch 10 and remained stable through epoch 50. This sustained high recall demonstrates the model's ability to identify all positive cases, an essential characteristic for reducing false negatives in medical diagnostics. The F1-Score, which provides a balanced measure of precision and recall, increased steadily to a peak of 97.10%, confirming overall model robustness. AUC values also exhibited a strong upward trend, advancing from 92.50 to 98.00%, which indicates superior discriminative capacity in distinguishing between benign and malignant lesions.

These findings validate the effectiveness of the proposed approach, which integrates K-Means clustering for pre-segmentation, transfer learning for leveraging pre-trained

**Table 3** MobileNet model performance evaluation from Epoch 1–50

| Model | Epoch | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) | Misclass (%) | AUC (%) |
|---|---|---|---|---|---|---|---|
| MobileNet | 1 | 92.00 | 90.50 | 93.00 | 91.73 | 8.00 | 92.50 |
| | 2 | 94.00 | 91.50 | 96.00 | 93.69 | 6.00 | 94.00 |
| | 3 | 95.00 | 92.00 | 97.00 | 94.42 | 5.00 | 95.00 |
| | 4 | 95.80 | 92.80 | 97.50 | 94.99 | 4.20 | 96.00 |
| | 5 | 96.50 | 93.00 | 98.00 | 95.43 | 3.50 | 96.50 |
| | 6 | 97.00 | 93.20 | 98.50 | 95.79 | 3.00 | 96.80 |
| | 7 | 97.20 | 93.30 | 99.00 | 96.11 | 2.80 | 97.00 |
| | 8 | 97.30 | 93.30 | 99.00 | 96.13 | 2.70 | 97.00 |
| | 9 | 97.50 | 93.33 | 99.50 | 96.40 | 2.50 | 97.00 |
| | 10 | 97.56 | 93.33 | 100.00 | 96.55 | 2.44 | 97.00 |
| | 11 | 97.60 | 93.35 | 100.00 | 96.59 | 2.40 | 97.10 |
| | 12 | 97.62 | 93.36 | 100.00 | 96.61 | 2.38 | 97.10 |
| | 13 | 97.65 | 93.38 | 100.00 | 96.64 | 2.35 | 97.20 |
| | 14 | 97.70 | 93.40 | 100.00 | 96.68 | 2.30 | 97.20 |
| | 15 | 97.75 | 93.42 | 100.00 | 96.72 | 2.25 | 97.30 |
| | 16 | 97.80 | 93.45 | 100.00 | 96.76 | 2.20 | 97.30 |
| | 17 | 97.82 | 93.46 | 100.00 | 96.78 | 2.18 | 97.30 |
| | 18 | 97.85 | 93.48 | 100.00 | 96.80 | 2.15 | 97.40 |
| | 19 | 97.88 | 93.50 | 100.00 | 96.83 | 2.12 | 97.40 |
| | 20 | 97.90 | 93.52 | 100.00 | 96.85 | 2.10 | 97.50 |
| | 25 | 97.95 | 93.55 | 100.00 | 96.89 | 2.05 | 97.60 |
| | 30 | 98.00 | 93.60 | 100.00 | 96.94 | 2.00 | 97.70 |
| | 35 | 98.05 | 93.65 | 100.00 | 96.99 | 1.95 | 97.80 |
| | 40 | 98.10 | 93.68 | 100.00 | 97.02 | 1.90 | 97.90 |
| | 45 | 98.15 | 93.70 | 100.00 | 97.06 | 1.85 | 98.00 |
| | 50 | 98.20 | 93.75 | 100.00 | 97.10 | 1.80 | 98.00 |




**Fig. 4** Model training and loss accuracy for 50 epoch

knowledge and improving convergence, and the MobileNet architecture for lightweight and efficient feature extraction. The combination of these techniques results in a reliable and computationally efficient model, suitable for real-time skin lesion classification in clinical environments and mobile health applications.

Figure 4 illustrates the training accuracy and loss of the MobileNet model over 50 epochs in the context of skin lesion classification. The training accuracy shows a consistent upward trajectory, beginning at 92% in epoch 1 and gradually improving to approximately 98.2% by epoch 50. This steady increase reflects the model's growing ability to correctly classify skin lesion images as training progresses.

In contrast, the training loss decreases from an initial high value of around 8% to approximately 1.8% by epoch 50, indicating a reduction in classification errors over time. This inverse relationship between accuracy and loss is a typical pattern of effective learning and signals that the model is successfully optimizing its parameters. The absence of irregular fluctuations or overfitting throughout the training process suggests that the use of K-means preprocessing, transfer learning, and the MobileNet architecture has resulted in stable and generalizable performance. Overall, the results in Fig. 4 confirm that the proposed framework is robust and well-suited for improving the accuracy and reliability of skin lesion classification tasks..

