Ingénierie des Systèmes d'Information
Vol. 31, No. 10, October, 2026, pp. 3481-3495
Journal homepage: http://iieta.org/journals/isi

Deep Learning for Vitiligo Image Analysis: A Systematic Review of Datasets, Segmentation, Classification, and Data Augmentation

Priyanka Pawar*, Rais Abdul Hamid Khan

School of Computer Science & Engineering, Sandip University, Nashik 422213, India

Corresponding Author Email: priyapawar2009@gmail.com

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/isi.311020

Received: 16 June 2026
Revised: 27 August 2026
Accepted: 3 September 2026
Available online: 31 October 2026

Keywords:
vitiligo, deep learning, lesion segmentation, image classification, data augmentation, dermatological datasets, systematic review

ABSTRACT

Vitiligo image analysis using deep learning (DL) has been investigated for lesion segmentation and image-based classification, but progress is constrained by limited datasets and inconsistent evaluation practices. This systematic review synthesizes vitiligo-specific research and related methodological studies on datasets, segmentation, classification, data augmentation, and performance metrics. Following a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-based selection process, searches were conducted in IEEE Xplore, PubMed, Scopus, and Web of Science, supplemented by additional sources and reference-list screening. The screening process identified 286 records and retained 111 studies. The search primarily covered 2015–2025, with earlier foundational studies included where relevant. The reviewed literature indicates a shift from thresholding, region-growing, and clustering methods toward convolutional neural networks (CNNs) and newer semi-supervised, weakly supervised, ensemble, and hybrid architectures. Transfer learning with pretrained networks, including VGG, ResNet, DenseNet, MobileNet, and EfficientNet, is widely reported for image classification. Common augmentation strategies include geometric transformations, photometric adjustments, MixUp, and generative models. Resources discussed include Vitiligo-SI, datasets hosted on Kaggle and Mendeley, DermNet materials, and the PUMCH inflammatory skin disease dataset. However, differences in image sources, acquisition conditions, class balance, annotation practices, skin-tone representation, and validation protocols limit direct comparison across studies. Reported performance should therefore be interpreted in the context of each study's dataset and evaluation design rather than as a common ranking of methods. Future work should prioritize well-documented, clinically representative datasets, standardized annotation and evaluation protocols, and external multicenter validation. Automated image analysis should be assessed as decision support and not treated as a substitute for clinical diagnosis.

1. INTRODUCTION

Vitiligo affects approximately 0.5–2% of the global population [1]. It is a chronic skin condition characterized by selective melanocyte destruction and associated psychosocial effects [2]. It is characterized by the selective destruction of melanocytes, resulting in depigmented macules or patches. The disorder has a complex etiology involving environmental factors, oxidative stress, autoimmune responses, and genetic predisposition. Based on the distribution and progression of lesions, vitiligo is clinically divided into multiple types [3]. Non-segmental vitiligo (NSV), the most common type, is characterized by symmetrical white patches and may occur in generalized or universal forms [4]. Unilateral depigmentation restricted to a dermatome is the hallmark of segmental vitiligo (SV) [4, 5]. SV often stabilizes after rapid initial progression. Localized patches are a feature of focal vitiligo. This variation makes automated image classification challenging and highlights the need for accurate classification to support effective monitoring and treatment planning [6].

According to a recent systematic review, the lifetime prevalence of vitiligo is approximately 0.36% (95% credible interval (CrI) 0.24–0.54), or about 28.5 million people worldwide. As per the same study, the prevalence was approximately 0.67% among adults (approximately 37.1 million adults) and 0.24% among children (approximately 5.8 million) in 2022 [7]. Atopic dermatitis was reported in 12.5% of patients with vitiligo and 8.4% of the matched general population [8].

Conventional diagnosis depends mainly on subjective visual assessment that might differ from one dermatologist to another. To improve diagnostic consistency, researchers have increasingly adopted deep learning (DL) and artificial intelligence (AI) techniques. These include convolutional neural networks (CNNs) [9] and transfer learning models (TLMs) [10]. Lightweight architectures [11] and transformer-based models [12] have also been explored for efficient and robust diagnosis. Segmentation networks [13, 14] have been employed to improve lesion localization, while generative adversarial network (GAN)-based models [15] have supported

data augmentation and image generation. Explainable artificial intelligence (XAI) techniques [16, 17] have further been incorporated to enhance the interpretability and transparency of diagnostic predictions. This change entails the development of specialized datasets, such as Vitiligo-SI, DermNet NZ, etc. [18]. It includes clinical and dermoscopic images of various skin types; however, issues like data imbalance and inconsistent conditions still exist.

Alavi and Kargari [19] and Shobana and Deny [20] proposed contrast enhancement and artifact removal preprocessing methods. These are essential for maximizing image quality. Data augmentation techniques, such as rotation and brightness adjustment, help reduce overfitting. GANs further enhance dataset diversity by generating synthetic images. Yu et al. [21] focused on a technique to resize photos to 224 × 224 pixels for CNNs. It is an example of how image resizing guarantees uniformity in size for batch training in neural networks. Asheer and Shilpa [22] used normalization techniques that work with mean subtraction to scale pixel values for quicker model convergence. In domains like medical imaging, Sharma et al. [23] used denoising techniques to eliminate noise while preserving details. Data augmentation increases the size and diversity of the dataset through transformations such as rotation and color jitter, thereby reducing the risk of overfitting. Contrast and brightness adjustments improve image visibility. Cropping focuses on the areas of interest in the image. Researchers tailor images for algorithms through color space conversion and highlight important regions using selective segmentation [24]. The dataset is usually divided into training, validation, and testing sets at a ratio of 70:15:15; edge detection improves feature extraction. Color balancing corrects illumination disparities, while compression reduces file size without compromising quality.

The integration of segmentation, data augmentation, and classification constitutes a critical framework for developing accurate and robust AI-based systems for automated vitiligo image analysis. To localize the depigmented lesion, segmentation helps precisely identify it, remove irrelevant background and focus on important regions. Classification then identifies the type of the lesion based on extracted features. Data augmentation increases the size and diversity of the dataset and may improve classification accuracy. However, it also introduces challenges related to image transformations and synthetic sample generation. This review examines the synergistic role of segmentation, classification, and augmentation in robust AI-based vitiligo diagnosis.

It is important to distinguish between AI-based image classification and clinical diagnosis of vitiligo. In this review, image-based classification refers to the automatic classification of images using learned visual features, and should not be interpreted as equivalent to clinical diagnosis. The clinical diagnosis of vitiligo may be made by a comprehensive evaluation including patient history, lesion distribution and morphology, physical examination, Wood's lamp examination, dermoscopy, and evaluation by a qualified dermatologist. Thus, the performance reported by AI-based image classification studies reflects the potential of automated image analysis. It should not be considered a substitute for clinical diagnosis, especially in the absence of external validation and prospective clinical evaluation.

