Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

Optimizing Skin Lesion Diagnosis Using Deep Learning and Novel Noise Image Filtering Algorithm

Adel A. Ahmed[1*], Omar M. Barukab[1], Ahmad A. Alzahrani[2]

[1] Information Technology Department, Faculty of Computing and Information Technology-Rabigh, King Abdulaziz University, Jeddah, 25729, Saudi Arabia
[2] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
E-mail: aaaabdullah1@kau.edu.sa

Received: 18 June 2025; **Revised:** 5 August 2025; **Accepted:** 5 August 2025

Abstract: Non-Reference Image Quality Assessment (NR-IQA) has made a significant contribution to medical imaging services such as X-ray, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scans. Integrating deep learning artificial intelligence algorithms into NR-IQA has revolutionized medical services, enhanced diagnostic accuracy, and improved patient care management. Among deep learning algorithms, Convolutional Neural Networks (CNNs) are extensively utilized for object detection, image recognition and classification, and semantic segmentation. However, the CNN algorithm has a crucial challenge with noisy datasets that can decrease decision-making accuracy. Thus, applying NR-IQA to CNN algorithms can enhance feature extraction and diagnosis accuracy. This research proposes an efficient skin lesion diagnosis by integrating NF-IQA into the CNN algorithm. The proposed system calculates NR-IQA using three optimal metrics: average information entropy, chromatic level factor, and average luminance. By filtering low-quality images, NR-IQA reduces the noisy images in both the training and testing datasets, which enables the CNN to focus on clinically relevant features. Additionally, it is designed to diagnose seven types of skin lesions. The proposed system performance is evaluated using Area Under the Curve-Receiver Operating Characteristic (AUC-ROC), confusion matrix, and classification report, which have been used to measure accuracy, loss, precision, recall, and F1 score. The proposed system outperforms the baseline methods and achieves accuracies of 100%, 93.3%, and 85% on balanced training, validation, and testing datasets, respectively. The main advantage of the proposed system is to improve the learning efficiency and robustness of CNN-based classifiers, and to offer rapid skin lesion diagnosis, thereby enhancing AI-driven medical care.

***Keywords*:** NR-IQA, CNN, skin lesion, Machine Learning (ML)

MSC: 68T07, 68T45

Abbreviation

ACLF	Average Chroma Level Factor
AE	Average Information Entropy
AKIEC	Actinic Keratosis
AL	Average Luminance

DOI: https://doi.org/10.37256/cm.6420257508

AUC	Area Under the Curve
BCC	Basal Cell Carcinoma
BKL	Benign Keratosis
CLF	Chromatic Level Factor
DF	Dermatofibroma
E	Information Entropy
IQAF	Image Quality Assessment Formulation
ME	Melanoma
ML	Machine Learning
NAL	Normalized Average Luminance
NR-IQA	Non-Reference Image Quality Assessment
Nv	Nevus
ROC	Receiver Operating Characteristic
VASC	Vascular Lesion

1. Introduction

Recently, skin lesions and cancers have become more widespread due to environmental pollution which includes radiation, chemical dust, and climate change. The most common skin lesions that exist today are Basal Cell Carcinoma (BCC), Nevus (NV), Dermatofibroma (DF), Melanoma (ME), Benign Keratosis (BKL), Actinic Keratosis (AKIEC) and Vascular Lesion (VASC). Figure 1 depicts those seven skin lesions that were obtained from the International Skin Imaging Collaboration (ISIC) archive. Due to the increasing number of skin lesion patients, dermatologists cannot deal with this pathological phenomenon which usually requires appointments and leads to long waiting times and crowded clinics. Delayed diagnosis can cause serious complications including metastasis or spread to other body parts. Furthermore, the rapidly widespread skin lesion among humans is akiec, which causes the dermatological phenomenon to become pandemic in some countries. To address this, collaborative efforts between dermatologists and AI-driven solutions are now essential.

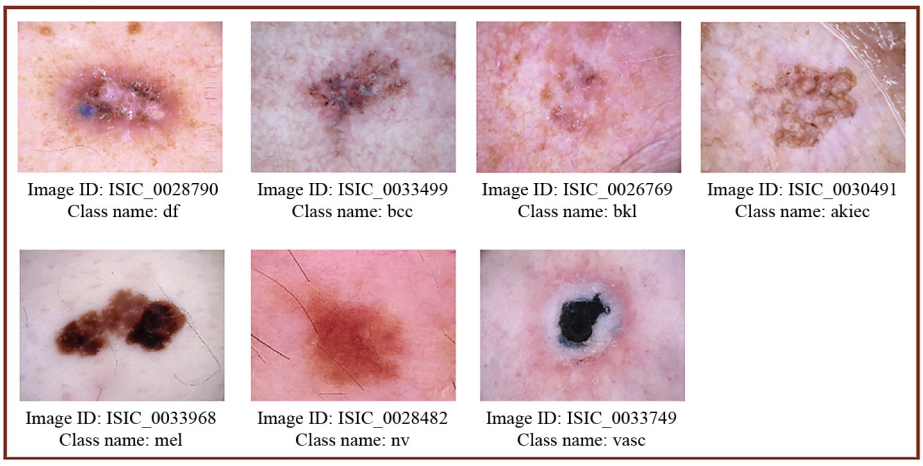


Figure 1. Seven skin lesions from HAM10000 dataset

One of the most effective diagnostic solutions for skin lesions is using Machine Learning (ML), which is utilized to create optimized diagnostic tools for analyzing medical images [1]. For example, pattern recognition is used to identify specific diseases using X-rays or MRI scans. Similarly, deep learning algorithms not only provide patients with initial diagnoses but also assist physicians in making faster, more accurate decisions.

Among deep learning algorithms and the availability of high-quality datasets, CNN is the most effective ML algorithm that is used to extract lesion features from skin medical images. In contrast, CNN might provide less accuracy if the dataset contains noisy and poor-quality images as will be proven in this research [2–4]. Even with the noisy dataset, CNN still has several advantages which can be described as follows [5–12]:

• **Image Analysis Proficiency:** CNN is an optimal solution for detecting objects, feature extraction, and processing and analyzing X-rays, MRI scans, and CT scan images. This makes CNN also suitable for skin lesion diagnostics.

• **Accuracy and Efficiency:** If CNN is applied to a high-quality dataset, it can diagnose complicated dermatology images with excellent accuracy, which has a significant impact on patient healthcare. Otherwise, it provides less accuracy with noisy datasets.

• **Automated Feature Extraction:** Most of the ML algorithms require manual feature extraction, which relies on heuristic data collection. This leads to time-consuming and prone to human error. CNNs automate feature extraction from the input images by learning the correlated features, which reduces human error and increases the efficiency of the diagnosis process.

• **Handling Large Datasets:** The diagnostic devices generate a huge number of images, and CNNs are perfect for handling and analyzing these enormous medical images with moderate time processing.

• **Customization and Adaptability:** CNNs can be adapted to specific medical applications, which enables them to be flexible and adaptable to a variety of medical data and diagnosis requirements because of their ability to be customized.

Although CNN is an effective ML technique for visual data analysis, it has drawbacks and restrictions. One significant drawback is vulnerability to noisy images or poor-quality datasets, which can skew the retrieved features and reduce accuracy. Furthermore, large, varied, and well-annotated datasets are required for CNN to generalize diagnosis efficiently, which can be costly and time-consuming. The most effective drawback of CNN is overfitting, which happens when CNN learns particular patterns from small training datasets rather than generalizable features. These challenges need careful integration with NR-IQA to ensure robust and reliable performance. The main goal of the proposed research is to provide fast, convenient, and dependable first assessments.

1.1 *Problem statement and research motivation*

Applying NR-IQA to filter noisy and poor-quality images from datasets poses significant challenges including decreasing the number of images in certain classes, which may compromise the accuracy of the CNN due to insufficient data. Furthermore, while numerous performance metrics exist for calculating NR-IQA, their selection requires careful consideration [13–17]. To address these challenges, the proposed method aims to enhance skin lesion diagnosis by combining NR-IQA with CNN to achieve high precision and to offer a practical and creative diagnosis solution. Consequently, the proposed system will reduce the congestion in hospitals and improve appointment availability for only the most critical cases.

