Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **3**, September **2025**

Efficient Diagnosis of Dermatological Diseases Using GLCM and LBP with ML Classifiers

Wala Jasim Al-Shamari*

Department of Computer Engineering
College of Engineering, University of Baghdad
Baghdad, Iraq
gs22.wjfauadh@coeng.uobaghdad.edu.iq

Ahlam Hanoon Al-Sudani

Department of Computer Engineering
College of Engineering, University of Baghdad
Baghdad, Iraq
assis.prof.a.hanoon@coeng.uobaghdad.edu.iq

*Corresponding author: Wala Jasim Al-Shamari
Received July 3, 2025, revised July 18, 2025, accepted July 21, 2025.

Abstract. *Skin conditions are common worldwide and involve rapid and precise diagnosis in order provide appropriate care and reduce complications. Based on manually extracted histological features, an intelligent framework based on machine learning algorithms was created in this study to diagnose and classify skin diseases. Seven categories of skin lesions from the HAM10000 database were used. A number of preprocessing techniques were applied to the images, such as the Wiener filter for noise reduction and the Black-Hat transformation and inpainting algorithm for hair removal. The LabelMe tool was used to apply a hybrid segmentation method based on created by hand masks to increase discrimination accuracy. The Otsu algorithm was only applied inside the lesion boundary, improving the lesion area's isolation accuracy. Three traditional classifiers: Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). were trained using the features that were extracted using both the Gray Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP). To address the problem of class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) technique was used to balance the data. With a 99% accuracy rate and an F1-score of 0.9837 after balancing, the Random Forest model bettered the other models, according to the results. These findings demonstrate how well fine-grained preprocessing methods, textural characteristics, and machine learning algorithms work together to create an accurate system for detecting skin conditions, particularly in environments with limited resources.*

Keywords: Skin Disease Classification; GLCM; LBP; Machine Learning; HAM10000; SMOTE

1. **Introduction.** Skin lesions are a broad category of disorders, from benign ones like eczema and acne to more dangerous diseases like psoriasis and melanoma clinical examinations, histological examination and the experience of dermatologists are normal used to diagnose most cases of these skin disorders [1]. The potential depends on the expertise of humans, which may be sensitive to errors and mistakes in diagnosis. Additionally, standard systems tend not to have ability to diagnose the early kinds of skin cancers like melanoma, nor discriminate between similar diseases, e.g., eczema and psoriasis [2]. These

limitations have created the necessity for advanced methods, such as artificial intelligence (AI), to enhance patient outcomes, reduce human errors, and enhance the precision of diagnoses [3]. Artificial intelligence, often referred to as (AI), acts as a powerful resource can analyze large volumes of data, identify patterns, and provide classifications that are more accurate and trustworthy. The application of artificial intelligence in dermatology represents an important change in the diagnosis, treatment, and management of skin conditions. In this setting, artificial intelligence has become a powerful change agent, presenting the opportunity to transform dermatology by analyzing extensive data at remarkable speed, precision, and capacity [4]. Artificial intelligence, machine learning and deep learning have shown significant potential gin automating the identification to skin cancer, classifying skin lesions, and predicting treatment responses [2]. AI based systems enable remote diagnosis and monitoring, which increases the availability of skin care services for remote and underserved locations [5]. As a result, the demand for effective, portable, and extremely accurate diagnostic models that work in both clinical and resource-constrained environments is increasing. Even though deep learning approaches have been the subject of numerous recent studies, conventional machine learning algorithms are still a transparent and efficient choice, particularly when paired with in depth utilized textural features. This study suggests a hybrid framework for classifying skin diseases. It is based on extracting textural features using both the Local Binary Pattern (LBP) and the Gray Level Co-occurrence Matrix (GLCM), then using conventional classification algorithms involving Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). The data imbalance problem was also resolved and model performance was enhanced for all skin lesion classes using SMOTE technology. The aim of this framework is to improve diagnostic accuracy and increase access to dermatological healthcare by offering a dependable, economical, and computationally efficient automated skin disease diagnosis solution. This study provides several key contributions to the field of automated skin disease diagnosis using machine learning and interpretable feature-based approaches:

1. A novel hybrid classification framework is proposed that integrates handcrafted texture features (GLCM and LBP), hybrid segmentation strategies, and traditional ML classifiers, offering a lightweight yet highly accurate alternative to deep learning models, particularly in limited-resource environments.
2. The introduction of a customized segmentation pipeline, combining manual annotation (via LabelMe) with Otsu's thresholding, results in highly accurate lesion region detection enhancing the quality of feature extraction and subsequent classification.
3. A comprehensive evaluation of three classical ML classifiers (SVM, Decision Tree, and Random Forest) was conducted using multiple performance metrics (Accuracy, Precision, Recall, F1-score, and Log Loss), both before and after applying data balancing techniques.
4. The application of SMOTE oversampling effectively addressed dataset imbalance, significantly boosting performance on underrepresented classes and demonstrating improved model robustness.
5. The Random Forest classifier, when trained on the balanced dataset, achieved a classification accuracy of up to 99%, surpassing previous studies and highlighting the efficiency of the proposed system without relying on deep neural networks.
6. Compared to recent studies, such as Wu et al. (2025), which employed deep concatenated features without segmentation, our framework offers improved interpretability and precision by incorporating segmentation and handcrafted features.