The remainder of this paper is organized as follows. Section 2 describes the review methodology. Section 3 presents vitiligo types, dataset availability, and related challenges. Section 4 discusses segmentation techniques, while Section 5 reviews data augmentation methods. Section 6 presents visualization techniques. Sections 7 and 8 discuss DL approaches and performance evaluation metrics, respectively. Section 9 outlines the challenges and future directions, Section 10 presents the discussion, and Section 11 concludes the paper.

2. REVIEW METHODOLOGY

2.1 Research questions

To systematically study the current state of the AI-based vitiligo classification methods, this research guides four research questions. These questions aim to examine existing methodologies, performance, and challenges associated with segmentation, classification, and augmentation.

RQ1: What AI techniques and architectures have been devoted to segmentation of vitiligo lesions? How effective are they?

RQ2: Which DL models have been used for automated vitiligo classification? Which approaches perform better?

RQ3: How are traditional data augmentation techniques used in research, what are their limitations, and how do they contribute to improving model robustness?

RQ4: Which datasets, standard resources, and evaluation metrics are most commonly used in AI-based vitiligo studies, and how do they influence the assessment of model performance?

These research questions provide a structured framework for evaluating existing literature, identifying current trends, comparing methodological approaches, and highlighting research gaps.

2.2 Survey methodology and article selection

The study conducted a systematic literature search across four major databases: IEEE Xplore, PubMed, Scopus, and Web of Science. Additional resources, including Kaggle, DermNet, and Mendeley, brought forth an initial collection of 286 records. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, the selection process followed four sequential phases: (1) identification, (2) screening, (3) eligibility, and (4) inclusion.

During the identification phase, the researchers retrieved 286 records from the selected databases. After removing 39 duplicate entries, 247 unique records remained for further consideration. During the screening phase, the researchers independently evaluated the titles and abstracts against the predefined inclusion and exclusion criteria and removed 95 irrelevant records. The remaining 152 articles proceeded to the next eligibility phase. Full-text assessment was performed on these articles. The screening process excluded a total of 41 articles. The researchers excluded these articles due to incorrect study design (n = 15), non-English language (n = 8), insufficient data reporting (n = 10), and unrelated conditions (n = 8). Following the final quality and relevance assessment, the researchers included 111 articles in the review. Included literature was categorized into three thematic clusters:

(1) Clinical and medical studies focusing on vitiligo pathology, epidemiology, and treatment (n = 18).

(2) DL classification studies focusing on transfer learning and XAI (n = 55).

(3) Vitiligo-specific AI studies covering segmentation, augmentation, dataset curation, and performance evaluation (n = 38).

The complete selection process is illustrated in Figure 1.

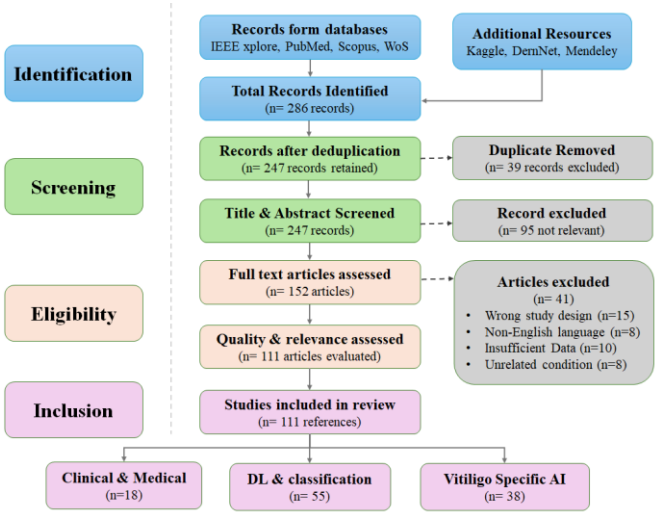

Figure 1. Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flow diagram for systematic literature selection process for artificial intelligence (AI)-based vitiligo classification studies

Researchers conducted a systematic literature search across major scholarly databases to identify studies on the application of AI and DL in vitiligo analysis. The researchers used Boolean operators to develop a primary search string incorporating keywords related to vitiligo, AI, and image analysis. The researchers reviewed the reference lists of the selected articles to identify additional relevant studies. The review mainly considered studies published between 2015 and 2025. The researchers also included highly relevant studies published before 2015 if they provided foundational or directly applicable contributions to vitiligo diagnosis, image analysis, or related AI methodologies. The included studies either directly addressed vitiligo or presented methodologies applicable to vitiligo image analysis.

The researchers excluded studies that did not align with the review objectives, lacked sufficient methodological details, or were not peer-reviewed. However, they included some non-vitiligo studies when they presented relevant methodologies, such as DL, image preprocessing, segmentation, data augmentation, or XAI techniques. The researchers assessed study quality using a checklist covering datasets, models, preprocessing, validation, evaluation metrics, and methodological challenges.

While reviews on the role of AI in dermatology are available, this work is a targeted PRISMA-based systematic review on DL for vitiligo image analysis. It synthesizes literature on CNN architectures, datasets, and methods, focusing on dataset diversity, skin-tone representation, and validation practices.

3. VITILIGO TYPES, DATASETS, AND THEIR CHALLENGES

3.1 Vitiligo types and their challenges

Vitiligo stands as one of the most distinctive and clinically complex depigmentation disorders. It is clinically categorized according to the distribution, extent, and progression of depigmented lesions. Al-Smadi et al. [25] described the major clinical types of vitiligo, including non-segmental, segmental, focal, universal, and mixed/unclassified types. In a previous study, Pawar and Hamid Khan [26] explored the distinctive characteristics of vitiligo. These include blue vitiligo, punctate vitiligo, follicular vitiligo, quadrichrome vitiligo, and inflammatory vitiligo, as shown in Figure 2.

Vitiligo has a wide range of clinical patterns that make classification difficult. NSV has symmetrical white patches, segmental affects one side of the body, focal isolated spots, universal involves most of the body parts, and mixed is the overlapping of most of the characteristics. Additional variants of vitiligo have also been identified based on their clinical appearance. Blue vitiligo is characterized by blue-gray lesions associated with deeper pigment loss. Punctate vitiligo presents as multiple tiny depigmented spots. Inflammatory vitiligo is characterized by a reddish border surrounding an active lesion. Follicular vitiligo involves pigment loss that begins around the hair follicles. Quadrichrome vitiligo exhibits four distinct shades of pigmentation, which may occur during active disease progression or repigmentation [6]. Classification is a very challenging task due to these diverse characteristics of vitiligo. Table 1 shows the major types and their challenges.