1.2 *Research contribution*

The following contributions are reported in this research:

• **Key Methodological Contributions**

- A novel, lightweight, and effective NR-IQA framework is proposed for evaluating dermoscopic images. The framework assesses image quality using three complementary quality metrics: information entropy, chromatic level factor, and average luminance.

- A heuristic exhaustive search method is employed to determine the optimal weight distribution for the three NR-IQA metrics, ensuring accurate and generalizable quality estimation.

• **Implementation and Performance Contributions**

- The proposed NR-IQA framework is used to filter out noisy images from the HAM10000 dataset before CNN training and evaluation. This preprocessing step significantly enhances CNN's performance, achieving up to 100% training accuracy, 93% validation accuracy, and 85% testing accuracy.

- A balanced dataset training approach is introduced, where a higher weight is assigned to underrepresented classes, addressing class imbalance and further improving classification performance. The proposed model outperforms several state-of-the-art CNN-based methods by up to 10% in test accuracy.

The rest of this paper is organized into the following: Section 2 presents the related works on skin lesion diagnosis. The proposed system model is explained in Section 3. Also, Section 4 describes the Experimentation and Performance Evaluation. Section 5 describes the future work, and finally, Section 6 presents the conclusion.

2. Related works on skin lesions diagnosis

Deep learning algorithms can be used in several medical applications such as skin lesion diagnosis, lesion segmentation, and the early identification of skin cancers [17]. Esteva et al. [18] showed that CNNs can reliably diagnose skin cancer from dermoscopic images. However, issues like noisy datasets and unbalanced classes affect CNN accuracy in real scenarios. Bi et al. [19] and Tschandl et al. [20] emphasized the necessity of image quality assessment to guarantee reliable diagnostic results, particularly when addressing noisy and imbalanced datasets. Recent studies by Sadia et al. [21] developed a Computer-Assisted Diagnosis (CAD) framework that can detect skin lesions at an early stage. The authors in [21] adjusted the CNN model with seven convolutional layers to diagnose three skin lesions with 87.64% accuracy. Furthermore, the research by Mohamed et al. [22] utilized a CNN model with advanced configuration for data augmentation, image resizing, and normalization to enhance CNN accuracy in skin cancer diagnosis. However, the authors claim an accuracy of more than 97%, which contradicts the accuracy calculation in the classification report of 76%. Moreover, several researchers used CNN to classify two skin lesions only (binary classification) with high accuracy. One of the recent binary image classifiers was called TurkerNet which was proposed by Turker et al. [23]. The study in [23] utilized deep learning with four distinct blocks to accurately classify skin images as benign (non-cancerous) or malignant. Similarly, Abbas et al. [24] utilized a deep learning model for classifying melanoma in AM and BN cases. However, the research study in [24] trained a deep learning model using a dermoscopic image dataset from the Yonsei University Health System in South Korea. Also, Kemal et al. [25] proposed a combination of CNN and one-versus-all models. In [25], a single CNN trained on the HAM10000 dataset achieved 77% accuracy on seven skin lesions while the second model achieved a precision of 85.7%. Shanthi et al. [26] utilized a CNN based on a previously trained model AlexNet and it applied to the DermNet dataset. The research study in [25] classifies four skin lesions: acne, keratosis, eczema herpeticum, and urticaria with each category including 30 to 60 samples. The diagnostic accuracy in [26] was 85.7%, 92.3%, 93.3%, and 92.8% for acne, keratosis, eczema herpeticum, and urticaria respectively. In a specific study [27], Bajwa et al. proposed a computer-aided diagnosis system that trained four deep learning layers to achieve optimal accuracy on the DermNet and ISIC-2018 datasets. The K-Fold method was used for data separation, and the class imbalance was resolved using a weighted loss method.

Despite these advances in state-of-the-art works, certain limitations have been reported in the previous literature studies [8, 18–27] which can be specified in two facts: first, the literature works applied only image enhancement such as segmentation and augmentation on poor quality images which have limited effect on the CNN accuracy. However, the proposed NR-IQA is lightweight and dynamically optimized using the threshold. Second, the literature works did not filter poor-quality images, but they tried to enhance them. In contrast, the proposed system filters the noisy images from the dataset in three phases: training, validation, and testing. Therefore, the skin lesion diagnosis using NR-IQA and CNN has great diagnostic precision and accuracy enhancement.

3. System proposed model

The architecture of the proposed skin lesion diagnosis consists of four components, which are input quality assessment, CNN-powered analysis, results evaluation, and output diagnostic in Graphical User Interface (GUI) as illustrated in Figure 2. The first component of the proposed system is a quality assessment model based on the innovative NR-IQA formula to filter noisy images from the input dataset. The output of the first component is high-quality dataset.

The second component is CNN-Power analysis, which is divided into two sub-components: training with validation and testing the high-quality dataset. The third component is the results evaluation for training, validation, and testing. Finally, the results of the diagnosis in the third step appeared in the GUI interface.

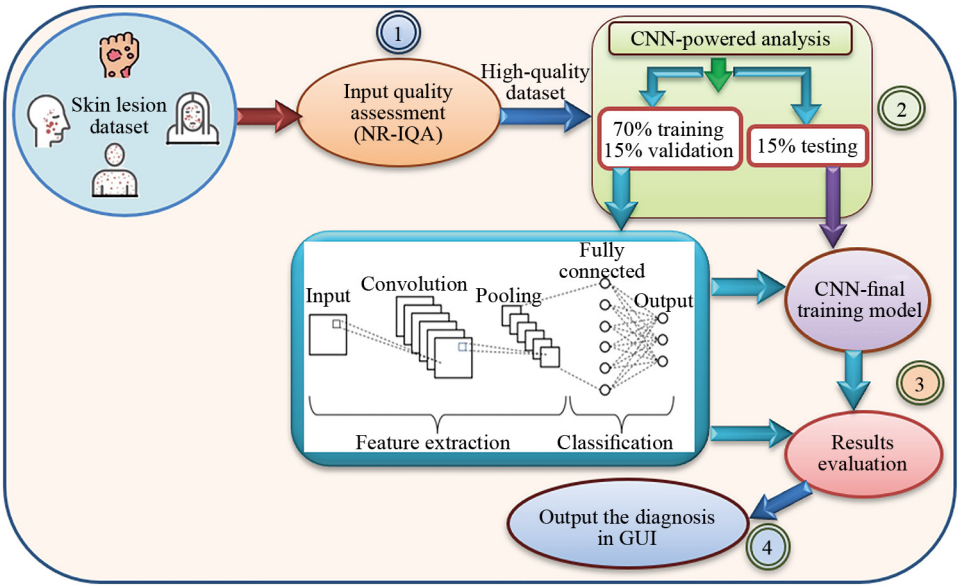


Figure 2. Proposed diagnostic system architecture diagram

3.1 *Input quality assessment*

In this phase, the training and validation datasets of skin lesions are assessed using the proposed NR-IQA formula, which uses three mathematical indicator metrics: information entropy, chromatic level factor, and average luminance. The main reason for selecting these three metrics is their ability to capture essential aspects of image quality that are highly relevant to medical diagnosis. Information entropy measures the quantity of texture and structural details in an image, which are critical for distinguishing between different types of skin lesions. Chromatic level factor evaluates the distribution of color components, which is particularly important for identifying color variations in lesion pigmentation, which is considered a key diagnostic indicator in dermatology. Average luminance quantifies the overall brightness of the image, which detects unclear or dark images that could affect feature extraction. Together, these three non-reference indicators provide a comprehensive assessment of image quality, and they ensure that only diagnostically useful images are retained for CNN training and inference. Furthermore, the recent state-of-the-art [28–31] specifies that the aforementioned three parameters have powerful perceptual features along with colorfulness and dark channel features.

The Average information Entropy (AE) reflects the quantity of information for each pixel in the skin lesion image. It presumes that a good-quality image needs to have sufficient information pixels. In contrast, the image is considered poor quality if the AE of the pixels is less than a certain threshold. The amount of information in a gray image is quantified by information Entropy (E), which can be expressed as follows [32, 33]:

$$E = -\sum_{k=0}^{L-1} P(r_k) \log_2 (P(r_k)) \tag{1}$$

where $P(r_k)$ represents the pixel distribution probability at the r_k gray level. The maximum value of E (8 bits with 256 gray levels) means that the image has an ideal equalization histogram. However, the image with a single gray level means

E is equal to zero. To assess the quality of a color image, the AE of the RGB components (Red, Green, and Blue) should be computed as:

$$AE = \sqrt{\frac{1}{3}\sum_{n=0}^{2} E^2(n)} \text{ where } n = 0,\ 1,\ 2 \tag{2}$$

where n represents one of the three color components of an image. If the AE has a very low or zero-bit representation, this means that the input image has poor information and hence is considered a low-quality image.