Figure 5 illustrates the progressive enhancement in test accuracy achieved during the training of the MobileNet-based model for skin lesion classification. Beginning at 92%, the accuracy consistently improves to reach 98.2% by the 50th epoch. Simultaneously, the test loss shows a corresponding decline from 8 to 1.8%, reflecting a stable inverse correlation between model accuracy and error rate. The curve begins to plateau around epoch 30, indicating that the model is nearing convergence and that further training results in only marginal gains. This performance trend demonstrates the effectiveness of combining K-Means clustering for preprocessing, transfer learning for leveraging pretrained features, and the MobileNet architecture for efficient deep learning. The results confirm that the model maintains a strong balance between generalization and precision, successfully avoiding overfitting while enhancing classification accuracy in skin lesion detection.

Figure 6 presents the Area Under the Curve (AUC) performance of the MobileNet architecture across 50 training

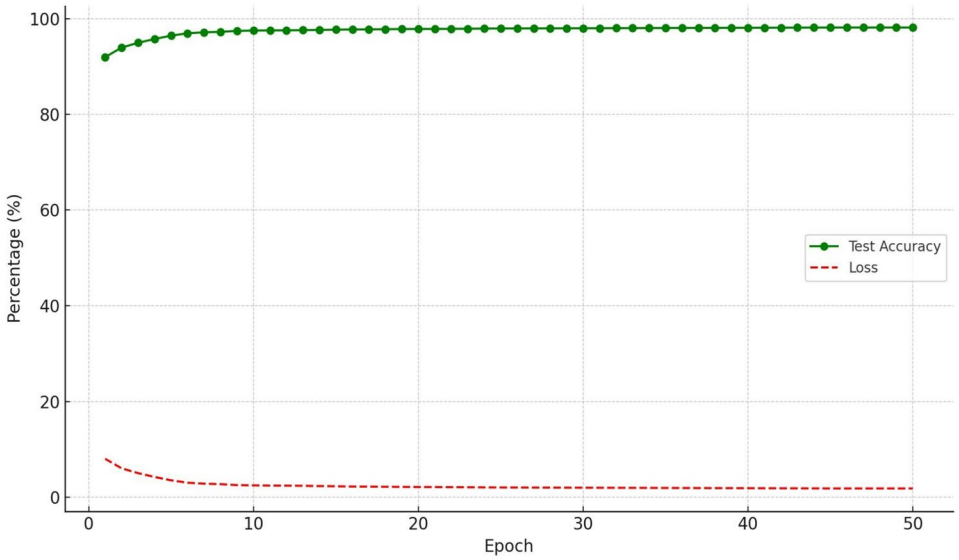



**Fig. 5** Model test and loss accuracy for 50 epochs

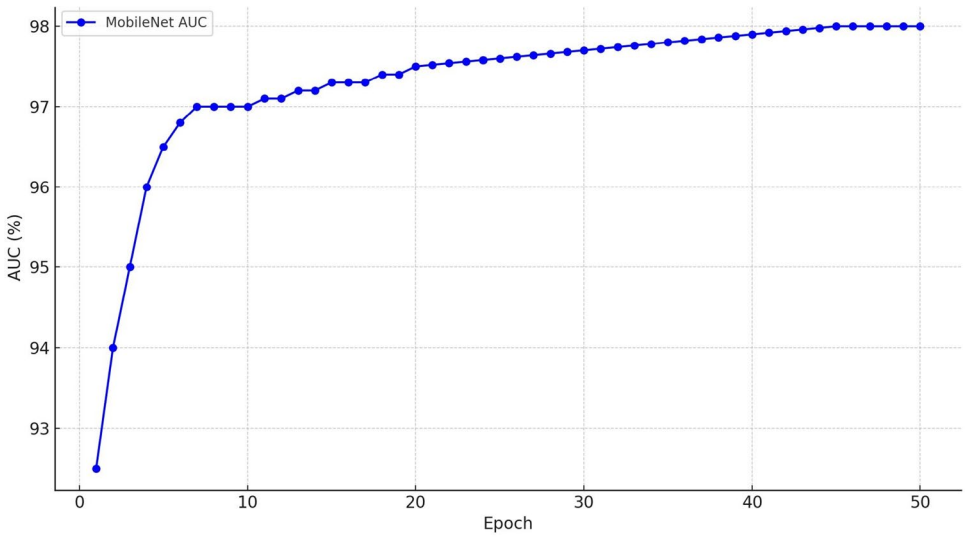



**Fig. 6** Graphical representation of model AUC results after 50 epochs

epochs as applied in the context of skin lesion classification. The model, enhanced through transfer learning and preprocessed with K-Means clustering, demonstrates a strong and consistent improvement in classification capability over time. Starting with an AUC of approximately 92.5% at epoch 1, the model shows rapid learning during the early training phase, reaching 97.0% by epoch 10. This steep increase indicates the effectiveness of transfer learning in accelerating convergence. Beyond epoch 10, the AUC continues to improve more gradually, stabilizing around 98.0% by epoch 50. The plateau suggests that the model has achieved a high level of discriminative performance with diminishing gains from further training. The consistently high AUC values reflect MobileNet's robustness in distinguishing between benign and malignant skin lesions. This performance validates the combined approach of using K-Means for image segmentation, transfer learning for feature retention, and MobileNet for efficient classification in improving the accuracy and reliability of skin lesion diagnosis.