Table 1. Types of vitiligo and their challenges

Vitiligo Type	Key Characteristics	Major Challenges
Non-segmental vitiligo (NSV)	Symmetric, bilateral depigmented patches; commonly on face, hands, knees, trunk; progressive course	High intra-class variability
Segmental vitiligo (SV)	Unilateral patches following dermatome distribution; rapid early progression then stability	Limited dataset availability due to rarity
Focal vitiligo	Small, isolated lesions; early-stage localized spots	Very small lesion size leads to poor feature extraction
Universal vitiligo	>80–90% body depigmentation; widespread progression	Full-body imaging is difficult due to privacy and variability
Mixed vitiligo	Combination of segmental and non-segmental features	Label ambiguity (experts disagree on class)
Punctate vitiligo	Numerous small confetti-like depigmented macules	Micro-lesions often missed by segmentation models
Inflammatory vitiligo	Red or raised borders around lesions indicating inflammation	Redness overshadows true lesion boundaries
Follicular vitiligo	Depigmentation begins around hair follicles before spreading	Requires dermoscopy-level resolution
Quadrichrome vitiligo	Four-color appearance: white, hypopigmented, normal, and hyperpigmented zones	Color transitions confuse pixel-based classifiers
Blue vitiligo	Blue-gray hue due to dermal melanin or deeper destruction	Rare – extremely limited training samples

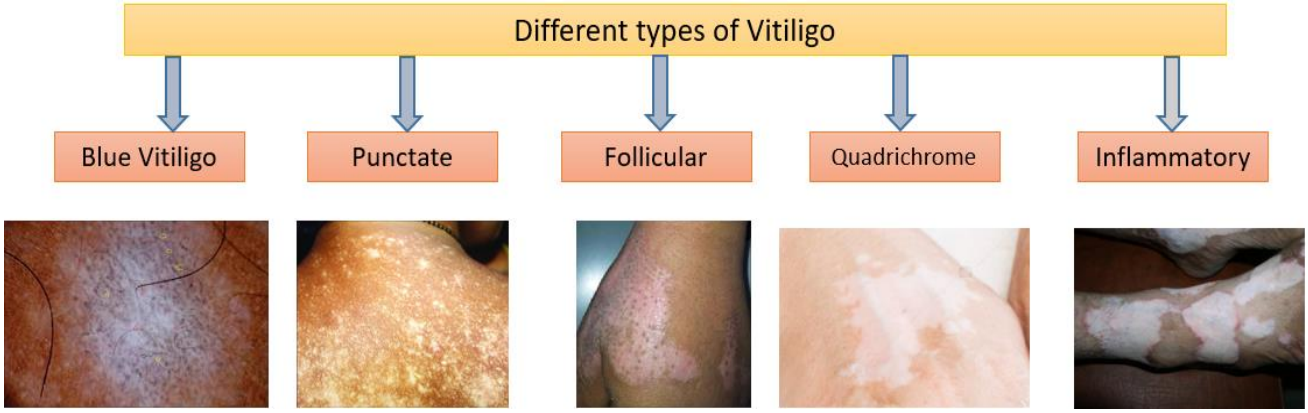

Figure 2. Vitiligo types based on their characteristics

3.2 Dataset availability and their limitations

The study [6] analyzed 11,483 images from 4,744 patients with Fitzpatrick skin types III and IV, various hypopigmented dermatoses (HD), including vitiligo (VI), pityriasis alba (PA), and other conditions. The ethics board approved the study, and all patients provided informed consent. Its performance was compared with that of 98 dermatologists. Pawar et al. [27] used the PUMCH-ISD public dataset available on Kaggle, with 463 vitiligo images as an input for a pretrained model. The dataset on Kaggle [28] consists of 4,098 skin disease images frequently used for vitiligo classification and detection. Collecting samples from multiple sources can result in inconsistent image quality and labeling accuracy. Researchers need to take care of filtering, cleaning and reannotation of images. Mendeley Data [29] having different skin disease classification dataset like acne, hyperpigmentation, Nail psoriasis, and vitiligo. The vitiligo dataset contains 2,016 images captured from different body parts. The dataset contains variable quality, inconsistent labels and limited metadata. Kaggle [30] contained 3,628 clinical images, including 2,100 vitiligo images and 1,528 healthy-skin images. These images represent various body parts and skin tones, with variations in lighting conditions and image resolution. Images are curated for binary classification and include close-up patches of common body regions such as the face, hand, and trunk. A total of 101 vitiligo images are available in the collection [31], featuring individuals of different ages and showing lesions across various body regions. Kaggle [32] consisted of 1,271 images from 891 vitiligo and 380 healthy skin images.

Figure 3 summarizes the use of vitiligo image datasets in AI/machine learning (ML) dermatology research. Although vitiligo image datasets are becoming increasingly accessible, building robust DL models is still a challenge. Challenges related to patient information, image modality, annotation quality, and limited skin-tone and demographic diversity may contribute to overfitting as well as biased model performance. Variations in image acquisition and limited annotation details can affect data consistency and label reliability.

Wang et al. [33] developed the dataset, comprising 1,950 clinical photographs and 7,798 dermoscopic images from 1,174 patients. It focuses on 8 inflammatory skin diseases: psoriasis, dermatitis (including atopic, contact, seborrheic types), lichen planus, pityriasis rosea, vitiligo, acne vulgaris, rosacea, and morphea. Along with this, the author used the publicly available multimodal dataset Derm7pt to develop a model. The dataset publicly available at Kaggle [34] has 583 clinical and dermoscopic images collectively.

The diversity in anatomical sites, image sources, and skin appearance across these resources offers a good foundation for creating robust AI models. Researchers collect samples from multiple sources, which can result in variations in image quality. They may also encounter inconsistencies in labeling accuracy. Hence, these datasets provide a good basis for methodological research. The future development of datasets can enhance clinical relevance with full patient metadata, standardized imaging conditions, disease stage information, and expert-validated annotations. Importantly, the available documentation does not always specify whether researchers acquired the images using standardized clinical or dermoscopic protocols. Therefore, the researchers characterize the datasets based on the modality and information reported by the original sources. They do not assume a specific imaging modality or annotation process. Future research should emphasize standardized cleaning and preprocessing of vitiligo images to improve image quality, reduce acquisition-related variability, and enhance the reliability and generalizability of AI models.