The main components of a color image are chrominance (color) and luminance (brightness). The chromatic level factor, CLF(n), is a measure of the chromatic level complexity of individual color components in an image. The normalized level is determined by [32, 33]:

$$CLF = \frac{1}{256}\sum_{m=0}^{256} \text{Count}\left[P(n,\ m)\right] \text{ where } n = 0,\ 1,\ 2 \tag{3}$$

In Equation (3), Count is used to count the chromatic spectrum, while $P(n,\ m)$ shows the distribution of pixel numbers for different color components. The Average Chroma Level Factor (ACLF) can be calculated as:

$$ACLF = \sqrt{\frac{1}{3}\sum_{n=0}^{2} CLF^2(n)} \text{ where } n = 0,\ 1,\ 2 \tag{4}$$

The highest level of the chromatic range for every color component is 256, and Equation (3) represents the percentage, which means the maximum is one. Higher ACLF corresponds to good image quality.

The third metric is an Average Luminance (AL). A high-resolution image must possess suitable brightness. Poor image quality results from excessive brightness or darkness. The AL for an ideal image with 256 gray levels on the 8-bit system has optimally visible luminance at 127.5. It can be calculated for the three-color components (Red, Green, and Blue) as [32, 33]:

$$AL = \sqrt{\frac{1}{3}\sum_{n=0}^{n}\left(\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} L(n,\ x,\ y)\right)} \text{ where } n = 0,\ 1,\ 2$$
$$L(x,\ y) = 0.299 \times R(x,\ y) + 0.587 \times G(x,\ y) + 0.114 \times B(x,\ y) \tag{5}$$

where $L(n,\ x,\ y)$ represents the color n luminance at the pixel $(x,\ y)$. If we consider AL_{op} as the optimal visible luminance, high image quality has AL_{op} closer to 127.5. To approximate the level of AL, it should be normalized as follows:

$$NAL = 1 - \frac{|AL - \text{AL}_{op}|}{\text{AL}_{op}} \tag{6}$$

When AL = AL_{op}, NAL = 1, signifying that the image's luminance has been optimized for 127.5, which means the optimal visual quality for that specific image. Therefore, NAL signifies the difference between AL and AL_{op}.

3.1.1 *Image quality assessment formulation*

The above three metrics can be combined into a single formula to provide an initial image quality assessment which is referred to as IQAF:

$$IQAF = \alpha AE/8 + \beta ACLF + \gamma NAL \text{ where } \alpha,\ \beta,\ \gamma > 0 \text{ and } \alpha + \beta + \gamma = 1 \tag{7}$$

The weightage values of α, β, and γ are estimated by exhaustive search through experimental work. As will be described in the experimental section, the optimal values for α, β, and γ are 0.3, 0.2, and 0.5 respectively. Furthermore, IQAF reflects the probability for image quality, which means the high image quality has IQAF near to 1 and vice versa. In the experiment section, we will show that the threshold of IQAF equal to 0.5. This means if IQAF has a greater value than 0.5, the skin image is considered an acceptable quality, and it can be used in the CNN model for dermatology diagnostics. In contrast, if IQAF has less value than 0.5, which causes the proposed system to exclude the skin image from the dataset.

3.2 *CNN-powered analysis*

CNN is the optimized method to extract the features of each type of skin lesion. The proposed system uses three layers of the CNN model: Rectified Linear Unit (ReLU) layer, pooling layer, and fully connected layer as can be illustrated in step 2 of Figure 2. The convolution operations are utilized by the convolutional layer to extract local features while the ReLU activation function introduces non-linearity by zeroing out negative values. Moreover, the pooling layer reduces the spatial dimensions of the feature maps to decrease computation. Finally, the fully connected layer integrates the high-level features and produces the class predictions. The proposed system tuned the parameters of the former layers with the advantage of hierarchical organization, which enables CNN to extract complicated patterns from skin images [26]. Moreover, the training process in CNN consists of two essential phases: forward propagation and backpropagation. The input data is passed through the three layers of the network in the forward propagation to give out the prediction. Backpropagation plays a crucial role in adjusting the network parameters by calculating the gradient of the loss function with respect to the weights. It is interesting to know that the Stochastic Gradient Descent (SGD) is used to iteratively update the network weights.

3.2.1 *CNN dataset*

The dataset refers to the origin or repository data that is collected for training, validating, and testing the CNN model. The quality, quantity, and relevance of data from these sources are vital for the high accuracy and optimal performance CNN model. The proposed system uses the publicly available dataset of skin lesions called HAM10000 that was gained from ISIC which is a cooperation between academics and industry to reduce skin cancer death [20, 34]. HAM10000 dataset consists of de-identified dermoscopic images of skin lesions collected retrospectively. No identifiable private data was accessed, and no human subjects were directly involved. It contains seven classes of skin lesions: mel, nv, bcc, akiec, bkl, df, and vasc. This research uses standard distribution of dataset into training, validation and testing with 70%, 15%, and 15% respectively.

3.2.2 *Preprocessing dataset*

The processing of the dataset is divided into two categories: preprocessing of the training and validation of the dataset and preprocessing of the testing dataset. It includes the following preprocessing [35–37]:

• **Data Cleaning based on NR-IQA:** removing any corrupt, noisy, or poor-quality images from the dataset.

• **Data Preparation:** organizing data into training, validation, and testing, and applying data augmentation to enhance the dataset's diversity.

• **Data Augmentation:** applying transformations such as rotations, flips, and color adjustments to increase the diversity of the training data.

• **Normalization:** scaling pixel values to a range that is suitable for the CNN model.
• **Resizing:** ensuring the dimensions of all images are the same, which is a requirement for input into the CNN model.

3.2.3 *Training and optimizing model*

The CNN model uses the augmented dataset to apply training, which involves adjusting and tuning the CNN model parameters such as batch size, epoch, learning rate, weight, validation size, etc. Those parameters are used to optimize the accuracy with which objects are detected or to minimize the loss function. Furthermore, optimization has been used to minimize prediction errors. A single complete pass through the whole training image is called an epoch. Training models typically involve running multiple epochs. This means the model will go through the entire training dataset multiple times. Each epoch allows the model to learn and adjust its weights based on the refined dataset, gradually improving its performance. Too few epochs might lead to underfitting, where the model hasn't learned enough from the data. In contrast, excessive epochs may result in overfitting, causing the model to excel on the training dataset yet underperform on new testing data.

3.3 *Results evaluation*

The performance of the training model is monitored using a validation dataset which is also used to prevent overfitting in the final model. The output of the training model is called the final or best model. The testing dataset is used to assess the final model, and it evaluates the generalization capabilities of the final model. The training model has been evaluated using the validation set to assess the proposed system's performance in terms of accuracy, loss, precision, recall, and F1 score.

3.4 *Display diagnosis in GUI*

The final step is displaying the testing results in the confusion matrix, which shows the accurate percentage of the seven-skin lesions as described in Section 4. Also, the comparison between the actual and predicted skin lesions is illustrated in the confusion matrix GUI.

3.5 *Operation of the proposed model*

The pseudocode of the proposed skin lesion diagnosis model is shown in Algorithm 1. The operation of the proposed system is described as follows:

Step 1: Split the Dataset

The dataset images are divided into 70% training, 15% validation, and 15% testing, which are labelled based on seven skin lesions: akiec, bcc, bkl, df, mel, nv, and vasc.

Step 2: Initialize CNN

The CNN model is initialized with the number of layers, type of filter, kernel, the initial weights, biases, and the optimizer SGD.

Step 3: Estimation of Weightage and Threshold of IQAF Equation

The first part of this step is performing the CNN model on the training and validation datasets. The coefficients α, β, and γ of the IQAF equation are determined using an exhaustive search on the validation results (accuracy and loss). After that, the optimal threshold of IQAF is estimated based on validation results as well.

Step 4: Filtering Dataset

The proposed system assesses the image quality in training, validation, and testing datasets using Equation (7). After that, the duplicated images will be removed from the filtered datasets.