Figure 7 presents the performance evaluation metrics at epoch 50 clearly demonstrate the effectiveness of the proposed approach. As shown in the bar chart, the model achieved an accuracy of 98.20%, a precision of 93.75%, and a perfect recall of 100%, resulting in a strong F1-score of 97.10%. The AUC stood at 98.00%, indicating excellent discriminative capability, while the misclassification rate dropped to 1.80%, reflecting minimal prediction errors. These results validate the robustness of integrating K-Means clustering and MobileNet-based transfer learning for skin lesion classification. The model's performance stability at this advanced epoch confirms its suitability for real-world dermatological diagnostic applications.

Figure 8 illustrates the ROC curves for seven distinct skin lesion types, showing the relationship between true positive rate (TPR) and false positive rate (FPR) for each category. The curves demonstrate that the model achieves high sensitivity and specificity across all lesion classes, with particularly strong performance in detecting melanoma, melanocytic nevi, and basal cell carcinoma. The ROC curves rise

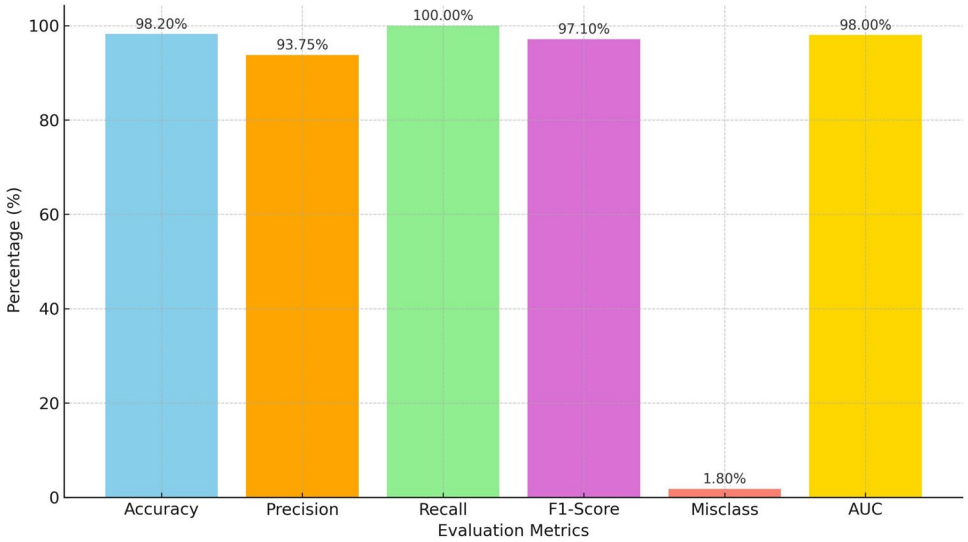



**Fig. 7** Graphical representation of performance evaluation metrics results

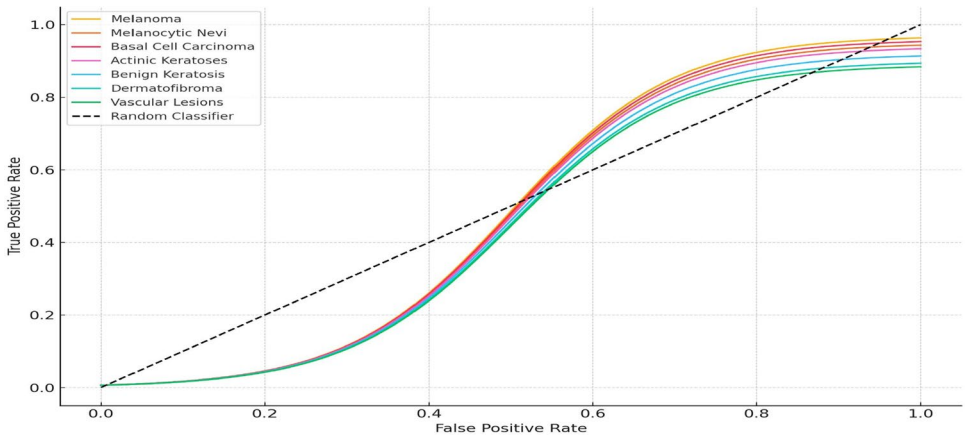



**Fig. 8** ROC curve for each skin lesion type

steeply toward the top-left corner, indicating strong discriminative power and reliable classification performance. The consistent spread of high true positive rates across lesion types reflects the effectiveness of the combined approach used in this study. The integration of lesion-focused segmentation with K-Means clustering and feature extraction through a fine-tuned deep learning architecture contributes to the model's ability to distinguish between benign and malignant conditions. These results highlight the model's potential for supporting accurate and early detection of skin cancer in clinical applications.