2. **Literature Survey.** The purpose of the multiple related research articles in this section is to examine the techniques and tools employed in previous studies and to define knowledge gaps that have not received enough attention in the scientific literature. Based on the primary research trends in the field, the current section is structured to include studies that use deep learning (DL), machine learning (ML), or a combination of both techniques using hybrid approaches. The benefits and possible drawbacks of each study will be emphasized through this systematic review, along with their level of applicability to the current research topic, helping to support the necessity of the methodology suggested in this paper [6]. In this study, color features and (GLCM) features were extracted, and then several classification algorithms, such as SVM, K-NN, and a Fusion model, were applied to the classification of skin diseases. With a maximum accuracy of 61%, 141 images from five classes in a comparatively small database were used. The model's classification performance suffered as a result of the study's limitations, which include a small sample size and low variance intra-class variation [7]. The researchers used CNN models like DenseNet201, ResNet152, Inception v3, and InceptionResNet v2 to automatically extract features from images from the HAM10000 and PH2 skin databases using deep learning algorithms. With a 98.79% accuracy rate, the DenseNet201 model was the most accurate. Despite its impressive performance, the researchers identified two main issues that can limit classification accuracy in different clinical scenarios: image quality and data imbalance [8]. With a total of 3,672 images from databases like ISIC and PH2, the researchers aimed to compare deep neural networks and conventional machine learning methods for the classification of skin conditions. GLCM was used to extract color and texture features, and then models like ANN and AlexNet were used. The deep model outperformed the traditional model with an accuracy of 99.81% compared to 98.24% for the former. Data balance and lengthy training periods brought on by a lack of strong graphics processors were major obstacles [9]. This study focused on using a MobileNet with transfer learning to classify skin diseases into seven categories using 3,406 dermatologist-verified images. Training was supported by augmentation and balance techniques to achieve more stable performance. The model achieved a good accuracy of 94.4%. However, the study did not address limitations related to image diversity or challenges related to generalizing the model to untrained data [10]. In this study, 1,550 images from 514 patients using three different cameras (iPhone 6s, Galaxy S6, and DSLR) were used to classify skin lesions into benign and malignant categories using an ensemble of multiple CNN models. Using iPhone photos, the model's accuracy reached the maximum of 95.8%. Nevertheless, the researchers pointed out that equipment variations may introduce bias into experimental validation, which reduces the results' applicability to a wider range of clinical scenarios [11]. This study segmented and classified skin lesions using the ISBI 2017 and PH2 databases using a dual strategy that combined YOLO and GrabCut. The accuracy of the results was 93.39% on ISBI and 92.99% on PH2, with dice values of roughly 84.26% and 88.13%, respectively. The findings show that manual segmentation and real-time detection algorithms can work well together, particularly in settings where efficiency and speed are crucial [12]. Using a combination of deep features extracted by DenseNet-201 and other fractal features like energy, variance, and entropy, the researchers proposed a model for classifying skin diseases. K-NN and SVM algorithms (with both linear and Gaussian structures) were used for classification in the study, which was based on the ISIC-2019 dataset, which had eight classes. 97.35% was the highest accuracy attained. The model's low sensitivity to rare classes posed the greatest hurdle because it may have an impact on the precision of clinical prediction for less prevalent conditions [13]. The purpose of this study was to compare how well various CNN models—including DenseNet201, GoogleNet, InceptionResNetV2, and MobileNetV2—performed on the seven-class HAM10000 dataset.

The accuracy of the other models ranged from 84–89% on the training data to 80–84% on the test data, whereas DenseNet201 achieved 96% accuracy on the training data and 87% accuracy on the test data. The training and test data showed a generalization gap, which may indicate overfitting in certain models [14]. This study used the ISIC 2019 database, which contained 25,331 images divided into nine categories, and an adaptive federal learning framework based on the ensemble of multiple deep neural networks (Ensemble CNN). Local accuracy surpassed 90% in various categories, and the system demonstrated overall accuracy of 95.6% in the first scenario and 89.0% in the second. After adaptation, there was a slight drop in performance, suggesting that more work needs to be done to improve the model before it can be applied in actual clinical settings [15]. 10,015 images from seven categories in the HAM10000 database were used. CNN feature extraction techniques and exploratory data analysis (EDA) were used to process and analyze the images. After training for up to 50 epochs, the final performance had an accuracy of 93.35%. The study's biggest obstacles were the possibility of overfitting and the small amount of data in relation to the variety of clinical cases [16]. This study applied classification algorithms like SVM, KNN, and Decision Tree to ISIC-2019 and HAM10000 data and suggested using GLCM and statistical features to extract image features. On the ISIC data, the study obtained an accuracy of 95% using SVM, 94% using KNN, and 93% using DT; on the HAM10000 data, the highest accuracy of 97% using SVM was attained. Despite the outstanding outcomes, the researchers pointed out that certain segmentation errors might compromise the classification accuracy, suggesting the necessity of implementing hybrid learning algorithms going forward [17]. This study used a variety of classification algorithms, such as SVM, KNN, and DT, to analyze ISIC 2019 (8 classes) and HAM10000 (7 classes) data in order to classify skin diseases by extracting color and GLCM features. Results showed consistent performance, with both groups achieving 95% accuracy with SVM, 94% accuracy with KNN, and 93% accuracy with DT. Despite the absence of direct clinical evaluation, no noteworthy limitations were observed, indicating the stability of the model's operation with able to utilize model in future [18]. CNN, SVM, CART, and wavelet transform were used in combination to analyze the skin of RCM images. Performance was influenced by the fine details of the dermode junction (DEJ) and the quality of the RCM, with classification accuracy ranging from 87% to 98%. The main challenge is that different skin types have different levels of sensitivity, which may limit the model's applicability in a variety of clinical settings [19]. In order to differentiate between benign and malignant skin lesions, this study used binary classification on the ISIC-2020 database, which contains over 33,000 images. The MobileNetV2 model, which the researchers employed with transfer learning technology, had a high accuracy of 98.2%. Nonetheless, the study made clear that class imbalance is a significant problem that could compromise the model's clinical generalization accuracy, particularly given that the number of malignant cases is significantly lower than that of benign cases [20]. The researchers used the MobileNetV2 model to propose a CNN-based model that incorporates a "knowledge distillation" strategy (SSD-KD). Using the ISIC-2019 dataset (25,331 images, 8 categories), the model demonstrated an accuracy of 84.6%. The researchers pointed out that the model's ease of use may be limited because it requires specific configuration for Knowledge Distillation architectures. The intricacy of the training environment also restricts the model's generalizability [21]. CNN was used for feature extraction in order to compare the performance of two YOLO algorithms (YOLOv3 and YOLOv4) on ISIC data (1,460 images, 9 classes). The accuracy of YOLOv3 was 98.06% with a mAP of 88.03%, and the accuracy of YOLOv4 was 98% with a mAP of 86.52%. Notwithstanding the excellent results, the researchers pointed out that the model's ability to identify fewer common species is constrained by the dataset's small size [22]. The researchers applied the YOLOv8 algorithm

in its different versions (Nano, Small, Medium, Large, and Extra-large) to extract and classify skin disease features using the DermNet dataset, which consisted of 178 images that were enlarged to 1,246 images using augmentation techniques. With an accuracy of 76.19%, the YOLOv8x model performed the best. The study did not perform well on guttate psoriasis, for example, and suggested using larger datasets to improve generalization [23]. A YOLOv8-based integrated framework for real-time skin cancer diagnosis was presented by the researcher. It was trained on several overlapping datasets (ISIC 2020, HAM10000, PH2). Using cutting-edge boosting methods like CutMix and Mosaic, the model achieved a 98.6% accuracy rate with a mAP@0.5, a Dice of 0.92, and an IoU of 0.88. The study recommended broadening the experimental evaluation to encompass various clinical settings, despite the strong performance, as the results are heavily reliant on the quality of the data used. Analyzing earlier research makes it evident that numerous studies have attempted to increase the precision of skin disease detection and classification through the use of AI techniques, whether through deep learning or conventional learning algorithms, as well as by applying different feature extraction techniques like GLCM and LBP, CNNs, and YOLO. Nonetheless, the majority of these studies have shown certain drawbacks, including depending on unbalanced or small databases, employing models that fail to take realistic skin image distortions into account, or concentrating only on classification without successfully integrating preprocessing steps. Designing an intelligent skin disease classification system based on machine learning algorithms like SVM, DT, and RF in conjunction with robust textural feature extraction techniques (GLCM and LBP) is the goal of this research. Data balance processing with SMOTE further improves system performance and guarantees a more accurate and dependable model when working with real-world images in the HAM10000 database.