Step 5: Training, Validation, and Testing

The CNN model is trained using balanced and imbalanced techniques. In the imbalanced training, the class weight and data augmentation are applied. However, balanced training used only data augmentation. The validation and testing results are estimated.

Step 6: Performance Evaluation

Finally, the evaluation parameters in terms of accuracy, loss, precision, recall, and F1 score are calculated, and the graphical evaluation based on AUC-ROC, confusion matrix, and classification is presented in both balanced and imbalanced testing.

Algorithm 1 Enhanced Skin Lesion Diagnosis Pseudocode

Skin Lesion Diagnosis Model

Input: Input_image, Data_source;

Output: GUI Predictions

1: **Start Algorithm (Derm_diagnostic)**
2: Initialize_CNN_Model (no. of layers, filter, kernel, weights, biases, optimizer (SGD))
3: Filter dataset using IQAF; // remove noisy images
4: Filter duplicated images from dataset; // remove duplicated images
5: Split remaining dataset into 70% training, 15% validation, 15% testing
6: CNN_train (Input_image, training_data, labels, epochs, learning_rate)
7: **for** epoch=1 to epochs **do**
8: **for** each batch in training_data **do** // # Forward pass
9: Feature_map = Convolution (Input_image);
10: Activated_output = ReLU (Feature_map);
11: Pooled_output = MaxPooling (Activated_output);
12: Flattened_output = Flatten (Pooled_output);
13: Predictions = Dense (Flattened_output); // Prediction (class scores or probabilities)
14: Loss = CrossEntropyLoss (Predictions, labels);
15: Gradients = Backpropagation (Loss);
16: Update_weights_biases (Gradients);
17: **end for**
18: Validation_model (Input_image, validate_data, labels, epochs, learning_rate);
19: Calculate performance metrics (Train_loss, Train_accuracy, etc);
20: **end for**
21: **Return** Predictions
22: Get (Final_model.keras)
23: Test_use (Final_model.keras)
24: Test_Model (Input_image, test_data, labels, epochs, learning_rate)
25: Calculate performance metrics and classification report ()
26: Print Confusion Matrix ()
27: **End;** // **Algorithm**

4. Experimentation and performance evaluation

In this study, an experimental framework has been designed locally in HP ProBook 640 G8 Notebook with 16 GB of RAM and the source code has been developed using Python programming version 3.10, TensorFlow version 2.18, and Keras version 3.7. The performance of the proposed skin lesion diagnosis system has been evaluated based on the HAM10000 dataset, which contains 10,015 images and is considered one of the largest collections of multi-sources dermatoscopic images of common pigmented skin Lesions [20, 34]. The proposed quality assessment filters 539 low-quality images from the HAM10000 dataset. This preprocessing step is important to ensure model performance is not compromised by poor-quality images. Furthermore, 2,333 duplicate images have been removed from the former dataset to ensure that the dataset is free of redundancy and prevents the model from overfitting due to repeated poor samples. The remaining images (7,143) are divided into 70% (5,036 images) training, 15% (1,033 images) validation, and 15% (1,072

images) testing, which are labelled based on seven skin lesions: akiec, bcc, bkl, df, mel, nv, and vasc. Table 1 shows the distribution of images in each skin lesion class. Also, the layer configuration of the CNN model is described in Table 2. It is interesting to know that the mean and standard deviation for the experiment have been utilized with an allowance of 5% margin of error; the average results were estimated within a confidence interval exceeding 95%.

Table 1. Distribution of images in categories

Lesion	AKIEC	BCC	BKL	DF	ME	NV	VASC	Total
HAM10000	327	514	1,099	115	1,113	6,705	142	10,015
Without duplication	234	358	735	75	637	5,541	102	7,682
After filter	223	300	716	69	609	5,138	88	7,143
Training	157	212	504	49	430	3,624	62	5,038
Validation	32	43	104	10	88	743	13	1,033
Testing	34	45	108	10	91	771	13	1,072

Table 2. Configuration and resizing images

Parameters	Configuration
Image size	224 × 224 px
Input layer	RGB images resized to 224 × 224 × 3.
Conv2D layer 1	32 filters of size 3 × 3, ReLU activation, followed by a 2 × 2 Max. Pooling layer.
Conv2D layer 2	64 filters of size 3 × 3, ReLU activation, followed by a 2 × 2 Max. Pooling layer.
Conv2D layer 3	128 filters of size 3 × 3, ReLU activation, followed by a Batch Normalization layer and a 2 × 2 Max. Pooling layer.
Flatten layer	Convert the 2D feature maps into a 1D feature vector.
Dense layer	128 units with ReLU activation, and L2 regularization (λ = 0.0001).
Output layer	Dense layer with SoftMax activation for multi-class classification across 7 skin lesion categories.
Optimizer	Adam with a learning rate of 0.0001.
Loss function	Categorical cross-entropy.
Dropout	Not used.
No. of epochs	50
Batch size	64

4.1 *Performance evaluation*

Standard assessment metrics in terms of accuracy, loss, precision, recall, and the F1 score were utilized to assess the effectiveness of the proposed algorithm during the experiments. The definitions of the above metrics are described in Table 3, where TP stands for True Positives, TN represents True Negatives, FP denotes False Positives, and FN represents False Negatives. The Accuracy (AC) measures how frequently the model is correct, while the Precision (PR) indicates how often the model forecasts a class as positive compared to the overall total of positives across all classes. The Recall (RE) specifies the frequency with which the model accurately predicts the class as positive, while the F1-score represents the weighted average of precision and recall.

Table 3. Performance metrics

Metric	Configuration
Accuracy (AC)	(TP+TN)/(TP+TN+FP+FN)
Precision (PR)	TP/(TP+FP)
Recall (RE)	TN/(TN+FN)
F1 score (F1)	2*TP/(2*TP+FP+FN)

4.2 *Estimation weightage and threshold of IQAF equation*

The first experiment in the proposed system is to estimate the optimal weightage of IQAF (α, β, and γ). This has been performed using an exhaustive search on the training and validation datasets, while reserving the testing dataset for final evaluation only. The substitution of α, β, and γ is between 0.1 and 0.9 under the condition $\alpha+\beta+\gamma=1$. Several experimental works have been conducted using the default parameters of the CNN training model, and the validation accuracy and loss have been used to estimate effectively α, β, and γ. Table 4 and Figure 3 show the validation accuracy and loss metrics for the most optimal cases of α, β, and γ. As can be seen in Table 4, the optimal case is 0.3, 0.2, and 0.5 for α, β, and γ respectively.

Table 4. Estimation optimal weightage based on validation results

Case No.	P(α, β, γ)	Max (Val_Accuracy)	Min (Val_Loss)
1	P (0.2, 0.2, 0.6)	0.72	0.73
2	P (0.3, 0.1, 0.6)	0.74	0.7
3	P (0.3, 0.2, 0.5)	0.93	0.41
4	P (0.4, 0.2, 0.4)	0.71	0.74
5	P (0.4, 0.3, 0.3)	0.65	0.82
6	P (0.5, 0.2, 0.3)	0.69	0.77
7	P (0.5, 0.3, 0.2)	0.60	0.94
8	P (0.38, 0.25, 0.37)	0.70	0.76
9	P (0.45, 0.1, 0.45)	0.72	0.72

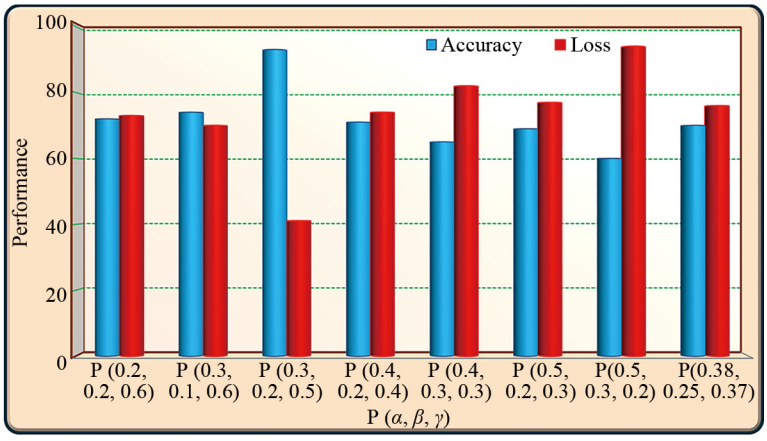


Figure 3. Performance of selected weightage

Table 5. Optimal threshold estimation based on validation results

Case No.	IQAF threshold	No. of noisy images	Max (accuracy)	Min (loss)
1	0.1	0	0.79	0.61
2	0.2	52	0.80	0.6
3	0.3	167	0.79	0.59
4	0.4	312	0.80	0.59
5	0.5	539	0.93	0.41
6	0.6	2,158	0.77	0.67
7	0.7	4,881	0.68	0.89
8	0.8	6,637	0.48	0.99
9	0.9	All images	-	-
10	1.0	All images	-	-