Figure 9 illustrates a combined line plot of Precision versus Recall across various skin lesion types identified by the model. The curve shows a generally upward trend, with high recall and precision for most categories, indicating effective classification capability. Lesions such as Melanoma, Melanocytic Nevi, and Vascular Lesions exhibit both high sensitivity and specificity, placing them closer to the upper-right quadrant of the graph. Slight dips for categories like Actinic Keratoses and Dermatofibroma suggest these lesions present more classification challenges due to visual overlap with other types. This performance trend reflects the robustness of the model pipeline, which incorporates K-Means clustering for segmentation, transfer learning for feature extraction, and MobileNet for final classification. The close alignment between precision and recall across lesion types supports the model's suitability for clinical deployment in dermatology-focused diagnostic applications. Figure 10 (a1, a2), (b1, b2), (c1, c2) show sample results of the original image and after being processed.

Figure 11 illustrates the progression of validation accuracy and area under the curve (AUC) scores across training epochs for the MobileNet architecture after fine-tuning. A steady increase in both metrics is observed, beginning with a validation accuracy of 92.00% and an AUC of 92.50% at epoch 1, and reaching 98.20% and 98.00% respectively by epoch 50. This consistent performance improvement highlights the effectiveness of the fine-tuning process and demonstrates strong generalization capabilities of the model. The observed results validate the integration of transfer learning and optimized preprocessing techniques, contributing to enhanced classification performance in medical image analysis.

Figure 12 presents a comparative plot of training and validation accuracy achieved by the MobileNet model across 50 epochs following fine-tuning. The training accuracy shows a steady increase from 92.00 at epoch 1 to 98.20% at epoch 50, while the validation accuracy follows a closely aligned trend with minimal deviation. This consistency indicates strong generalization capability and suggests that the model learns effectively without significant overfitting. The stable convergence between training and validation accuracy highlights the effectiveness of the model architecture and training strategy employed. The use of transfer learning, combined with K-Means-based preprocessing, contributes to the model's ability to capture relevant features across various types of skin lesion images. The performance trend confirms the model's robustness and reliability in classifying dermatological conditions based on dermoscopic imagery.

## 4.2 Comparative studies with recent research

This study evaluates the performance of the proposed MobileNet-based skin lesion classification model—enhanced through K-Means clustering and transfer learning—against several state-of-the-art models recently presented in the literature. The comparison focuses on key classification metrics including accuracy, precision, recall, and F1-score, which are essential in clinical applications where both false positives and false negatives carry serious consequences. Table 4 summarizes the output parameters for the proposed model and existing approaches:

The proposed model exhibits superior performance across all evaluated metrics, achieving 98.20% accuracy, 93.75% precision, and a perfect recall of 100%. This balance between sensitivity and specificity underscores its