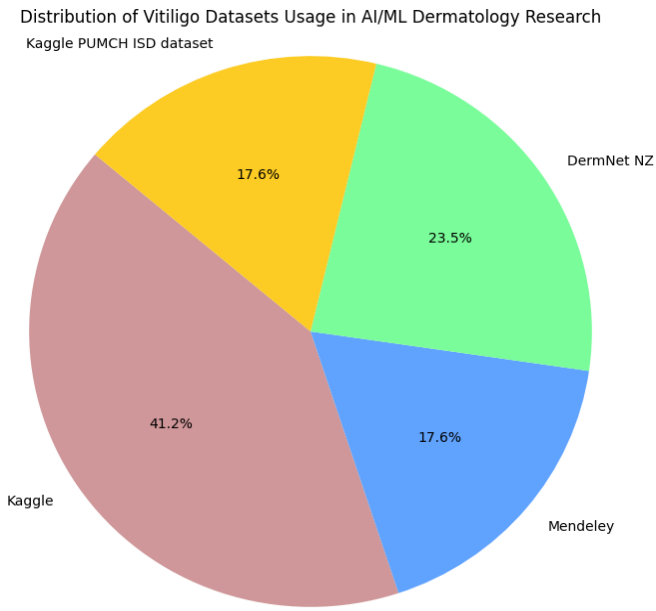

Figure 3. Distribution of vitiligo dataset usage in artificial intelligence (AI)/machine learning (ML) dermatology research

4. SEGMENTATION TECHNIQUES

Segmentation is an important step to identify the exact lesion from the image. Thresholding-based segmentation is a simple technique that separates the image into regions based on pixel intensity. Jia et al. [35] highlighted global thresholding, which works on a single threshold value and classifies pixels into different regions. Bradley and Roth [36] focused on adaptive thresholding applied with varying thresholds to account for uneven lighting. Guo et al. [37] suggested the fastest Otsu's method for segmentation. This method automatically determines the optimal threshold by minimizing intra-class variance. Figure 4 shows an example of threshold segmentation where lighting affects the resultant output because the method solely depends on pixel intensity. Selecting an appropriate threshold value is also challenging. Figure 5 illustrates edge-based segmentation using the Sobel, Prewitt, Canny, and Laplacian of Gaussian (LoG) operators. In this example, the Sobel filter produces the best results; however, the performance may vary depending on the characteristics of the input images. Region-based segmentation is shown in Figure 6, which groups pixels into meaningful regions based on color, intensity, or texture. It generally performs well on images with homogeneous areas, but the challenging task is to choose the appropriate seed point to define suitable similarity.

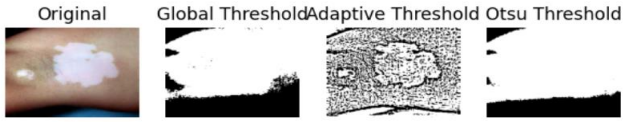

Figure 4. Sample image segmentation using threshold-based segmentation technique

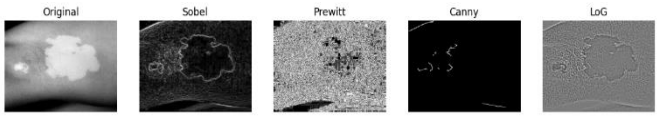

Figure 5. Sample of image segmentation using edge-based segmentation

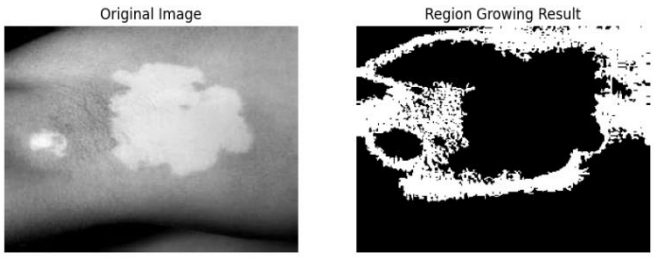

Figure 6. Sample of image segmentation using region-based segmentation

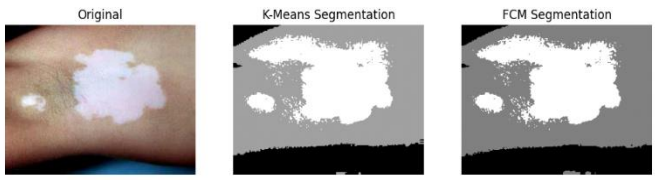

Figure 7. Sample image segmentation using K-means and fuzzy C-means (FCM) segmentation

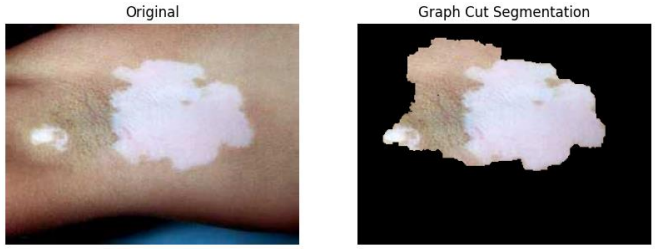

Figure 8. Sample of image segmentation using graph-based segmentation

Figure 7 shows examples of image segmentation with K-means and fuzzy C-means (FCM) based segmentation. K-means image is divided into hard clusters based on intensity, producing distinct grey regions where vitiligo patches appear in separate clusters with sharp boundaries, along with noise. FCM output provides smoother, more natural-looking segmentation because it assigns pixels to clusters with varying membership levels. It allows gradual transitions between skin tones and better representation of lesion boundaries.

Figure 8 shows the poor result from graph cut segmentation due to sensitivity to initialization, low contrast between the lesion and surrounding skin, and complex backgrounds. Graph-cut segmentation relies heavily on color differences and an approximately correct bounding rectangle around the lesion; consequently, irregularly shaped lesions may be only partially segmented.

Liu et al. [38] proposed a skin lesion segmentation network which is the combination of attention mechanism and redundancy reduction strategy on different International Skin Imaging Collaboration (ISIC) datasets with the highest F1-score 7.75%. The proposed samDA model achieved a mean intersection over union (IoU) of 75.2% in the performance of edge segmentation [39]. Mazzetto et al. [40] reviewed deep learning methods for vitiligo lesion segmentation and severity scoring and concluded that automated image analysis can replace time-consuming manual measurement, with reported accuracies of up to 97.4%. Chen et al. [41] created an AI method to assess skin pigmentation that segments photographed lesions and scores them by the facial vitiligo area scoring index (F-VASI). The authors achieved an accuracy of 81% by using a U-Net with a CNN. Wang et al. [42] propose GloW-VSNet, demonstrated on two public datasets along with two private datasets, University of British Columbia (UBC)-EEM-Vitiligo (UEV) and UBC-VIT. Compared with existing models, the proposed model achieved the highest Dice score, which is 87%, IoU 78% and Accuracy 96%. Li et al. [43] proposed a model that used a tailor-made contrast enhancement scheme and long-range comparison for better performance. The model achieved a mean error of 1.17% and a maximum error of 3.11%. Li et al. [44] introduced an image synthesis algorithm for segmentation on 843 vitiligo images from the internet with CNN. The proposed model achieved only 10.6% error.