The second experiment has been implemented to estimate the optimal value of the IQAF threshold, which will be used to filter the noisy images from the training, validation, and testing datasets. As can be seen in Table 5, the threshold value varies between 0.1 and 1, which also varies the total number of excluded images from the training and validation datasets. Table 5 shows that the optimal threshold with high validation accuracy and less validation loss is 0.5. The impact of the proposed NR-IQA assessment method can be shown in Figure 4, which displays the significance of good quality in the training, validation, and testing datasets. Figure 4a depicts the excluded images from the HAM10000 dataset because of unclear and noisy information that will definitely affect the accuracy of skin lesion diagnosis. However. Figure 4b illustrates the effectiveness of the proposed NR-IQA to select optimal information images from the HAM10000 dataset.

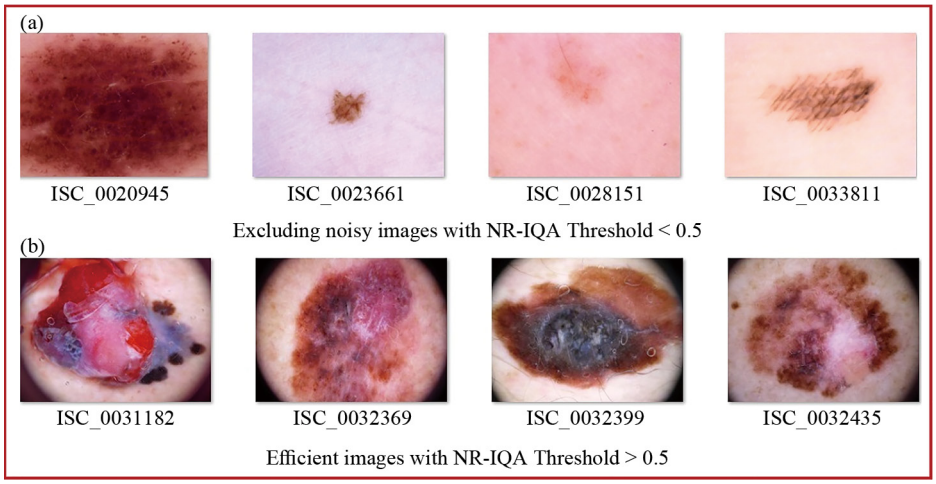


Figure 4. Impact of proposed NR-IQA assessment method: (a) excluding noisy image; (b) efficient informative images

4.3 *Training performance evaluation*

The initial implementation phase is training the proposed model in two different cases: balanced and imbalanced. Balanced training means all seven classes have the same weight regardless of the number of images in each class. However, imbalanced training means the weight of each class depends on the number of images and it can be expressed in the following equation:

$$W_C = \frac{N}{C \times n_C} \tag{8}$$

where W_C is the class weight, N is the total number of images, C is the number of classes, and n_C is the number of images in the C class. As can be seen in Equation (8), the class with fewer samples will gain a higher weight, which ensures higher importance during training. To address class imbalance in the training dataset, two strategies were employed. First, class weights were calculated based on the distribution of training labels using compute_class_weight from the sklearn library, and incorporated into the model training process. Second, data augmentation techniques including random rotations, translations, zooming, and horizontal flipping were applied using Keras' Image Data Generator to increase the diversity of training images and reduce overfitting. However, the balanced training used only data augmentation.

4.3.1 *Training performance analysis*

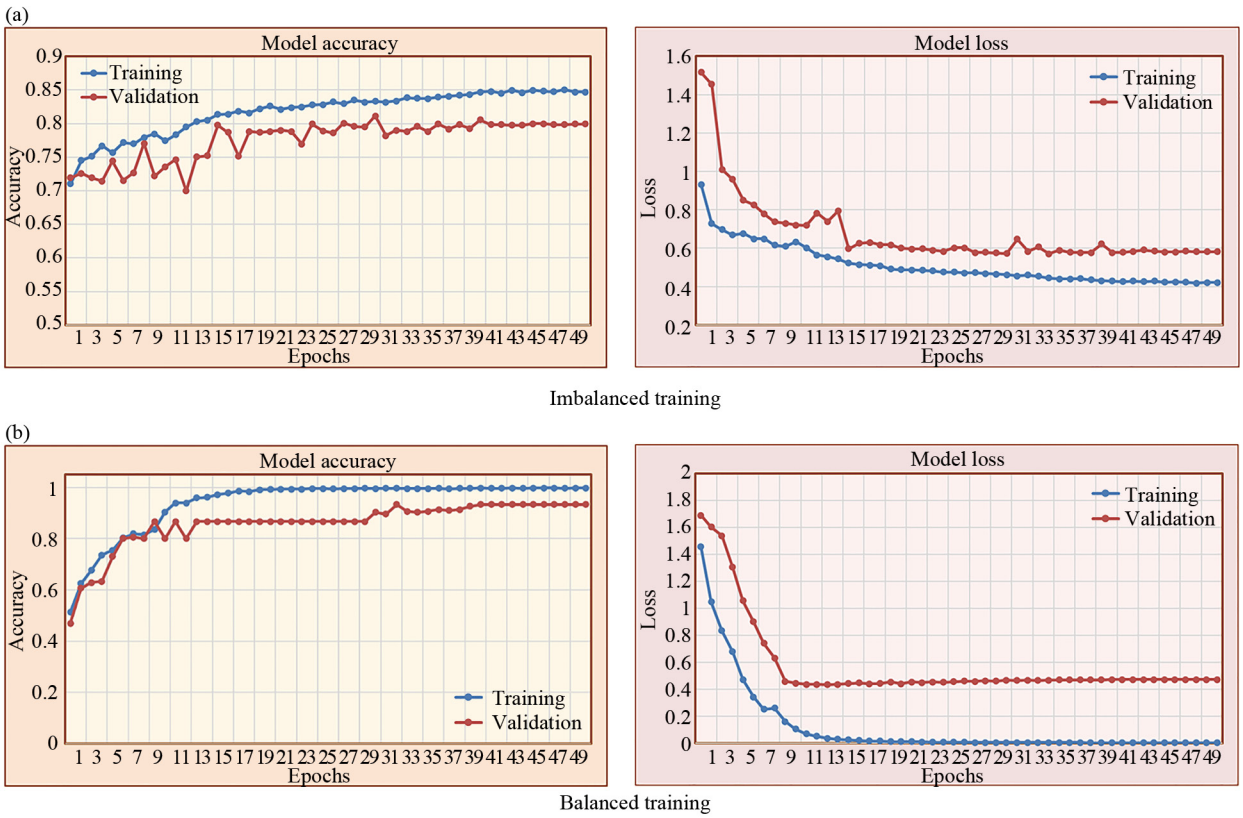


Figure 5. Accuracy and loss curves of the training and validation: (a) imbalanced; (b) balanced

Figure 5 presents the accuracy and loss curves during training for both balanced and imbalanced dataset experiments. In this figure, balanced training stably increased until it reached 100% accuracy compared to 85% in imbalanced training. Furthermore, balanced validation achieved 93.3% accuracy compared with 81% in imbalanced training. However, the loss curve decreased steadily in both balanced and imbalanced models. The balanced model achieved 40% training loss compared to 42% in the imbalanced model. In addition, the balanced model achieved 41.5% validation loss compared to 61.6% in the imbalanced model. The impressive performance of balanced training is mainly achieved due to the following reasons: Firstly, imbalanced training has classes with a higher number of samples, which influence the training

more significantly as the loss function focuses on reducing errors for these prevalent classes. Secondly, an imbalanced training model becomes expert at predicting the majority class and less effective with minority classes. As a result, the imbalanced training model provides an inadequate generalization on the validation set. Finally, the balanced training enhances its capacity to generalize throughout the dataset.