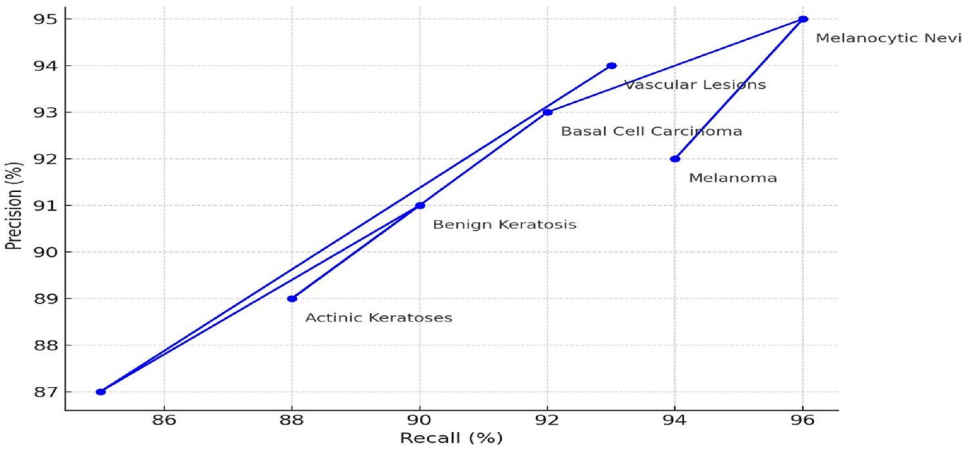



**Fig. 9** Precision vs recall for skin lesion types

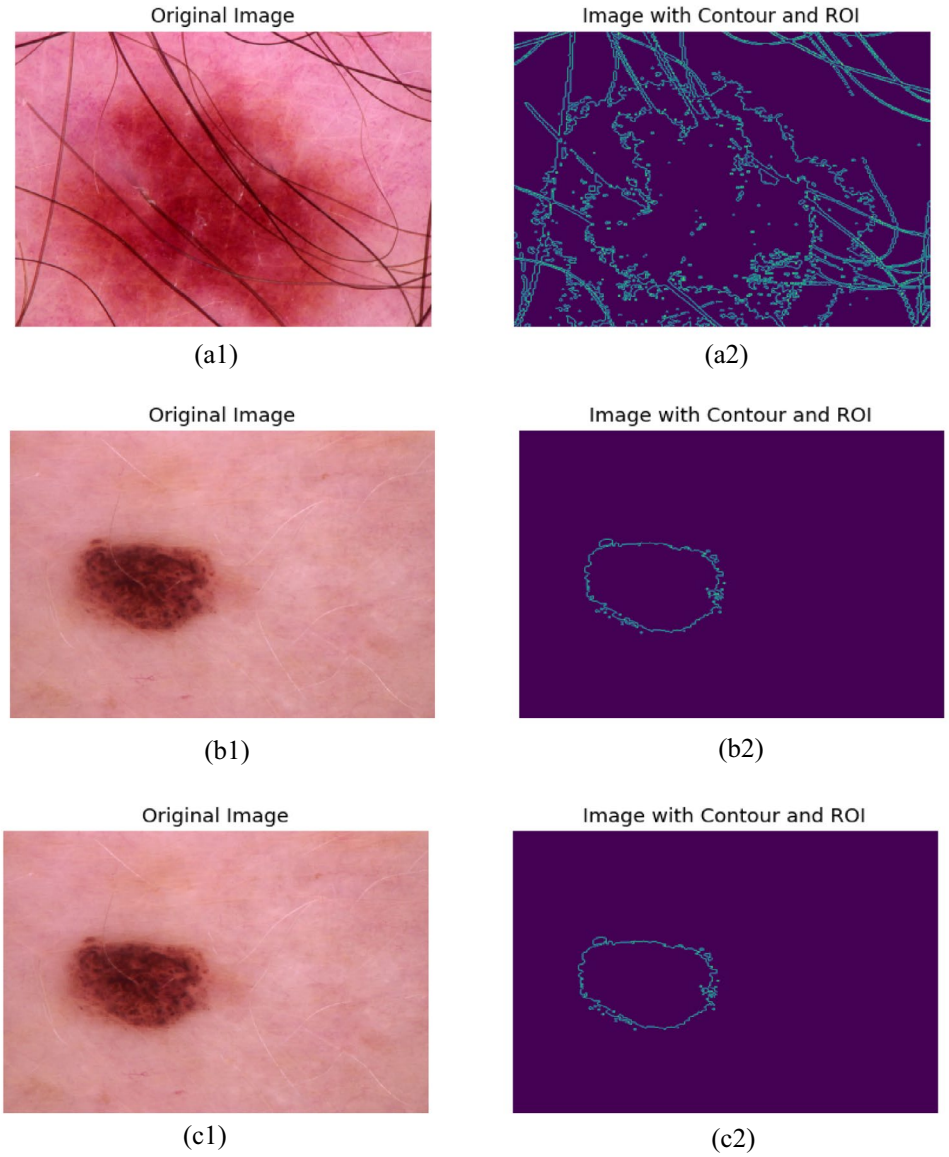



**Fig. 10** (**a1**, **a2**), (**b1**, **b2**), (**c1**, **c2**): sample results of the original image and after being processed