5. AUGMENTATION TECHNIQUES

In the survey, it is observed that there are many datasets available but with limited images. To increase the size of the dataset, augmentation plays a crucial role. Applying transformations such as rotation, flipping, or color changing will help to extend the dataset without requiring additional

real-time data. Park et al. [45] presented the effectiveness of different augmentation techniques that will help enhance dataset size. Shorten and Khoshgoftaar [46] mainly focused on smart augmentation, which utilizes the concept of neural augmentation. Lama et al. [47] introduced a lesion-aware MixUp augmentation technique, specifically focusing on dermatological images. Table 2 shows the different augmentation techniques used by researchers and their performance.

This systematic review is focused on the applications of DL for vitiligo image analysis. It includes a selection of works from other medical imaging domains for methodological context. It reviews segmentation architectures and augmentation techniques, highlighting their evolution and potential for vitiligo research. The focus is primarily on vitiligo-specific studies, but cross-domain references are also included to enhance the understanding of relevant methodological advances.

Table 2. Different augmentation techniques with their performance

Reference	Augmentation Technique	Performance
[47]	Lesion-Aware MixUp augmentation	The proposed technique improved segmentation performance by 0.08%, increased the Jaccard score from 0.7947 to 0.8013, and increased the Dice score by 0.6% from 0.8714 to 0.8766.
[48]	Generative adversarial network (GAN)-based synthetic medical image augmentation	The proposed technique gives a classification performance of 78.6% sensitivity and 88.4% specificity. The results increased to 85.7% sensitivity and 92.4% specificity by adding synthetic data augmentation.
[49]	Random data augmentation-based enhancement	COVID-19 chest X-ray, KiTS19 dataset, and for red, green, and blue (RGB) imagery with: LC25000 dataset is used with improved accuracy.
[50]	GAN-Based augmentation methods for dermoscopy images	Integrates content loss, structural similarity index measurement, and smooth L1 loss. The classification accuracy is increased by 3.7% compared to the existing model.
[51]	Multidimensional data augmentation method	Combines affine transformations with random oversampling to enhance data diversity and address class imbalance issues. Improved segmentation performance, with the Mask-RCNN model achieving the best pixel-wise segmentation for both lesion types, yielding Dice similarity coefficient (DSC) values of 0.829 for pulmonary nodules and 0.879 for liver tumors, closely matching state-of-the-art results.
[52]	KeepMask and KeepMix,	Suitable for multi-class segmentation along with binary segmentation. The authors used the Dice similarity coefficient as an evaluation metric.
[53]	MediAug	Combination of six augmentation methods like MixUp, YOCO, CropMix, CutMix, AugMix, SnapMix evaluated on brain tumor MRI and eye disease datasets. Classification accuracy: 79.19% for brain tumor.

6. VISUALIZATION IN VITILIGO CLASSIFICATION

Because of the different characteristics of vitiligo, classification is a tedious task. Visualization helps AI experts to classify the disease. Accurate classification helps dermatologists to understand the behavior of the disease by choosing appropriate treatment and to monitor changes in lesions. Table 3 gives insights into various visualization techniques used by the researchers and their significance. Various DL models help to classify vitiligo automatically by analyzing the color patterns, texture, shape of the patch, etc. Zhong et al. [54] used 4,320 dermoscopic images, among them 2,678 vitiligo images and 1,642 images of hypopigmented skin diseases, to compare ResNet34, ResNet50, ResNet101, and Swin Transformer. ResNet50 achieved an accuracy of 88.49%; ResNet101 achieved an accuracy of 87.18%. Swin Transformer Large achieved an accuracy of 93.82%. Figure 9 shows the usage of performance metrics other than accuracy, such as area under the curve (AUC), F1-score, Dice coefficient, Jaccard Index, and mean Dice similarity coefficient (DSC) from some of the available research papers.

Table 3. Different classification models, their performance, and visualization used by the author

Reference	Dataset	Model	Performance	Visualization
[54]	Kaggle vitiligo vs. healthy-skin dataset	MobileNet, ResNet50, EfficientNetB4, InceptionV3, Xception, Swin Transformer, Firefly-DenseNet201	Swin Transformer Large: accuracy 93.82%, AUC 94%, sensitivity 94.02%, specificity 93.5%	ROC curves and confusion matrices across five models
[55]	Large-scale dataset of 2,974 patients	Two-stage CNN classification + LLM-based report generation	AUC of 0.9906, with vitiligo achieving the model's highest per-class performance	LLM-generated diagnostic reports plus classification confidence outputs
[56]	1,267 wide-field vitiligo images	Swin R-CNN	Average precision of 84.72%, average recall of 77.81%	Lesion segmentation mask overlays
[57]	Chinese in-house dataset (2,876 images) plus worldwide public dataset (1,341 images) of vitiligo	VGG13, ResNet18, DenseNet121	VGG13 achieved an F1-score of 0.9684, surpassing dermatologists' 0.9221	Not reported
[58]	Public dataset of 10,015 images	4-CNN	93.6% accuracy	ROC curve is used for visualization

[59]	Vitiligo images captured under Wood's lamp illumination (in-house)	CycleGAN (image generation) + Attention-Aware DenseNet with Residual Deconvolution (ADRD, super-resolution) + ResNet50 (classification)	85.69% classification accuracy	CycleGAN-generated Wood's lamp images and super-resolved outputs shown as visual comparisons
[60]	1,309 images including segmental and generalized vitiligo	CNN (binary classification)	Binary classification accuracy of 95%	CNN-generated depigmentation maps compared against Wood's lamp imaging
[61]	Vitiligo images of 833 participants	Computer-aided repigmentation assessment	Repigmentation rate quantification	Repigmentation-rate visual tracking over treatment course
[62]	International Skin Imaging Collaboration (ISIC) archive, biopsy-proven dermoscopic images of melanoma and nevi. 4,204 biopsy-verified images	Deep CNN	Sensitivity of 82.3%	ROC curves and performance comparison plots
[63]	Dermatological vitiligo lesion images	VGG19-ResSE (VGG19 + squeeze-and-excitation blocks + residual connections) vs. YOLOv8, compared	Accuracy: 95.5%, Dice coefficient: 0.92 (segmentation); YOLOv8 comparator: F1-score 0.92	ROC curves and confusion matrices
[64]	696 images in total: 368 vitiligo-infected	InceptionV3, VGG16, VGG19, SqueezeNet	InceptionV3 + Logistic Regression: Accuracy 98.0%, AUC 0.997. VGG19 + Logistic Regression: Accuracy 98.0%, AUC 0.998. VGG16 + Neural Network/KNN: Accuracy 97.7%, AUC 0.998.	ROC curve
[65]	Dermatological vitiligo skin images	KNN, SVM, Gradient Boosting, CNN, and ANN, InceptionV3 + CNN, InceptionV3 + Random Forest	InceptionV3 + Random Forest: 99.90% accuracy, AUC 1.0; InceptionV3 + CNN: 99.80% accuracy, AUC 1.0	Bar charts comparing AUC across models
[66]	129,450 clinical images for training. 16,114 images for initial validation	Deep CNN with Inception-v4	Top-1 accuracy: 0.71 and Top-3 accuracy: 0.93	Confusion matrices to show class-wise performance for the 26 skin disease categories.