4.4 *Testing performance evaluation*

The output of the training phase in both balanced and imbalanced is called Final/Best model.keras which is used later to test the enhanced CNN method using the 1,072 unseen images. The important thing to test the proposed model on unseen dataset is to ensure the performance evaluation is fair and reflects real testing. Table 6 shows the classification reports including performance metrics, macro average and weighted average which are used to evaluate both models: balanced and imbalanced testing. The macro average computes the metric for each class individually and then takes the average (without considering class imbalance). This means it treats all classes equally, regardless of their size. However, the weighted average computes the metric for each class individually, but it weights the metric by the number of samples in each class which means it gives more importance to larger classes. In addition, Figure 6 shows the confusion matrix for balanced and imbalanced testing experiments.

4.4.1 *Test performance and generalization*

Table 6. Classification metrics report of the proposed model

Class (support)	Imbalanced testing			Balanced testing		
	Precision	Recall	F1-score	Precision	Recall	F1-score
AKIEC (34)	0.53	0.29	0.38	0.56	0.44	0.49
BCC (45)	0.47	0.49	0.48	0.57	0.7	0.63
BKL (108)	0.53	0.55	0.54	0.50	0.56	0.53
DF (10)	0.17	0.10	0.12	0.92	0.11	0.2
ME (91)	0.56	0.37	0.45	0.54	0.4	0.46
NV (771)	0.88	0.94	0.91	0.94	0.98	0.96
VASC (13)	0.89	0.62	0.73	0.97	0.77	0.86
Accuracy		0.8			0.85	
Total Loss		0.62			0.43	
macro avg	0.57	0.48	0.51	0.71	0.57	0.59
weighted avg	0.79	0.8	0.79	0.83	0.85	0.84

The classification metrics in Table 6 and the confusion matrix in Figure 6 validate the benefits of balanced training. As can be shown in Table 6, the weighted average accuracy improved from 80% in imbalanced to 85% in balanced testing while the overall loss drops from 62% to 43% as well. This is mainly due to the balanced model giving higher weight to the fewer images classes which increases the accuracy of the extracted features and the skin diagnosis. For skin lesions like df, vasc, and NV, balanced testing shows a significant increase in precision and F1-score compared to imbalanced testing, which reflects better diagnosis and prediction of those classes. For example, vasc's F1-score increased from 73% to 86%, while df's precision rose from 17% to 92%. Majority classes like NV perform exceptionally performance with the highest F1-score improvement from 91% to 96% due to increased precision and recall. The macro-average F1-score shows a significant enhancement with rising from 51% to 59%. The findings in this section highlight the advantages of using balanced testing in the skin lesion diagnosis, which enhance both the model's generalization and overall performance. Area Under the Curve (AUC) is an important measurement score that is used to evaluate the proposed classifier's performance at 0.5 decision threshold across a range of True Positive Rate (TPR) and False Positive Rate (FPR) pairs. Furthermore,

the Receiver Operating Characteristic (ROC) curve graphically represents those pairs of TPR and FPR at that threshold. Figure 7 illustrates the ROC curve of the training and testing datasets. In Figure 7a, the AUC scores for each class are almost 100% which means that the proposed classifier learned the training data very well. Furthermore, Figure 7b depicts drops in the AUC scores on the testing, but the differences between training and testing AUCs range from + 0.030 to + 0.136. Hence, the value remains greater than 86% which provides good generalization.

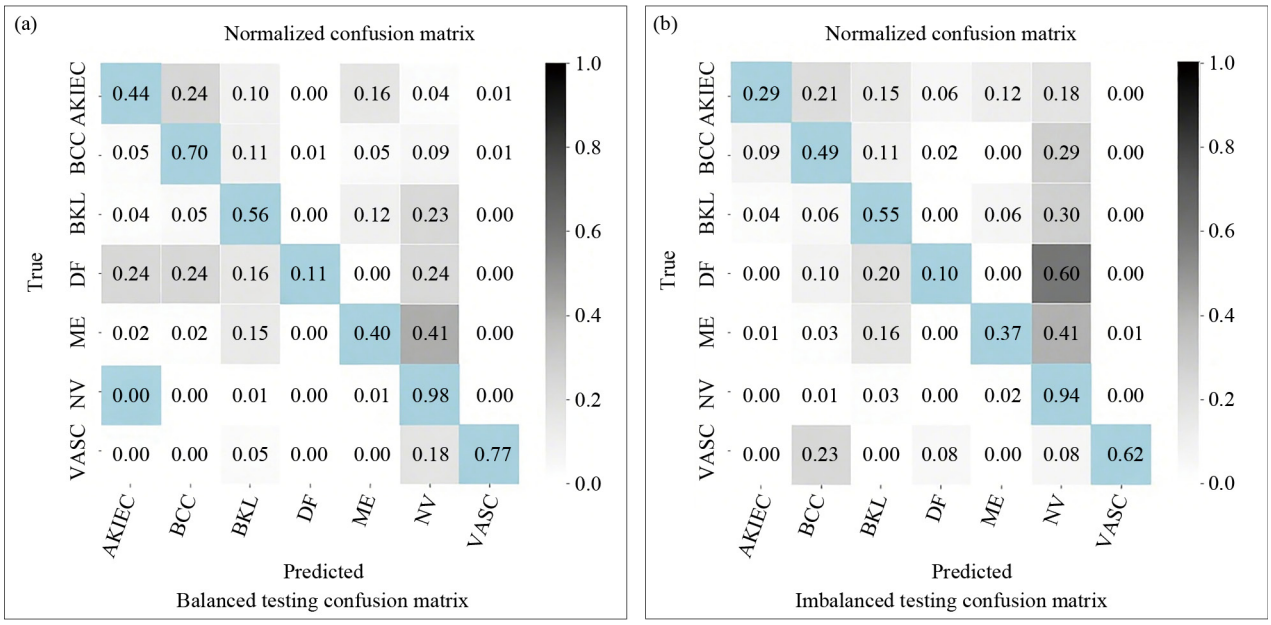
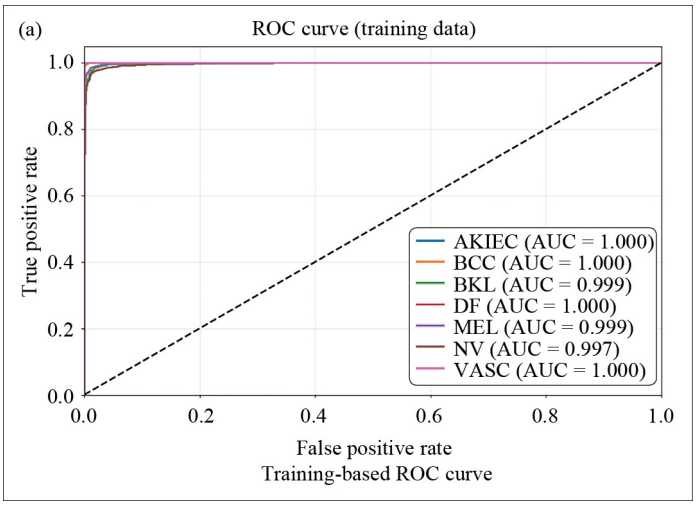