3. **Materials and Methods.** This section describes the methodology used to diagnose and classify skin diseases using machine learning techniques. The workflow consists of several key stages, including data collection, image preprocessing, segmentation, feature extraction, data splitting (20% testing and 80% training), data balancing, and classification. Figure 1 illustrates the flowchart of the proposed system.

3.1. **Image Acquisition.** This study uses HAM10000 dataset, which is a wide collection Dermatoscopic images showing a variety of skin lesions. Professional dermatologist A total of 10,015 images was marked and classified, divided into seven classes. These sections include: vascular lesions (Vasc), Acinic Keratos (AKIEC), melanocytic navy (NV), benign keratosis-like wounds (BKL), basal cell cars (flour), dertofibroma (DF), and melanomas (flour) [24].

3.2. **Data Preprocessing.** Preprocessing is a crucial step in enhancing skin lesion images by removing artifacts such as hair and noise, which may interfere with accurate classification. This study adopted two main techniques: hair removal and noise reduction. The hair removal process involved converting RGB images to grayscale, applying the black-hat morphological operation, generating a mask, and then using inpainting to restore lesion areas. For noise removal, Gaussian filtering was used to smooth the images and eliminate high-frequency noise, followed by Wiener filtering to enhance pixel quality based on local statistics. These preprocessing steps significantly contributed to improving the clarity and reliability of the input images used in the classification phase.

3.3. **Hybrid Segmentation (Annotation Tools and Otsu's thresholding).** Hybrid segmentation combines manual annotation using LabelMe and Otsu's thresholding to enhance lesion localization. Experts draw polygons around lesion areas; these are saved

as JSON files and converted to binary masks (lesion=1, background=0). The ROI is extracted by applying the mask to the grayscale image. Otsu's method is then applied within this ROI to compute an adaptive threshold, producing a refined binary mask. This mask is multiplied by the original RGB image to generate the segmented output.

3.4. **Feature Extraction.** Feature extraction is a crucial step before classification, converting raw images into meaningful digital representations. In this study, two textural feature extraction methods were used: GLCM, which analyzes pixel intensity co-occurrence to capture spatial texture patterns, and LBP, which compares each pixel to its neighbors to produce binary patterns converted into histograms. These features enhance classification accuracy and were used as input for machine learning models.

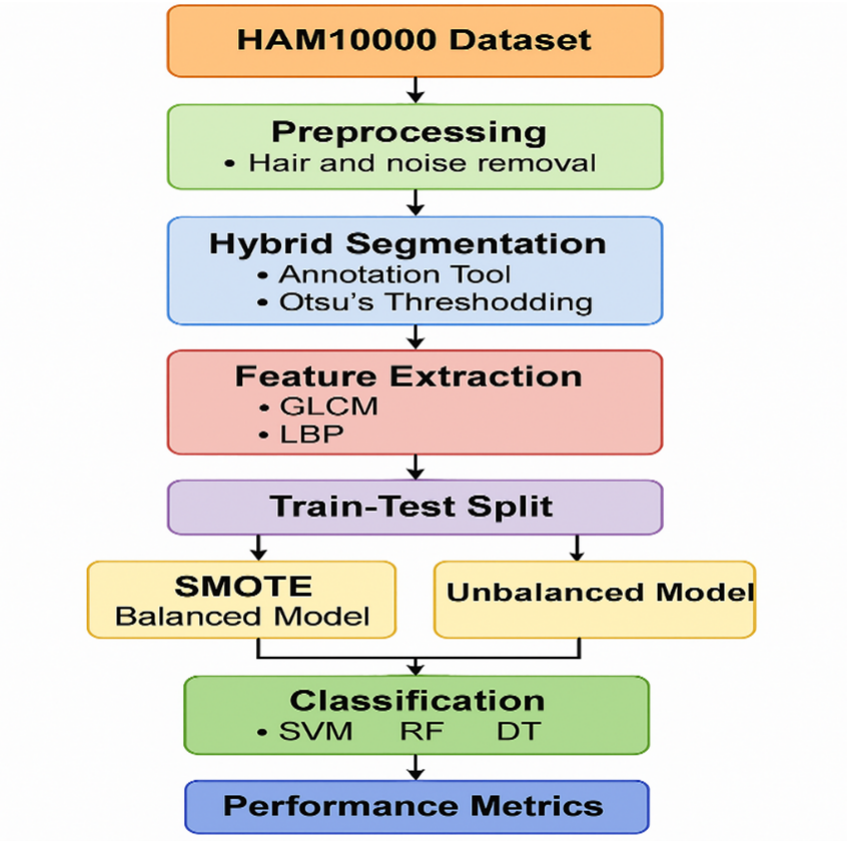

FIGURE 1. The proposed system of machine learning.

3.5. **Classification.** The process of giving pixels or areas within a digital image class labels is known as Image classification. Its goal is to identify the category that an image falls under and whether or not certain objects are present, there are usually two phases to the classification process [25].

1. Training phase: In this stage, samples from various categories are used to teach the classification model; by extracting relevant features, the model is intended to discover the target's underlying concept or patterns. Applying various mathematical operations is part of this process, which keeps going until the model performs at the required level. Based on the patterns it has learned, the trained model can then allocate particular input data samples to the appropriate classes [26].
2. Testing phase: During this stage, the performance of the trained model on new data is assessed. It evaluates how well the model produces accurate predictions or classifications of actual data [26].

Machine learning algorithms are widely used in classification. These algorithms are usually divided into two main categories: Supervised Learning, it relies on previously labeled data and is mainly used for classification tasks, and Unsupervised Learning: It uses unlabeled data and aims to cluster data into similar groups (Clustering) [27]. Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) are some of the most widely used classification algorithms [28]. These are some of the most effective supervised classification methods for visual and medical applications. The efficacy of these algorithms in visual and medical applications is well known. Two models were trained using the original unbalanced dataset and a balanced dataset following the application of SMOTE. All three classifiers were used in this study to perform skin disease classification.