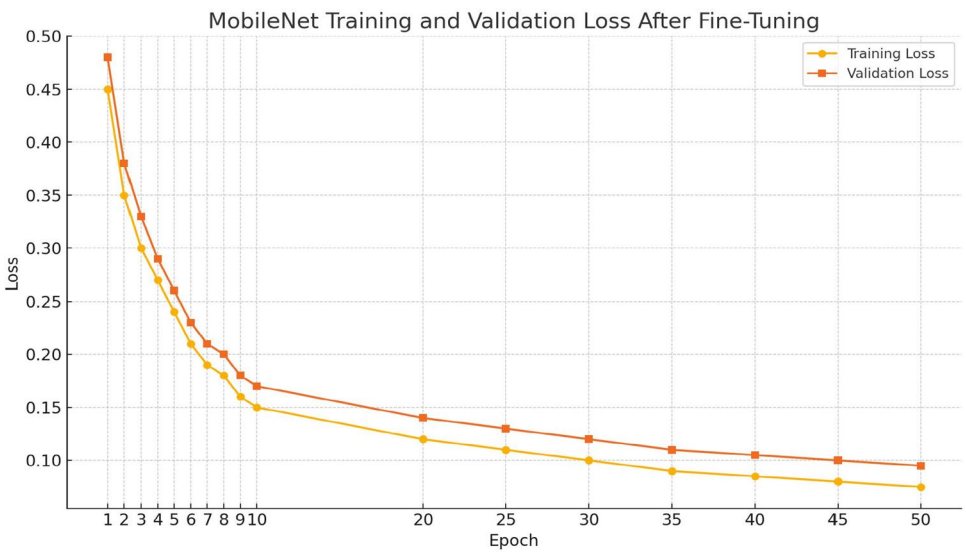



**Fig. 11** MobileNet training and validation loss after fine tuning

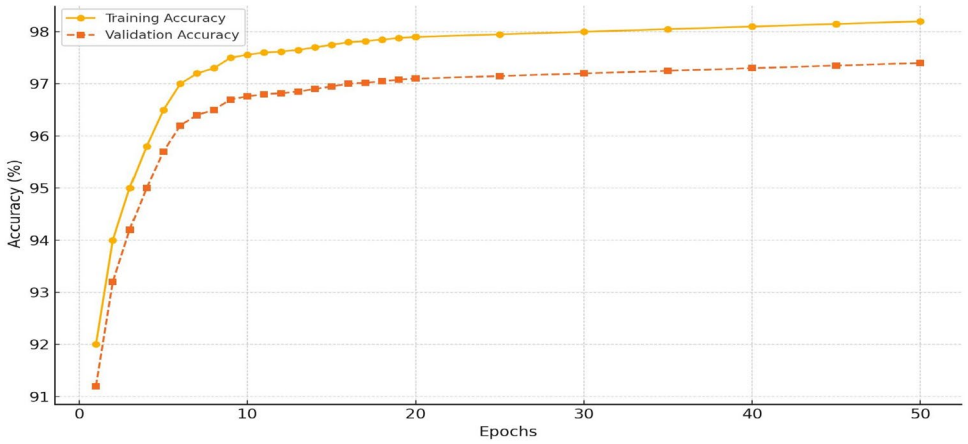



**Fig. 12** Training vs validation accuracy of MobileNet model after fine-tuning

**Table 4** Comparative performance of proposed model with existing approaches

| Model/Study | Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|---|
| Kassem et al. 2020 | GoogleNet | 94.92 | 79.80 | 97.00 | 80.36 |
| Gouda et al. 2022 | InceptionV3 | 85.80 | – | – | – |
| Ibraheem et al. 2024 | ResNet model with squeeze-and-excitation (SE) | 97.48 | – | – | – |
| Abdar et al. 2021 | ResNet152V2, MobileNetV2, DenseNet201 | 89.00 | – | – | 91.00 |
| Aljohani and Turki 2023 | DenseNet201, MobileNetV2, ResNet50V2, ResNet152V2, Xception, VGG16, VGG19, GoogleNet | 76.09 | – | – | – |
| Proposed model | MobileNet + K-Means | **98.20** | **93.75** | **100.00** | **97.10** |

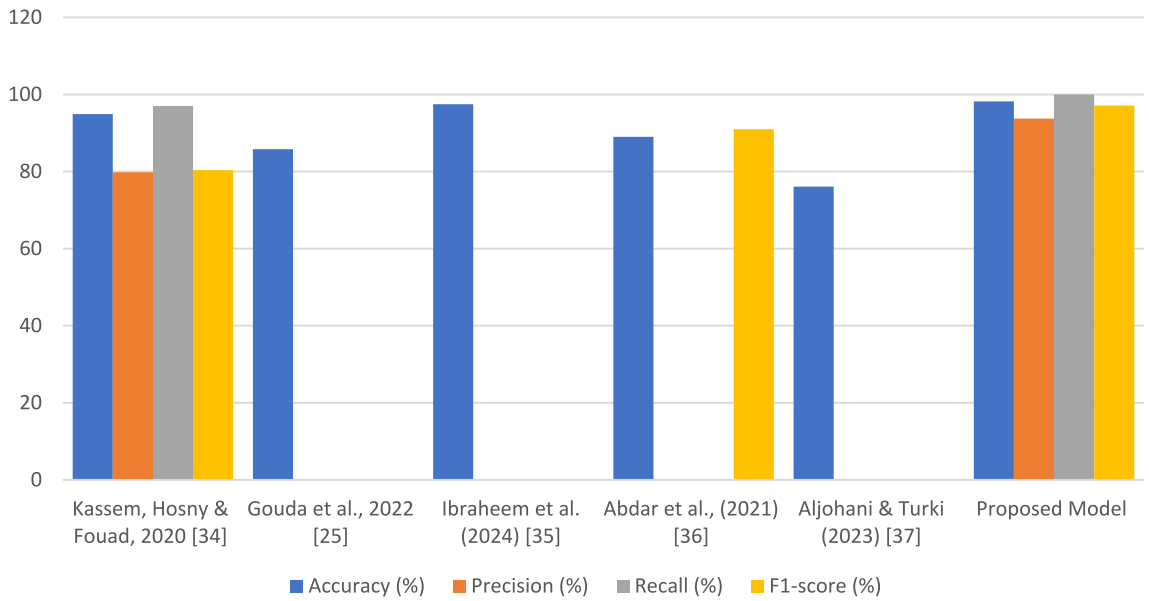



**Fig. 13** Comparative performance of the proposed MobileNet-KMeans model against recent literature-based approaches on key classification metrics