Note: GAN = generative adversarial network, ROC = receiver operating characteristic, LLM = large language model; kNN = k-nearest neighbors; SVM = support vector machine; ANN = artificial neural network.

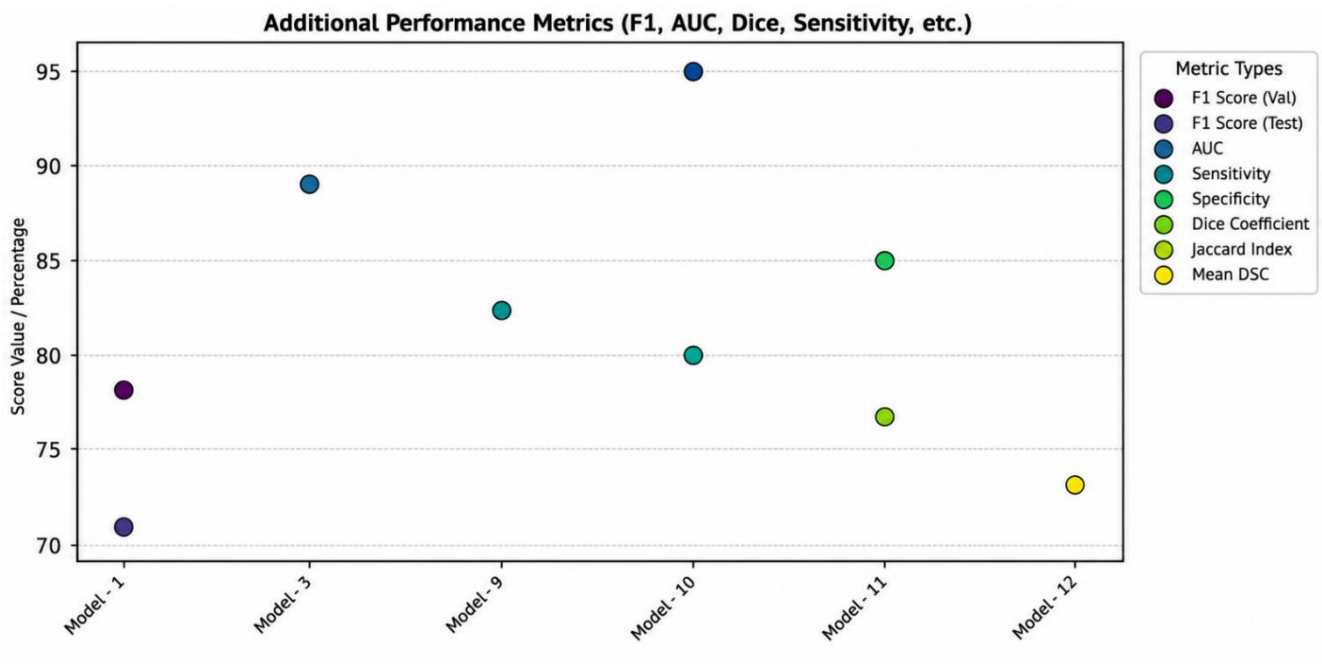

Figure 9. Comparison of additional performance metrics

7. DEEP LEARNING APPROACHES

DL is a subset of ML which employs a multilayer artificial neural network as shown in Figure 10. It helps to learn hierarchical representations of large datasets. Vakalopoulou et al. and Montesinos López et al. [67, 68] introduce a fundamental principle of DL that mainly focuses on CNN perception. CNNs can be applied to both 2D and 3D images. 2D CNNs process grayscale and red, green, and blue (RGB) images through convolution using kernels to extract features; 3D CNNs handle volumetric data and focus on spatiotemporal features.

CNN's inbuilt feature engineering techniques make it more popular compared to traditional classification techniques. Many studies have employed pre-trained DL architectures, including AlexNet [69], GoogLeNet [70, 71], and VGGNet (VGG16 and VGG19) [72-74]. Other commonly used architectures include ResNet (ResNet50 and ResNet101) [75, 76], DenseNet [77-79], MobileNet [80-82], and EfficientNet [83-85]. In addition, fine-tuning [86-88] and transfer learning [89-91] have been widely adopted to improve model performance and enable effective learning from limited dermatological datasets. Researchers use data augmentation techniques to improve classification model accuracy. In recent years, DL has been state-of-the-art in image classification in the healthcare domain. AlexNet achieves 95.95% accuracy by utilizing the UC Merced, Google Image dataset of SIRI-WHU, and WHU-RS.

The researchers derived the performance metrics presented in Table 4 from independent studies. These studies used different datasets, sample sizes, class distributions, preprocessing pipelines, and validation protocols. Therefore, the reported results should be compared with caution because the evaluation criteria may also differ. Thus, the reported metrics should not be considered direct comparative benchmarks. The table provides a consolidated overview of the methodologies and performance achieved under respective experimental conditions. Therefore, any comparison should be considered qualitatively to disclose methodological trends and research directions, instead of as definitive proof of the precedence of one approach over another.