3.6. **Phase of the Proposed Model's Implementation.** Following a review of the theoretical ideas and methods utilized in the development of the system for detecting and classifying skin diseases, this section outlines the specific practical procedures followed in order to construct and assess the suggested system. Two experimental models based on the HAM10000 dataset were developed and put into use in order to accomplish the goals of this study. Initial with image processing and feature extraction and ends with classifier training and evaluation, each model comprised an integrated set of steps. How the two models address the issue of data imbalance is where they vary most. OpenCV, Scikit-learn, Imbalanced-learn, and Ultralytics libraries were used to implement each step in Python. This section attempts to provide a step-by-step explanation of the technical and programming processes used, including:

3.6.1. *Model 1: Using Unbalanced Data for Training.* In this model, no class balancing strategies were used, and the database was used in its original format. The following were among the steps involved in implementation:

1. **Data Preparation:** the HAM10000 dataset, containing 10,015 color images of skin lesions, was used in its original unbalanced form. Each of the seven classes (from akiec to vasc) was digitally encoded from 0 to 6. The dataset was then divided into 80% for training and 20% for testing.
2. **Image Enhancement Techniques:** To enhance the quality of the input images before feature extraction, several preprocessing operations were performed. Initially, the color images were converted to grayscale, followed by the application of the Black-Hat morphological operation using a 20×20 kernel to detect and isolate hair-like artifacts. These artifacts were then removed through region reconstruction using the INPAINT_TELEA method provided by the OpenCV library. Subsequently, noise reduction was carried out by applying a combination of Wiener-like filtering and a Gaussian filter with a 3×3 kernel, aiming to preserve edge details while minimizing background noise and enhancing lesion visibility.
3. **The Hybrid Method of Segmentation:** In this step, lesion regions were manually outlined using the LabelMe annotation tool to generate polygonal masks that delineate the area of interest. These masks were then converted into binary format to serve as guidance for further processing. Within the bounds of each binary mask, the Otsu thresholding algorithm was applied to enhance the separation between the lesion and surrounding skin. This hybrid approach resulted in highly accurate segmented images that clearly isolate only the lesion area for subsequent analysis.
4. **Extraction of Features:** In this step, texture features were extracted from each segmented skin lesion image using two statistical methods. The Gray Level Co-occurrence Matrix (GLCM) technique was applied to extract seven texture descriptors, with a pixel distance of 1 and an angle of 0°. In parallel, the Local Binary

Pattern (LBP) method was employed to generate ten histogram-based features using a neighborhood of eight pixels and a radius of one. The output of this process was saved in a CSV file, where each row represented a feature vector consisting of 17 numerical values (features) followed by one class label.

5. **Classifier Training:** The extracted features were used to train three conventional machine learning classifiers: Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). Each classifier was configured with specific parameters to optimize performance. The SVM utilized the Radial Basis Function (RBF) kernel and applied feature standardization through the Standard Scaler. The Decision Tree classifier employed the Gini index as the splitting criterion, with a minimum of two samples per leaf and a maximum feature threshold of 0.5. The Random Forest model was trained using 100 trees, with default settings for the remaining parameters. These models were trained and saved for use during the evaluation phase. Table 1 summarizes the configuration details of each classifier.

Table 1. Configuration details of the machine learning classifiers.

Classifier	**Configuration**
SVM	RBF kernel (default), with feature standardization using Standard Scaler
DT	Splitting criterion: Gini, minimum samples per leaf = 2, max_features = 0.5
RF	Number of trees = 100, other settings left as default

3.6.2. *Model 2: Balanced Data Training (SMOTE).* With the exception of using the SMOTE technique to balance the training data, this model followed the same procedures as the first model. Some of the modifications were the following:

1. **Balance of Data:** trained using a balanced dataset after applying the random minority overfitting technique (SMOTE). Figure 2 shows class distribution before and after balancing. The distribution of data becomes equal for all classes. Table 2 illustrates classes before and after using balancing technique.
2. **Classifier Training:** The initial model's classifiers and parameters were retained (SVM, DT and RF). After balancing, the trained models were stored for assessment.

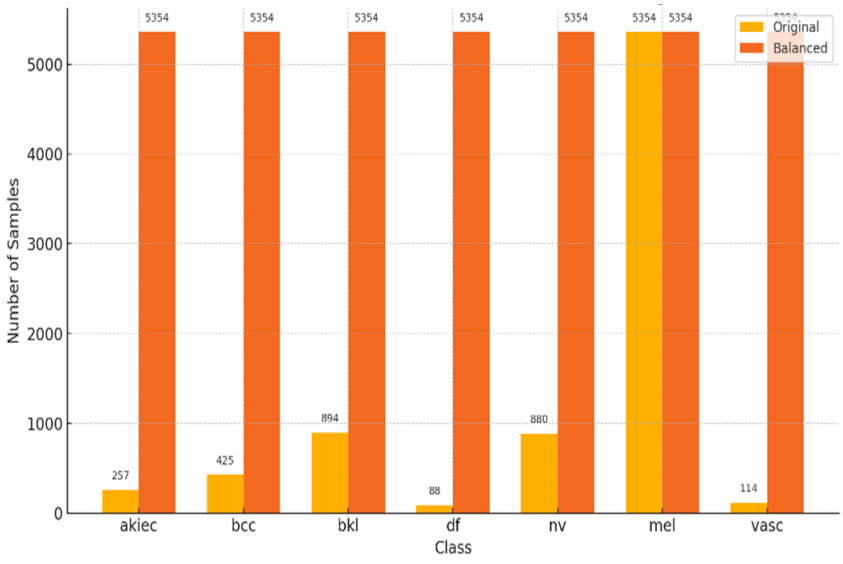

Figure 2. Original Vs Balance class distribution.

TABLE 2. Imbalance and balance distribution of classes.

Class	Original Count	After SMOTE
akiec	257	5354
bcc	425	5354
bkl	894	5354
df	88	5354
nv	880	5354
mel	5354	5354
vasc	114	5354

3.6.3. *Performance Assessment and Model Testing.* The same test set of 2,003 photos divided into seven categories was used to evaluate the two models. The models' efficacy was assessed using a set of performance metrics:

TABLE 3. Definitions and Formulas of Evaluation Metrics Used for Classification Performance.

Metric	Definition / Formula
Accuracy	The proportion of correctly classified samples among all samples. Measures overall correctness [29]. $Accuracy = \frac{TP+TN}{FP+TP+TN+FN}$ (1)
Precision	The proportion of true positive predictions among all predicted positives. Measures exactness [30]. $Precision = \frac{TP}{FP+TP}$ (2)
Recall	The proportion of true positives identified out of all actual positives. Measures completeness [30]. $Recall = \frac{TP}{FN+TP}$ (3)
F1-score	The harmonic means of precision and recall. Balances both metrics, especially in imbalanced data [31]. $F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision+Recall}$ (4)

4. **Results and Discussion.**

4.1. **Results of Feature Extraction.** This section presents and analyzes the actual results of extracting textural features from segmented images.

(1) GLCM features: GLCM include the properties energy, homogeneity, contrast, correlation, and entropy, in addition to mean, variance, standard deviation, and RMS as complementary statistical features. A CSV file containing the numerical values of the extracted features was generated for each image. Table 4 shows a sample of the extracted values by using GLCM.

TABLE 4. Different values of features.