Figure 6. Confusion matrix of the balanced and imbalanced testing experiment



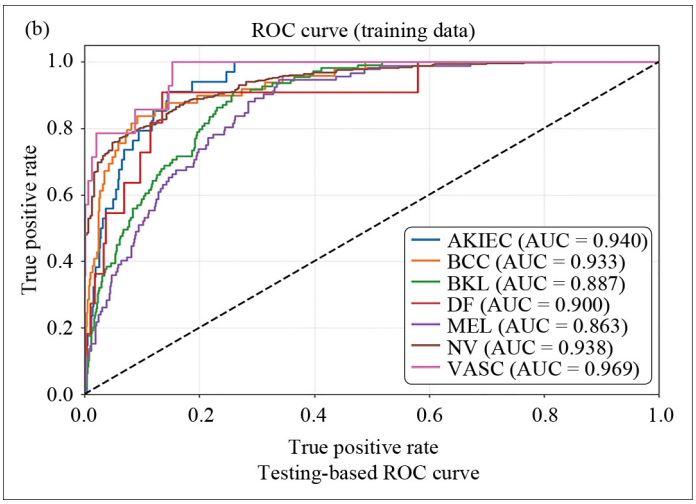


Figure 7. ROC and AUC for balanced training and testing

4.5 *Comparison of proposed model with complete dataset and related works*

This section shows the importance of the proposed NR-IQA by comparing the performance of the suggested NR-IQA with the unfiltered dataset (proposed work without NR-IQA) and the state-of-the-art related works. The comparison metrics include Accuracy (AC) and F1-Score (F1_S) across different skin lesions. Table 7 and Figure 8 show the achieved results of the proposed NR-IQA model compared to seven related works including the unfiltered dataset. As illustrated in Table 7, the proposed system with NR-IQA experiences 10.6% higher AC and F1_S compared to CNN without NR-IQA (unfiltered dataset). Also, the proposed method achieves the highest AC (85%) and F1_S (84%) compared to [22], which has 76% for both AC and F1_S, and [8], which has AC (77%) and F1_S (76.8%). This is mainly due to the proposed lightweight NR-IQA filters noisy images from datasets before applying the CNN model. Moreover, Shanthi et al. [9] and Sadia et al. [21] report higher accuracy of 89.5% and 87.6%, respectively, but they classified 4 and 3 types of skin lesions respectively. Finally, two methods classify only two types of skin lesions, such as Turker et al. and Abbas et al., which report high AC (92%) F1_ S (91%), and AC (90%) F1_S (90%) respectively. These results show the advantages of using NR-IQA to improve AC and F1_S for CNN when addressing the challenges of 7 types of skin lesions while maintaining strong performance metrics.

Table 7. Comparison between the proposed method and related works

Authors	Classifier	Classes	Accuracy (AC)	F1-score (F1_S)
Sadia et al. [21]	CNN model	3	87.6%	86%
Mohamed et al. [22]	CNN model	7	76%	76%
Turker et al. [23]	TurkerNet model	2	92%	90%
Abbas et al. [24]	CNN model	2	91%	90%
Kemal et al. [25]	CNN and CNN-OVA model	7	77%	76.8%
Shanthi et al. [26]	AlexNet model	4	89.5%	86%
Proposed work without NR-IQA	CNN model	7	76%	74%
Proposed work	CNN model with NR-IQA	7	85%	84%

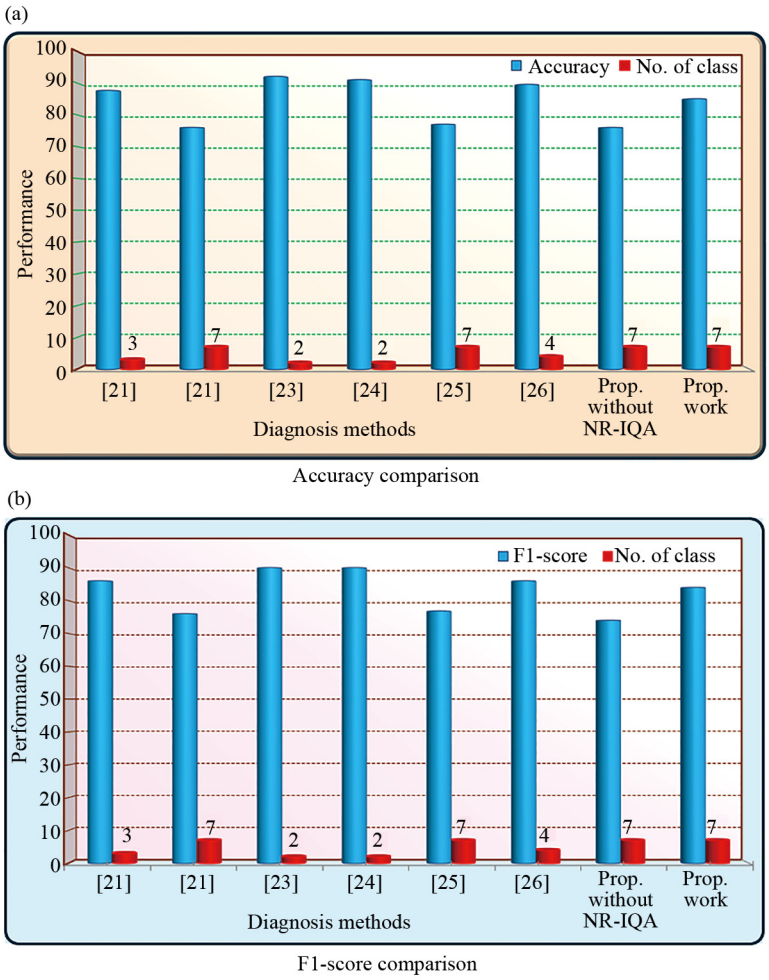


Figure 8. Skin lesion methods comparison with state-of-the-art: (a) accuracy; (b) F1-score

5. Limitations and future works

While the proposed NR-IQA-based filtering approach has shown promising results in improving CNN-based skin lesion diagnosis, it has some limitations:

• **Model Dependence:** The current framework is evaluated only on CNN-based models. It does not assess the performance of the NR-IQA filter on Transformer-based or hybrid architectures, which may offer better generalization in medical imaging.

• **Limited NR-IQA Scope:** The proposed NR-IQA metric uses three image quality indicators (entropy, chromatic level factor, and luminance), and it does not consider other less effective NR-IQA such as Average Bandwidth Factor (ABWF), Average Gradient (AG), sharpness, contrast distortion, and structural noise.

• **Limited Comparison with Other Filtering Pipelines:** The proposed filtering method is not directly compared against existing quality filtering or hybrid enhancement pipelines.

The proposed method focuses on CNN-based architectures to evaluate the impact of NR-IQA on skin lesion diagnosis; however, some recent state-of-the-art models are non-CNN models such as Vision Transformers (ViTs) and Swin Transformers. These models use self-attention mechanisms and have shown superior performance in many medical image analysis tasks. In future work, we aim to apply and evaluate the proposed NR-IQA filtering method within these Transformer-based frameworks. This will allow us to examine whether NR-IQA can improve performance in models

that leverage global context rather than local convolutions, which will produce the generalizability of our approach. Furthermore, the layers of the CNN model will be optimized, and Grad-CAM/SHAP will be used for model interpretability to enhance model transparency and support clinical decision-making. Finally, comprehensive comparisons with other filtering or hybrid strategies such as multi-stage denoising pipelines or ensemble-based quality assessment models will be investigated to demonstrate the superiority and robustness of the proposed method.

6. Conclusions

CNNs have been considered the most effective approach for medical image analysis due to their ability to independently learn and extract hierarchical features from raw image data. However, it has significant drawbacks when the dataset contains noisy or poor-quality images that can affect the retrieved features and accuracy. The proposed system introduces a novel, lightweight, and optimized NR-IQA framework formulated using three key metrics: information entropy, chromatic level factor, and average luminance. It enhances the skin lesion diagnosis system by filtering out noisy images from the dataset in training, validation, and testing. The findings show the performance of the enhanced CNN model, which outperforms state-of-the-art methods on balanced datasets and achieves a training accuracy of 100%, a validation accuracy of 93.3%, and a testing accuracy of 85%. It should be noted that despite these results, the dataset may not fully represent the diversity encountered in real-world clinical settings. The images used appear to be collected under relatively uniform conditions. This limitation could affect generalization to unseen populations or acquisition environments. Future work should involve testing the model on external datasets or under varied acquisition settings to validate its robustness. The main advantage of the proposed system is improving the learning efficiency and robustness of CNN-based classifiers and offering rapid skin lesion diagnosis, which will improve AI medical care.

Code and data availability

The data may be made available at reasonable request to the corresponding author, subject to institutional and ethical review board approval. The code used for this study is not publicly accessible but may be provided to qualified researchers upon reason able request to the corresponding author. No physical materials were used in this study.

Acknowledgment

This Project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (GPIP: 1791-830-2024). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

Conflict of interest

The authors declare that there are no financial interests, commercial affiliations, or other potential conflicts of interest that could have influenced the objectivity of this research or the writing of this paper.

References

[1] Ghosh H, Rahat IS, Mohanty SN, Ravindra JVR, Sobu A. A study on the application of machine learning and deep learning techniques for skin cancer detection. *World Academy of Science, Engineering and Technology International Journal of Computer Systems Engineering*. 2024; 18: 51-59. Available from: https://doi.org/10.5281/zenodo.10525954.