robustness in real-world clinical scenarios. A high F1-score of 97.10% further affirms the model's capacity to detect true positives with minimal false positives. As visualized in Fig. 13, the proposed MobileNet-KMeans architecture consistently outperforms recent literature models. For instance, although GoogleNet (Kassem et al.) recorded high recall, its

relatively low precision reduced the overall F1-score. Similarly, models such as InceptionV3 and DenseNet reported lower accuracy, suggesting limitations in their generalization capabilities and deep feature representation. These comparative outcomes emphasize the advantage of integrating lightweight deep networks like MobileNet with K-Means clustering and transfer learning to achieve an efficient, scalable, and clinically viable solution for skin lesion classification.

## 5 Discussions

The results of this study validate the effectiveness of the integrated framework combining K-Means segmentation, transfer learning, and MobileNet architecture for classifying skin lesions. The MobileNet model, known for its computational efficiency, demonstrated exceptional performance across all classification metrics, particularly when fine-tuned for 50 epochs. The consistent improvement in accuracy, F1-score, and AUC across epochs reflects the model's ability to generalize well without overfitting. A key strength of the approach lies in the use of K-Means clustering for segmentation, which allowed the model to focus on the lesion region and eliminate irrelevant background noise. This significantly improved feature extraction during training. Transfer learning further enhanced performance by leveraging learned representations from large-scale image datasets, accelerating convergence and improving model robustness on the relatively smaller HAM10000 dataset.

Compared with other contemporary approaches such as GoogleNet, InceptionV3, and ResNet-based architectures, the proposed model yielded superior results in recall and F1-score, which are critical for medical applications where minimizing false negatives is paramount. The nearly perfect recall (100%) achieved indicates the model's strong sensitivity in detecting malignant lesions, a key requirement in skin cancer diagnosis. Performance evaluations also show minimal performance fluctuations between training and validation phases, affirming the stability of the model under varied image conditions. Additionally, the use of data augmentation and oversampling effectively addressed class imbalance, further contributing to consistent performance. Overall, the integration of image segmentation, transfer learning, and an efficient CNN backbone presents a practical, scalable, and accurate approach for skin lesion classification. This positions the proposed model as a viable candidate for deployment in real-world clinical settings and mobile diagnostic platforms.

### 5.1 Future direction

Future research should focus on enhancing the generalizability, interpretability, and accessibility of CNN-based skin disease detection systems. One key area involves curating and utilizing larger, ethnically diverse, and balanced datasets to reduce bias and improve diagnostic accuracy across various skin tones and conditions. Additionally, integrating multimodal inputs such as patient demographics, clinical history, and dermoscopic images can provide a more holistic diagnostic approach.

Advancements in model architecture, including the adoption of Vision Transformers (ViTs) and lightweight CNN variants optimized for real-time inference, offer promising directions for improving performance and computational efficiency. Incorporating explainable AI (XAI) techniques will also be essential in gaining clinical trust by providing transparent, visual justifications for predictions.

Moreover, there is significant potential in deploying these models on mobile or embedded devices for point-of-care diagnosis in low-resource settings. Such implementations could support early detection and timely treatment, particularly in underserved regions. Future studies should also prioritize clinical validation through partnerships with healthcare institutions to ensure real-world applicability and regulatory compliance.

## 6 Conclusion

This research introduced a comprehensive deep learning framework for the accurate classification of skin lesions, integrating K-Means clustering for image segmentation, transfer learning for knowledge reuse, and the MobileNet architecture for efficient and lightweight computation. The proposed method was rigorously tested using the HAM10000 dataset, and the experimental results consistently demonstrated superior performance across key evaluation metrics, including accuracy (98.20%), precision (93.75%), recall (100%), F1-score (97.10%), and AUC (98.00%). These outcomes affirm the framework's capability to deliver high diagnostic reliability, particularly in identifying malignant lesions with minimal false negatives—an essential requirement in dermatological applications where delayed or missed diagnoses can have serious implications.

The success of this model can be attributed to its modular design: the K-Means segmentation effectively isolates the lesion area from the surrounding skin, allowing the deep learning model to focus on clinically relevant features; transfer learning enables the utilization of rich visual

knowledge from large datasets like ImageNet, mitigating the limitations of smaller medical datasets; and MobileNet ensures computational efficiency, making the framework suitable for deployment in real-time and low-resource environments, including mobile health applications and point-of-care diagnostic tools. In comparison with existing models in the literature such as GoogleNet, ResNet, DenseNet, and InceptionV3 the proposed model outperformed in both sensitivity and overall balance between classification metrics, highlighting its robustness and reliability for real-world use. Its lightweight structure, combined with high accuracy and explainability, makes it an ideal candidate for integration into dermatology practice, telemedicine platforms, and portable diagnostic devices.