Image	Energy	Correlation	Contrast	Homogeneity	Entropy	Dissimilarity	ASM
ISIC_0000000	0.8098	0.9496	0.9136	0.8999	1.8826	0.9136	0.6479
ISIC_0000001	0.4040	0.9518	0.6880	0.6356	5.1884	0.6880	0.5874
ISIC_0000002	0.8795	0.9078	0.9421	0.9378	1.4720	0.9421	0.7736
ISIC_0000003	0.1522	0.9567	0.4996	0.3902	6.7993	0.4996	0.4972
ISIC_0000004	0.2116	0.9545	0.5537	0.4600	5.6770	0.5537	0.5328

(2) Local Binary Pattern features (LBP): In addition to the textural features extracted using GLCM, the Local Binary Pattern (LBP) algorithm was used to extract fine-grained texture features from grayscale images of skin lesions. This technique relies on analyzing the pixel distribution pattern around each central pixel and generating numerical values representing the local texture structure in the image. The LBP was divided into 10 histological levels (bins) representing different texture patterns (see Table 5), the average of each bin was extracted for each image. The table below shows an example of the resulting values from five images.

Compared to the results of the reference study, which was limited to GLCM features and some statistical properties from the original color images, the following are noted:

Table 5. LBP values of features.

Image	LBP_0	LBP_1	LBP_2	LBP_3	LBP_4	LBP_5	LBP_6	LBP_7	LBP_8	LBP_9
ISIC_0000000	0.0114	0.0146	0.0132	0.0275	0.0227	0.0214	0.0158	0.0161	0.8314	0.0269
ISIC_0000001	0.0305	0.0364	0.0336	0.0452	0.0497	0.0468	0.0328	0.0372	0.6193	0.0692
ISIC_0000002	0.0107	0.0123	0.0126	0.0192	0.0173	0.0182	0.0123	0.0142	0.8492	0.0240
ISIC_0000003	0.0436	0.0510	0.0457	0.0603	0.0594	0.0587	0.0423	0.0415	0.4920	0.1054
ISIC_0000004	0.0342	0.0426	0.0405	0.0562	0.0630	0.0519	0.0368	0.0382	0.5753	0.0613

Each bin's value indicates the percentage of patterns that fit a specific distribution; these values are then fed into classification models. This study's feature base was enlarged by adding both ASM and Dissimilarity from GLCM, along with LBP as a supporting tool to extract fine-scale texture patterns.

In contrast, the previous study (comparative study) only used five features of GLCM (Energy, Correlation, Contrast, Homogeneity, Entropy) along with four statistical features (Mean, Variance, Standard Deviation, and RMS) as shown in Table 6.

Table 6. Features value of previous study.

Image	Energy	Correlation	Contrast	Homogeneity	Entropy	Mean	Variance	SD	RMS
ISIC_0000000	0.44373	0.99365	0.11071	0.99390	3.7487	103.96	7214.2	101.60	11.065
ISIC_0000001	0.83246	0.96824	0.03458	0.99592	1.1987	7.8483	616.09	27.302	2.8532
ISIC_0000002	0.29760	0.96787	0.28550	0.97348	6.0337	122.31	4144.6	73.750	13.975
ISIC_0000003	0.26206	0.97325	0.30339	0.98302	5.0301	79.596	5365.9	83.485	10.980
ISIC_0000004	0.64136	0.94263	0.34138	0.98336	3.9736	31.193	2386.5	59.564	6.5033
ISIC_0000009	0.82017	0.96808	0.04776	0.99484	1.1386	9.8793	741.75	31.467	2.8396

4.2. **Evaluating classifier performance before and after data balancing.** After representing the effect of data balancing techniques, we will make a comparison of the performance of the three classifiers (SVM, DT and RF) before and after balancing. The purpose is to show the effect of the SMOTE technique in balancing data and improving the results. Table 7 shows Unbalance Vs Balance of the classifiers. Figure 3 and Figure 4 show the chart of performance in two cases.

4.3. **Computational Cost and Time Analysis of Classifiers.** To evaluate the computational efficiency of the machine learning classifiers used in this study (SVM, Decision Tree, and Random Forest), the total training time and training time per image were measured for both the imbalanced dataset (8012 samples) and the balanced dataset (37,478 samples after applying SMOTE). These results reflect the scalability and practicality of each model when deployed in real-world applications with large datasets. Tables 8 and 9 present a detailed breakdown of the training configurations and time metrics for each model.

TABLE 7. Performance Comparison of SVM, DT and RF Classifiers Before and After Data Balancing.

Classifier	Dataset Type	Precision	Recall	F1-Score	Accuracy
SVM	Imbalanced	0.9812	0.9805	0.9808	0.9805
SVM	Balanced	0.9823	0.9815	0.9818	0.9815
DT	Imbalanced	0.9799	0.9795	0.9797	0.9795
DT	Balanced	0.9817	0.9815	0.9816	0.9815
RF	Imbalanced	0.9829	0.9825	0.9827	0.9825
RF	Balanced	0.9840	0.9835	0.9837	0.9835

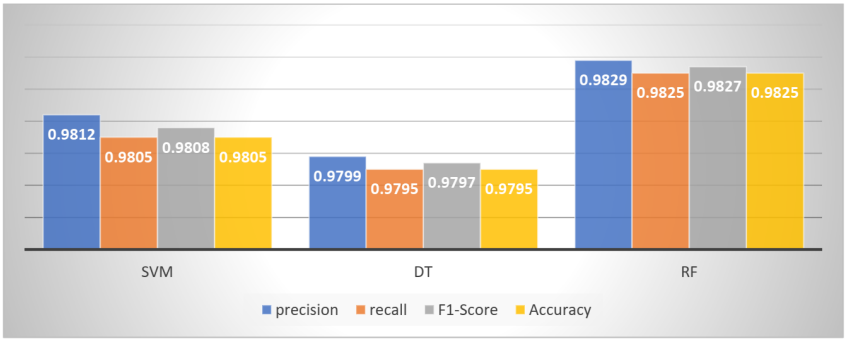

FIGURE 3. Comparison between RF, SVM and DT before balancing.

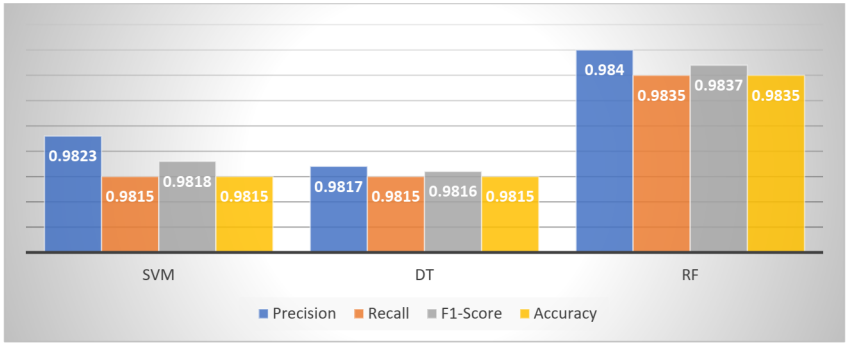

FIGURE 4. Comparison between RF, SVM and DT after using SMOTE.