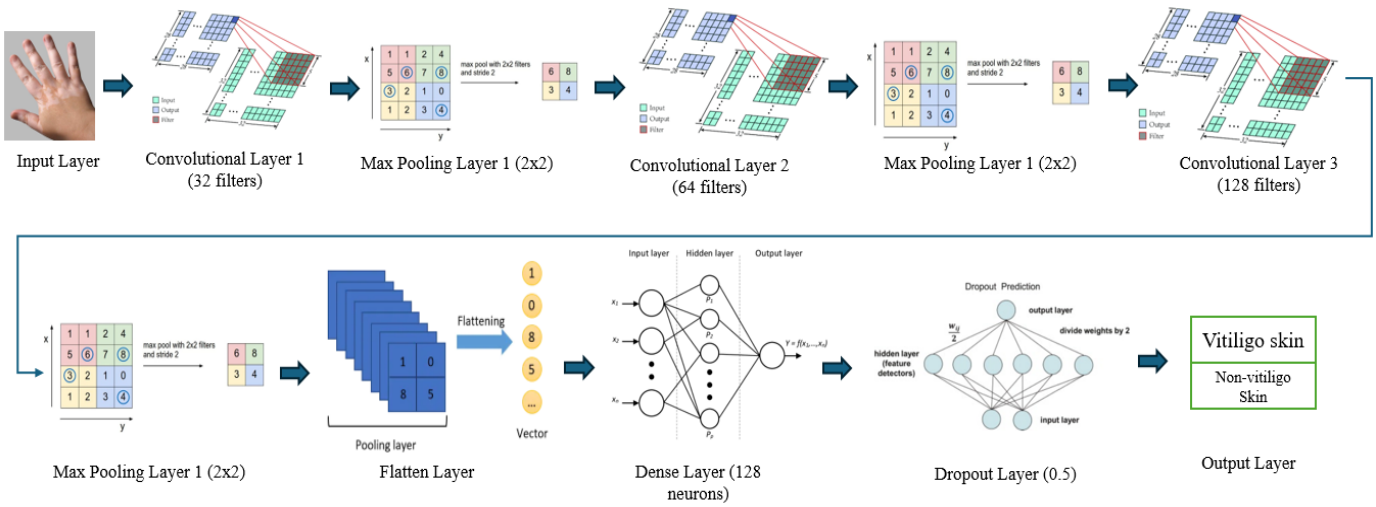

Figure 10. Convolutional neural network (CNN) layered architecture

Table 4. Performance and limitations of deep learning (DL) architectures

Reference	DL Method / Architecture	Dataset Used	Performance Evaluation	Key Limitations
[92]	Ensemble: EfficientNet-B4 + DenseNet201 + ResNet50 with CBAM and U-Net preprocessing	HAM10000 (trained), validated on ISIC-2019 and PH2	Accuracy 98.95%; Balanced accuracy 98.7%; Sensitivity (melanoma) 99.6%	Limited labelled dataset and potential ensemble complexity; risk of overfitting to training set
[92]	DenseNet variants (used in ensemble)	HAM10000 (ensemble)	Ensemble metrics reported above; standalone DenseNet metrics not provided	Memory/compute trade-offs and potential overfitting in small classes
[93]	HyperFusion-Net (MPViT + attention U-Net fusion)	Four public ISIC datasets (>60k images)	Accuracy 93.24%; AUC 95.80%; Segmentation Dice 94.5%	Computational cost of multi-path fusion; needs segmentation pipeline; limited clinical validation
[93]	U-Net used with classification pipeline (segmentation and classification)	ISIC / dermoscopic datasets (used as segmentation backbone in several studies)	Segmentation Dice reported for Hyper Fusion-Net (94.5%); individual U-Net classifier metrics not specified in snippets	Segmentation dependency can add pipeline fragility and extra annotation needs
[93]	DeepLabV3+ (segmentation baseline mentioned)	ISIC datasets (as baseline comparator)	Not Reported	Large model size and segmentation-to-classification integration complexity
[94]	CNN-Transformer-GRU fusion with StyleGAN3 augmentation	HAM10000	Accuracy 90.61%; Precision 91.11%; Recall 95.28%; AUC 97%	Synthetic augmentation biases, architectural complexity, and higher compute requirements

[95]	Novel lightweight DCNN (proposed DCNN)	ISIC 2016, ISIC 2017, ISIC 2020	Accuracy: ISIC2016 81.41%; ISIC2017 88.23%; ISIC2020 90.42%	Lower performance vs. larger models, limited external testing
[96]	InceptionV3 (transfer learning)	MedNode, PH2, HAM10000	Accuracy: MedNode 97.1%; PH2 97.2%; HAM10000 96.2%	Dataset bias (high performance on specific datasets may not generalize)
[96]	GoogleNet / Inception-style (baseline)	MedNode, PH2, HAM10000	InceptionV3 numbers reported; GoogleNet specific numbers not provided	Architecture-specific tuning required; dataset biases
[97]	VGG16 (pretrained baseline)	MedNode, PH2, HAM10000	Not Reported	High parameters count and computational cost; class imbalance effects
[97]	Fuzzy GrabCut + stacked CNN (GC-SCNN)	Dataset not precisely stated in snippets	Not Reported	Unclear external validation and limited reported numeric comparisons
[92, 98]	ResNet50 (single model, and used inside ensembles)	HAM10000 and others	Not Reported	Parameter complexity and potential domain shift sensitivity.
[98]	Weighted ensemble transfer learning approach	Dataset(s) referenced in paper (melanoma classification studies)	Not Reported	Model weighting choices may introduce selection bias and reproducibility concerns
[99]	Multi-CNN comparison (eight contemporary CNNs)	PH2, ISIC 2016, ISIC 2017, HAM10000	Paper reports per-model results but exact numbers are not extractable from supplied snippet—insufficient evidence for specific metrics	Heterogeneous evaluation protocols across models complicate comparisons
[100]	Generic CNN (classical CNN approaches vs traditional ML)	ISIC datasets (study comparing CNN vs classical ML)	The study reported that CNN outperformed Logistic Regression, SVM, RF, and KNN	Interpretability and dataset variability challenges noted in comparisons
[101]	Review aggregated methods (DL in ISIC 2016/2017/2020 etc.)	ISIC 2016, ISIC 2017, ISIC 2020 and others	Review summarizes many reported metrics across studies but does not give a single standardized numeric table in supplied snippet—insufficient evidence for per-method numbers	Cross-study heterogeneity, dataset bias, and limited clinical translation emphasized

Note: The reported performance metrics come from various published studies. Due to differences in datasets, class distributions, preprocessing techniques, validation protocols, and evaluation settings, the results serve only for qualitative comparison and should not be interpreted as direct quantitative benchmarks. CNN = convolutional neural network, KNN = k-nearest neighbors, ML = machine learning, DL = deep learning, CBAM = convolutional block attention module, MPViT = multi-path vision transformer, GRU = gated recurrent unit, DCNN = deep convolutional neural network, RF = random forest, ISIC = International Skin Imaging Collaboration.