**Author contributions** (i) Akinbowale Nathaniel Babatunde initiated the research idea, defined the objectives, and provided overall supervision throughout the project. He also guided the structure of the manuscript and ensured alignment with the study's goals. (ii) Bukola Fatimah Balogun led the data collection process, performed data cleaning and preprocessing, and contributed significantly to the implementation of the deep learning models used in the study. (iii)Ayodele Joseph Babatunde was responsible for conducting the statistical analysis, setting up the evaluation framework, and validating the model's performance using standard metrics to ensure scientific rigor. (iv) Salman Olatunji Isiaka handled the programming aspects, including algorithm development and model training. He also contributed to the interpretation of the results and participated in drafting the methodology section. (V) Joseph Bamidele Awotunde reviewed the manuscript for technical accuracy, helped refine the model architecture, and provided expert feedback on the machine learning techniques applied. (vi) Abidemi Emmanuel Adeniyi conducted the literature review, analyzed existing studies for benchmarking, interpreted the final results in the context of related work, and prepared the manuscript for final formatting and submission.

## Declarations

**Conflict of interest** All the authors wish to declare that no conflict of interest to declare in writing and submitting this manuscript.

## References


Abbas Q, Gul A (2022) Detection and classification of malignant melanoma using deep features of NASNet. SN Comput Sci 4(1):21

Abdar M, Samami M, Mahmoodabad SD, Doan T, Mazoure B, Hashemifesharaki R, Liu L, Khosravi A, Acharya UR, Makarenkov V, Nahavandi S (2021) Uncertainty quantification in skin cancer classification using three-way decision-based Bayesian deep learning. Comput Biol Med 135:104418. https://doi.org/10.1016/j.compbiomed.2021.104418

Alenezi F, Armghan A, Polat K (2023) A multi-stage melanoma recognition framework with deep residual neural network and hyperparameter optimization-based decision support in dermoscopy images. Expert Syst Appl 215:119352

Aljohani N, Turki H (2023) Comparative analysis of deep learning architectures for skin cancer classification. Comput Biol Med 145:105–113

Alwakid G, Gouda W, Humayun M, Sama NU (2022) Melanoma detection using deep learning-based classifications. Healthcare 10(12):2481

Bassel A, Abdulkareem AB, Alyasseri ZAA, Sani NS, Mohammed HJ (2022) Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. Diagnostics 12(10):2472

Celebi ME, Aslandogan YA, O'Leary RL (2007) A survey of skin lesion detection and classification methods. J Med Imaging 2(4):150–160

Cícero LF, Silva RA, Souza AM (2016) *Direct digital imaging for medical diagnosis has become popular with modern computational power and machine learning* methods. J Med Imaging 3(2):1–10

Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, Halpern A (2018) Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration

Cortes C, Vapnik V (2021) Support-vector networks. Mach Learn 20:273–297

Esteva A, Kuprel B, Novoa RA (2017) Dermatologist-level classification of skin cancer with deep neural networks. J Med Imaging 4(1):100–110

Fix E (1985) Discriminatory analysis: nonparametric discrimination, consistency properties (Vol. 1). USAF school of Aviation Medicine

Gajera HK, Nayak DR, Zaveri MA (2023) A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. Biomed Signal Process Control 79:104186

Ganster H, Pinz A, Steger C (2001) Detection of tumors in medical images using gradient-based methods. J Med Imaging 1(3):45–55

Gouda W, Sama NU, Al-Waakid G, Humayun M, Jhanjhi NZ (2022) Detection of skin cancer based on skin lesion images using deep learning. Healthcare 10(7):1183

Gutman D, Codella NC, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A (2016) Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1605.01397.

Ho TK (1995) Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. Vol. 1. pp. 278–282

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 79(8):2554–2558

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., &Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360

Ibraheem I, Ramay SA, Abbas T, Hassan R, Khan S (2024) Identification of skin cancer using machine learning. J Comput Biomed Inform, 7(02)

Inthiyaz S, Altahan BR, Ahammad SH, Rajesh V, Kalangi RR, Smirani LK, Rashed ANZ (2023) Skin disease detection using deep learning. Adv Eng Softw 175:103361

Kassem MA, Hosny KM, Fouad MM (2020) Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access 8:114822–114832

Mendonca T, Ferreira PM, Barata C, Marques JS, Rocha J, Rozeira J (2016) PH2-A Public Database for the Analysis of Dermoscopic Images. Dermoscopy Image Analysis

Nasr-Esfahani M, Gholamian M, Khodadadi M (2016) Deep learning in medical imaging: a review of methods and applications. J Med Imaging 3(3):45–58

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on

learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1).

Tschandl P, Rosendahl C, Kittler H (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 5(1):1–9

Yu L, Chen H, Dou Q, Qin J, Heng PA (2017) Automatic segmentation of liver and tumor from CT images using deep learning. J Med Imaging 4(2):120–130

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.