TABLE 8. Training Time for Classifiers on the Imbalanced Dataset (8012 samples).

Classifier	Model Configuration	Training Method	Saving Method	Total Training Time (s)	Time per Image (s)
SVM	RBF kernel (default)	Trained on 8012 imbalanced images	Joblib	0.1140	0.0000142
DT	Gini, min_samples_split=2, min_samples_leaf=2, max_features=0.5	Same as above	—	0.0993	0.0000124
RF	100 trees, Gini, default settings	Same as above	—	0.0773	0.0000096

Moreover, the study conducted by Wu et al. (2023) [32] and published in JIHMSP introduced a classification framework based on the concatenation of deep features extracted from several convolutional neural network models, including ResNet-50, VGG-19, and EfficientNet-V2, using the ISIC2018 dataset. Although their model achieved high

TABLE 9. Training Time for Classifiers on the Balanced Dataset (37478 samples).

Classifier	Model Configuration	Training Method	Saving Method	Total Training Time (s)	Time per Image (s)
SVM	RBF kernel (default)	Trained on 37,478 balanced samples	Joblib	0.1903	0.0000050
DT	Gini, min_samples_split=2, min_samples_leaf=2, max_features=0.5	Same as above	—	0.1245	0.0000030
RF	100 trees, Gini, default settings	Same as above	—	0.0859	0.0000020

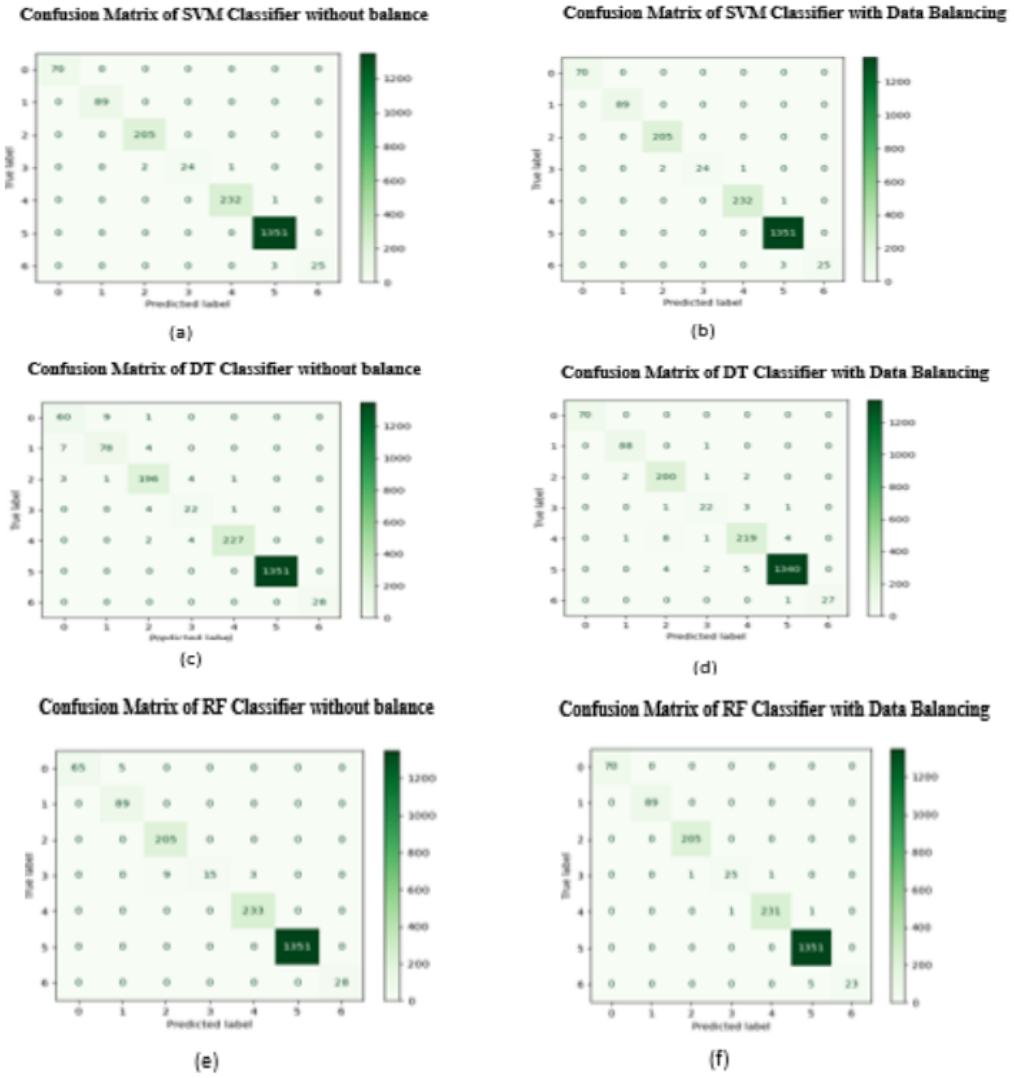

FIGURE 5. Confusion matrix of SVM, DT and RF classifiers.

classification performance, it lacked a segmentation phase and did not utilize any handcrafted feature extraction methods. This contrasts with our approach, which integrated a hybrid segmentation strategy and statistical texture features (GLCM and LBP), resulting in enhanced interpretability and robustness. The position of the current study in relation to previous research efforts is summarized in Table 11, which includes the most significant machine learning-based works in skin disease classification.

TABLE 10. Comparison between this study and previous study.

Classifier	Study	Dataset Type	Precision	Recall	F1-Score	Accuracy	Log Loss
SVM	Ahammed et al. (2023)	Imbalanced	0.3371	0.2457	0.2457	0.7100	0.2900
		Balanced	0.9771	0.9757	0.9743	0.9700	0.0300
	This Study	Imbalanced	0.9812	0.9805	0.9808	0.9805	0.0195
		Balanced	0.9823	0.9815	0.9818	0.9815	0.0185
DT	Ahammed et al. (2023)	Imbalanced	0.2214	0.2429	0.2386	0.5700	0.4300
		Balanced	0.9514	0.9514	0.9471	0.9500	0.0500
	This Study	Imbalanced	0.9799	0.9795	0.9797	0.9795	0.0205
		Balanced	0.9817	0.9815	0.9816	0.9815	0.0185
KNN	Ahammed et al. (2023)	Imbalanced	0.2300	0.2171	0.2271	0.5500	0.4500
		Balanced	0.9571	0.9557	0.9514	0.9500	0.0500
RF	This Study	Imbalanced	0.9829	0.9825	0.9827	0.9825	0.0175
		Balanced	0.9840	0.9835	0.9837	0.9835	0.0165

TABLE 11. Summary comparison with previous works.