[2] Mehmood M, Abbasi SH, Aurangzeb K, Majeed MF, Anwar MS, Alhussein M. A classifier model for prostate cancer diagnosis using CNNs and transfer learning with multi-parametric MRI. *Frontiers in Oncology*. 2023; 13: 1225490. Available from: https://doi.org/10.3389/fonc.2023.1225490.

[3] Chumachenko D, Yakovlev S. Artificial intelligence algorithms for healthcare. *Algorithms*. 2024; 17: 105. Available from: https://doi.org/10.3390/a17030105.

[4] Islam MM, Hossain MB, Akhtar MN, Moni MA, Hasan KF. CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack. *Algorithms*. 2022; 15: 287. Available from: https://doi.org/10.3390/a15080287.

[5] Debelee TG. Skin lesion classification and detection using machine learning techniques: A systematic review. *Diagnostics*. 2023; 13: 3147. Available from: https://doi.org/10.3390/diagnostics13193147.

[6] Birkenfeld JS, Tucker-Schwartz JM, Soenksen LR, Avilés-Izquierdo JA, Marti-Fuster B. Computer-aided classification of suspicious pigmented lesions using wide-field images. *Computer Methods and Programs in Biomedicine*. 2020; 195: 105631.

[7] Mahmoud NM, Soliman AM. Early automated detection system for skin cancer diagnosis using artificial intelligent techniques. *Scientific Reports*. 2024; 14: 9749. Available from: https://doi.org/10.1038/s41598-024-59783-0.

[8] Hashmani MA, Jameel SM, Rizvi SSH, Shukla S. An adaptive federated machine learning-based intelligent system for skin disease detection: A step toward an intelligent dermoscopy device. *Applied Sciences*. 2021; 11(5): 2145. Available from: https://doi.org/10.3390/app11052145.

[9] Kaur R, GholamHosseini H, Lindén M. Advanced deep learning models for melanoma diagnosis in computer-aided skin cancer detection. *Sensors*. 2025; 25: 594. Available from: https://doi.org/10.3390/s25030594.

[10] Kaur R, GholamHosseini H, Sinha R, Lindén M. Melanoma classification using a novel deep convolutional neural network with dermoscopic images. *Sensors*. 2022; 22(3): 1134. Available from: https://doi.org/10.3390/s22031134.

[11] Rasel MA, Obaidellah UH, Kareem SA. Convolutional neural network-based skin lesion classification with variable nonlinear activation functions. *IEEE Access*. 2022; 10: 83398-83414. Available from: https://doi.org/10.1109/ACCESS.2022.3196911.

[12] Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R. The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules*. 2020; 10: 1123. Available from: https://doi.org/10.3390/biom10081123.

[13] Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence*. 2021; 15: 1-22. Available from: https://doi.org/10.1007/s12065-020-00540-3.

[14] Yadav SS, Jadhav SM. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*. 2019; 6: 113. Available from: https://doi.org/10.1186/s40537-019-0276-2.

[15] Yang R, Yu Y. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. *Frontiers in Oncology*. 2021; 11: 638182. Available from: https://doi.org/10.3389/fonc.2021.638182.

[16] Deepa N, Chokkalingam SP. Deep convolutional neural networks (CNN) for medical image analysis. *Social Science Research Network*. 2019; 3894929. Available from: https://doi.org/10.2139/ssrn.3894929.

[17] Gurucharan MK. *Basic CNN architecture*. Available from: https://www.upgrad.com/blog/basic-cnn-architecture/ [Accessed 24th November 2024].

[18] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017; 542(7639): 115-118. Available from: https://doi.org/10.1038/nature21056.

[19] Bi L, Kim J, Ahn E, Feng D, Fulham M. Dermoscopic image segmentation via multi-stage fully convolutional networks. *IEEE Transactions on Biomedical Engineering*. 2017; 64(9): 2065-2074. Available from: https://doi.org/0.1109/TBME.2017.2712771.

[20] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*. 2019; 5: 180161. Available from: https://doi.org/10.1038/sdata.2018.161.

[21] Malik SG, Jamil SS, Aziz A, Ullah S, Ullah I, Abohashrh M. High-precision skin disease diagnosis through deep learning on dermoscopic images. *Bioengineering*. 2024; 11(9): 867. Available from: https://doi.org/10.3390/bioengineering11090867.

[22] Musthafa MM, Mahesh TR, Vinoth Kumar V, Guluwadi S. Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification. *BMC Medical Imaging*. 2024; 24: 201. Available from: https://doi.org/10.1186/s12880-024-01356-8.

[23] Tuncer T, Barua PD, Tuncer I, Dogan S, Acharya UR. A lightweight deep convolutional neural network model for skin cancer image classification. *Applied Soft Computing*. 2024; 162: 111794. Available from: https://doi.org/10.1016/j.asoc.2024.111794.

[24] Abbas Q, Ramzan F, Ghani MU. Acral melanoma detection using dermoscopic images and convolutional neural networks. *Visual Computing for Industry, Biomedicine and Art*. 2021; 4: 25. Available from: https://doi.org/10.1186/s42492-021-00091-z.

[25] Polat K, Koc KO. Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all. *Journal of Artificial Intelligence Systems*. 2020; 2: 80-97. Available from: https://doi.org/10.33969/AIS.2020.21006.

[26] Shanthi T, Sabeenian RS, Anand R. Automatic diagnosis of skin diseases using convolution neural network. *Microprocessors and Microsystems*. 2020; 76: 103074. Available from: https://doi.org/10.1016/j.micpro.2020.103074.

[27] Bajwa MN, Muta K, Malik MI, Siddiqui SA, Braun SA, Homey B, et al. Computer-aided diagnosis of skin diseases using deep neural networks. *Applied Sciences*. 2020; 10(7): 2488. Available from: https://doi.org/10.3390/app10072488.

[28] Oszust M, Bielecka M, Bielecki A, Ste I, Obuchowicz R, Piórkowski A. Blind image quality assessment of magnetic resonance images with statistics of local intensity extrema. *Information Sciences*. 2022; 606: 112-125. Available from: https://doi.org/10.1016/j.ins.2022.05.061.

[29] Dugonik B, Dugonik A, Marovt M, Golob M. Image quality assessment of digital image capturing devices for melanoma detection. *Applied Sciences*. 2020; 10: 2876. Available from: https://doi.org/10.3390/app10082876.

[30] Varga D. No-reference image quality assessment based on the fusion of statistical and perceptual features. *Journal of Imaging*. 2020; 6(8): 75. Available from: https://doi.org/10.3390/jimaging6080075.

[31] Chen X, Zhang Q, Lin M, Yang G, He C. No-reference color image quality assessment: From entropy to perceptual quality. *EURASIP Journal on Image and Video Processing*. 2019; 2019: 77. Available from: https://doi.org/10.1186/s13640-019-0479-7.

[32] Liu YH, Yang KF, Yan HM. No-reference image quality assessment method based on visual parameters. *Journal of Electronic Science and Technology*. 2019; 17(2): 171-184. Available from: https://doi.org/10.11989/JEST.1674-862X.70927091.

[33] Stępień I, Oszust M. No-reference image quality assessment of magnetic resonance images with multi-level and multi-model representations based on fusion of deep architectures. *Engineering Applications of Artificial Intelligence*. 2023; 123: 106283. Available from: https://doi.org/10.1016/j.engappai.2023.106283.

[34] Kaggle. *Skin lesion analyzer tensorflow.js Web App*. Available from: https://www.kaggle.com/code/vbookshelf/skin-lesion-analyzer-tensorflow-js-web-app/input [Accessed 10th June 2025].

[35] Chiu TM, Li YC, Chi IC, Tseng MH. AI-driven enhancement of skin cancer diagnosis: A two-stage voting ensemble approach using dermoscopic data. *Cancers*. 2025; 17: 137. Available from: https://doi.org/10.3390/cancers17010137.

[36] Thwin SM, Park HS. Enhanced skin lesion segmentation and classification through ensemble models. *Engineering*. 2024; 5: 2805-2820. Available from: https://doi.org/10.3390/eng5040146.

[37] Cino L, Distante C, Martella A, Mazzeo PL. Skin lesion classification through test time augmentation and explainable artificial intelligence. *Journal of Imaging*. 2025; 11: 15. Available from: https://doi.org/10.3390/jimaging11010015.