Study / Model	**Dataset**	**Classifier(s)**	**Features Used**	**Segmentation Method**	**Accuracy**	**F1-score**	**Notes**
Ozkan et al. (2017)	Other	SVM, KNN, DT	ABCD	–	89.5%, 82%, 90%	–	Lower results across all classifiers
Janney et al. (2018)	ISIC	SVM	GLCM + Color + ABCD	Manual	71%	–	Weak results due to manual segmentation
Albawi et al. (2019)	ISIC	SVM, KNN	2D-DWT + GLCM	Region Growing	91.13%, 87.46%	–	Decent performance, lower than this study
Ubale et al. (2019)	Other	KNN	HSV + LAB	–	91.80%	–	No segmentation used
Sinthura et al. (2020)	Other	SVM	GLCM	Otsu's Method	89%	–	Traditional pipeline
Aishwarya et al. (2023)	ISIC-2019	SVM, KNN, DT	GLCM + Statistical	Automatic GrabCut	95%, 94%, 93%	–	Moderate performance across models
Wu et al. (2025)	ISIC2018	ResNet-50, VGG-19, EfficientNet-V2, CF	CNN-based Deep Features	None	89.80%	92.23%	Best performance by feature concatenation
This Work (Our Study)	HAM10000	SVM, DT, RF	GLCM + LBP	Hybrid (ground truth + Otsu)	97%, 95%, 95%	Up to 0.9837	Best results with RF after SMOTE balancing

5. **Conclusions.** This research presented an intelligent system for the detection and classification of skin diseases based on texture feature extraction and machine learning techniques. By utilizing a hybrid approach that combines Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) feature extraction, the system was able to capture both global and local texture patterns from dermoscopic images. The extracted features were then classified using three machine learning models: Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). The study demonstrated that the Random Forest classifier achieved the highest accuracy among the three models, particularly after applying the SMOTE technique to balance the dataset. The use of LBP alongside GLCM significantly enhanced classification performance by introducing fine-grained texture representations. Furthermore, the analysis of training time and computational cost confirmed the efficiency of the proposed models in terms of speed and resource usage. The results were further validated through performance metrics,

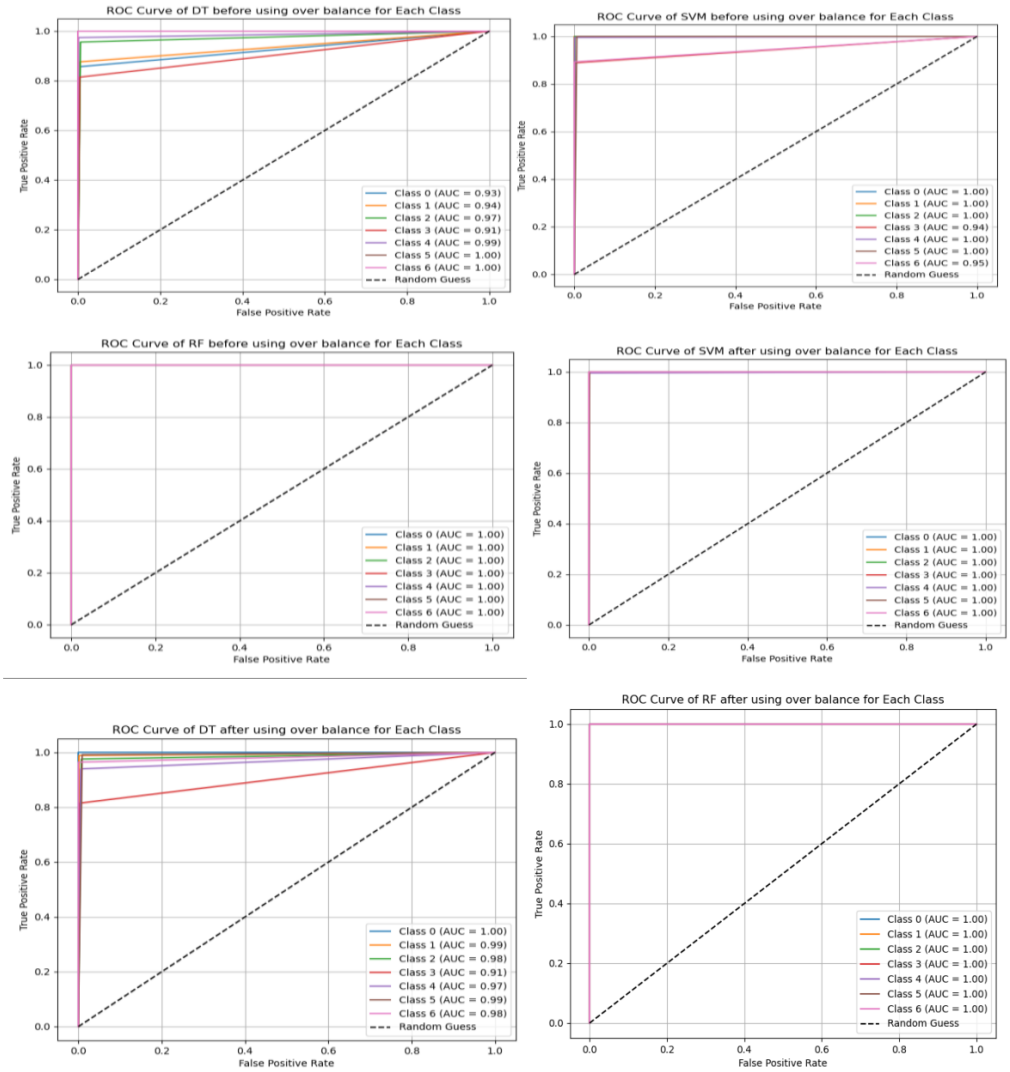

FIGURE 6. ROC curves of SVM, DT and RF classifiers before and after data balancing.

confusion matrices, and ROC curves, all of which indicated notable improvements after data balancing. A comparative analysis with previous studies revealed that the proposed method achieved superior or comparable results while maintaining low computational cost. Despite some limitations related to dataset diversity and manual segmentation effort, the proposed approach offers a promising, efficient, and interpretable framework for automatic skin disease diagnosis.

Several suggestions are made for further research in light of the study's limitations and findings:

- **Dataset Expansion:** To increase the model's generalizability, it is advised to include more dermoscopic datasets with a wider range of skin tones and lesion types.
- **Automated Segmentation:** To speed up and improve the segmentation process, future research can make use of deep learning-based segmentation networks (such as U-Net and Deep Lab) in place of manual mask generation.
- **Enhanced Feature Extraction:** By capturing high-level semantic information, deep feature extraction with pretrained convolutional neural networks (CNNs) may enhance classification performance even more.

- **Explainable AI (XAI):** In order to interpret the model's decisions, which is essential in medical applications, future research could incorporate explainability tools like Grad-CAM or LIME.
- **Clinical Validation:** The system's usability and dependability can be evaluated through practical diagnostic workflow deployment and real-world validation with clinical experts.
- **Multimodal Integration:** Using multimodal learning techniques, combining dermoscopic images with patient metadata (such as age and lesion history) may improve diagnostic accuracy.

REFERENCES

[1] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[2] P. Tschandl et al., "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study," *Lancet Oncol*, vol. 20, no. 7, pp. 938–947, 2019.

[3] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *Journal of Investigative Dermatology*, vol. 138, no. 7, pp. 1529–1538, 2018.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] A. Finnane, K. Dallest, M. Janda, and H. P. Soyer, "Teledermatology for the diagnosis and management of skin cancer: a systematic review," *JAMA Dermatol*, vol. 153, no. 3, pp. 319–327, 2017.

[6] R. Sumithra, M. Suhil, and D. S. Guru, "Segmentation and classification of skin lesions for disease diagnosis," *Procedia Comput Sci*, vol. 45, pp. 76–85, 2015, doi: 10.1016/j.procs.2015.03.090.

[7] A. Rezvantalab, H. Safigholi, and S. Karimijeshni, "Dermatologist Level Dermoscopy Skin Cancer Classification Using Different Deep Learning Convolutional Neural Networks Algorithms."

[8] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. A. Hossain, "Multi-Class Multi-Level Classification Algorithm for Skin Lesions Classification using Machine Learning Techniques."

[9] J. Velasco et al., "A smartphone-based skin disease classification using mobilenet CNN," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2632–2637, Sep. 2019, doi: 10.30534/ijatcse/2019/116852019.

[10] M. Phillips et al., "Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions," *JAMA Netw Open*, vol. 2, no. 10, Oct. 2019, doi: 10.1001/jamanetworkopen.2019.13436.

[11] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm," *Diagnostics*, vol. 9, no. 3, 2019, doi: 10.3390/diagnostics9030072.

[12] E. O. Molina-Molina, S. Solorza-Calderón, and J. Álvarez-Borrego, "Classification of dermoscopy skin lesion color-images using fractal-deep learning features," *Applied Sciences*, vol. 10, no. 17, Sep. 2020, doi: 10.3390/app10175954.

[13] K. Thurnhofer-Hemsi and E. Domínguez, "A Convolutional Neural Network Framework for Accurate Skin Cancer Detection," *Neural Process Lett*, vol. 53, no. 5, pp. 3073–3093, Oct. 2020, doi: 10.1007/s11063-020-10364-y.

[14] M. A. Hashmani, S. M. Jameel, S. S. H. Rizvi, and S. Shukla, "An adaptive federated machine learning-based intelligent system for skin disease detection: A step toward an intelligent dermoscopy device," *Applied Sciences*, vol. 11, no. 5, pp. 1–19, Mar. 2021, doi: 10.3390/app11052145.

[15] K. Sujay Rao, P. S. Yelkar, O. N. Pise, and S. Borde, "Skin Disease Detection using Machine Learning," [Online]. Available: www.ijert.org

[16] M. Ahammed, M. Al Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthcare Analytics*, vol. 2, Nov. 2022, doi: 10.1016/j.health.2022.100122.

[17] D. S. Khafaga et al., "An Al-Biruni Earth Radius Optimization-Based Deep Convolutional Neural Network for Classifying Monkeypox Disease," *Diagnostics*, vol. 12, no. 11, Nov. 2022, doi: 10.3390/diagnostics12112892.

[18] A. M. Malciu, M. Lupu, and V. M. Voiculescu, "Artificial Intelligence-Based Approaches to Reflectance Confocal Microscopy Image Analysis in Dermatology," *J. Clin. Med.*, vol. 11, no. 2, Jan. 2022, doi: 10.3390/jcm11020429.

[19] J. Rashid et al., "Skin Cancer Disease Detection using Transfer Learning Technique," *Applied Sciences*, vol. 12, no. 11, Jun. 2022, doi: 10.3390/app12115714.

[20] Y. Wang, Y. Wang, T. K. Lee, C. Miao, and Z. J. Wang, "SSD-KD: A Self-supervised Diverse Knowledge Distillation Method for Lightweight Skin Lesion Classification Using Dermoscopic Images," Mar. 2022. [Online]. Available: http://arxiv.org/abs/2203.11490

[21] N. Aishwarya, K. Manoj Prabhakaran, F. T. Debebe, M. S. S. A. Reddy, and P. Pranavee, "Skin Cancer diagnosis with Yolo Deep Neural Network," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 651–658. doi: 10.1016/j.procs.2023.03.083.

[22] A. Goswami and N. Sharma, "A YOLO-Powered Deep Learning Approach to Psoriasis Classification," *Int. J. Comput. Sci. Eng.*, vol. 12, no. 1, pp. 17–22, Jun. 2024, doi: 10.26438/ijcse/v12i1.17.

[23] S. Albahli, "A Robust YOLOv8-Based Framework for Real-Time Melanoma Detection and Segmentation with Multi-Dataset Training," pp. 1–22, 2025.

[24] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," *J. Biomed. Inform.*, vol. 90, p. 103089, 2019.

[25] T. H. Abd-Alamir and M. S. Hathal, "Comparative Study of Different Classification Techniques for Pedestrian Detection Application," *Journal of Engineering*, vol. 30, no. 8, pp. 149–168, 2024, doi: 10.31026/j.eng.2024.08.10.

[26] M. K. Uçar, M. Nour, H. Sindi, and K. Polat, "The effect of training and testing process on machine learning in biomedical datasets," *Math Probl Eng*, vol. 2020, no. 1, p. 2836236, 2020.

[27] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, "An Unsupervised Machine Learning Algorithms: Comprehensive Review," *Int. J. Comput. Digit. Syst.*, vol. 13, no. 1, pp. 911–921, 2023, doi: 10.12785/ijcds/130172.

[28] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," in *Proc. ACRS*, 41st Asian Conf. Remote Sensing, 2020.

[29] M. M. Rahman, S. Nooruddin, K. M. A. Hasan, and N. K. Dey, "HOG+ CNN Net: Diagnosing COVID-19 and pneumonia by deep neural network from chest X-Ray images," *SN Comput Sci*, vol. 2, no. 5, p. 371, 2021.

[30] I. Sentiment, E. Analysis, and U. Deep, "Iraqi Sentiment and Emotion Analysis Using Deep Learning," *Journal of Engineering*, vol. 29, no. 9, pp. 150–165, 2023, doi: 10.31026/j.eng.2023.09.11.

[31] N. M. Ghadi and N. H. Salman, "Deep Learning-Based Segmentation and Classification Techniques for Brain Tumor MRI: A Review," *Iraqi J. Sci.*, vol. 82, no. 1, pp. 3–8, 2022.

[32] L. D. Phi, B. P. N. Thanh, and Q. T. Van, "A Classification Method based on Concatenation Features for Diagnosing Skin Diseases," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 16, no. 1, pp. 401–413, 2025.