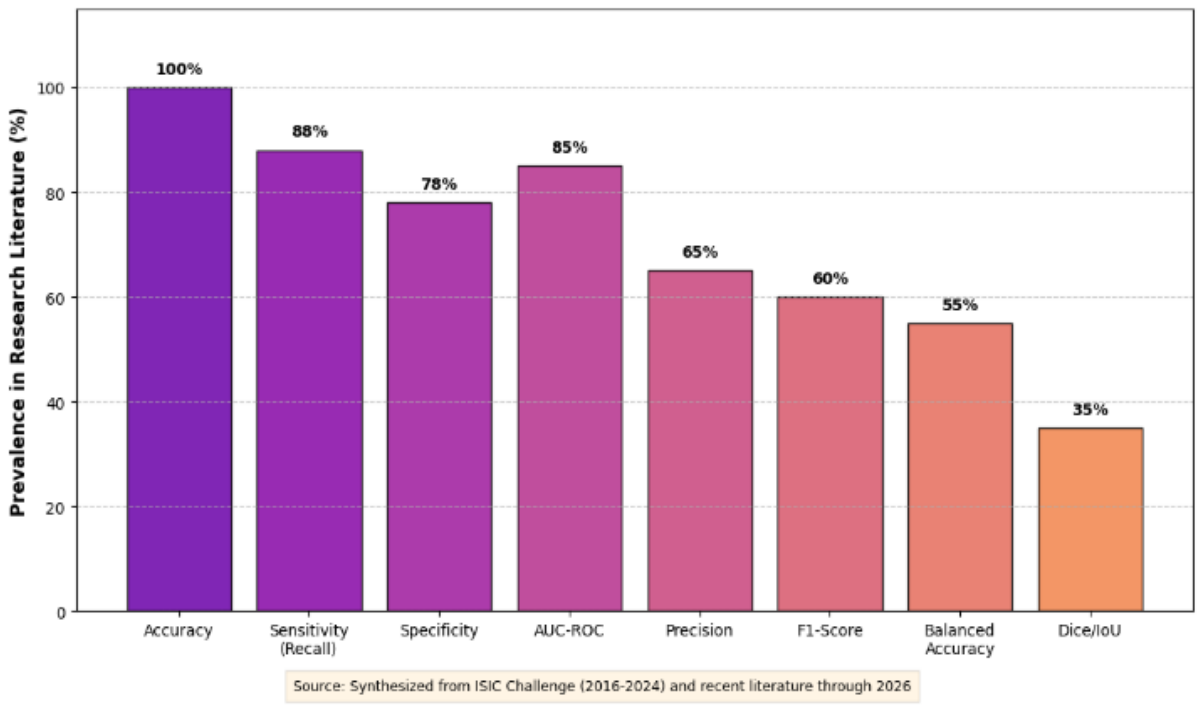

Figure 11. Performance evaluation metrics in skin image classification (2016–2026)

8. PERFORMANCE EVALUATION METRICS

As per Table 4 and as shown in Figure 11, more than 50% of researchers mainly focused on accuracy, sensitivity, specificity, and F1-score parameters. But studies [102-105] also focused on the Dice/IoU parameter, Low et al. [106] worked on Jaccard Index (JI). As vitiligo patches have irregular borders, small-sized patches, and varying contrast, Jaccard and Dice quantify the clinical relevance of segmentation and support objective comparison of DL models. Accuracy alone may overlook boundary quality. It can be influenced by the large number of background pixels. Dice and IoU provide better evaluation by focusing on lesion pixels and measuring spatial overlap. Evaluating the model using a combination of the above metrics helps to enhance the model performance.

9. CHALLENGES TO BE FACED AND FUTURE DIRECTIONS

9.1 Image appearance issues

Lesion and skin appearance variability can hinder segmentation and feature extraction. Studies [107-109] highlight factors such as illumination variation, vague lesion boundaries, confounders of skin surface, and clinical look-alikes. These factors make consistent pixel-level detection difficult and increase false positives and negatives.

9.2 Data and generalizability

Data challenges and bias in datasets are central obstacles to deploying the model. The report documents restricted subject numbers, uneven skin type representation, and a drop in cross-domain performance. Studies [107, 110] discussed skin type bias, small or narrow cohorts, the need for richer datasets, and annotation scarcity.

9.3 Algorithmic evaluation hurdles

The model must localize lesions accurately and provide clinically meaningful severity metrics. Gopalakrishnan et al. and Tanvir et al. [107, 111] discussed segmentation performance variability, computational demands, the lack of standardized severity metrics and inter- and intra-observer variability. Comparative studies report high within-set scores but expose weaknesses when tested across settings or when computational costs rise.

Recent research highlights several concrete development paths to overcome current gaps and move toward clinically useful tools.

(1) Unsupervised and ensemble segmentation approaches can reduce annotation requirements and improve localization robustness. Preliminary studies have reported high internal accuracy and Dice scores.

(2) Transformer and self-attention architectures have been explored to improve contextual feature modeling. These approaches can be combined with multimodal inputs, such as clinical metadata and multiple imaging modalities, along with explicit uncertainty calibration to produce more reliable disease-stage diagnoses.

(3) Ensemble and hybrid pipelines can combine detection, segmentation, and colorimetric morphometry to yield objective severity measures suitable for longitudinal tracking.

(4) Benchmarking and dataset initiatives driven by multicenter data collection and standardized evaluation protocols, as recommended by systematic reviews, to enable fair comparison and safer translation.

10. DISCUSSION

This systematic review examined various research articles related to segmentation, vitiligo classification, and augmentation using DL. Earlier studies relied mainly on clinical observations and consensus-driven classification. These approaches may be less efficient when applied to digital images. Furthermore, there is a lack of reliable and automated methodologies for vitiligo analysis. The following discussion synthesizes the evidence gathered from the 112 reviewed studies to address each research question and identify the critical gaps in the current literature.

RQ1: AI segmentation techniques for vitiligo analysis.

With respect to the first question, the reviewed literature shows an evolution from classic to DL-based segmentation approaches. Early methods, including Otsu thresholding, adaptive thresholding [36, 37], region growing [40], and FCM clustering [41, 42], offered computationally efficient solutions for lesion segmentation. However, these approaches often struggled with variations in lesion appearance, skin tone, and image noise. CNN-based architectures have significantly improved lesion boundary detection and segmentation accuracy. On the other hand, semi-supervised [104], weakly supervised [105], and ensemble-based models achieve good performance to address the challenge of limited annotated data.

RQ2: DL models for automated vitiligo classification.

The reviewed literature indicates that transfer learning using pretrained models is one of the most effective strategies for vitiligo classification when working with limited datasets. Most widely used architectures like ResNet [76], EfficientNet [83, 84], VGG [73, 74], and MobileNet [80, 82] have shown excellent classification performance with fine-tuning on vitiligo-specific datasets. Ensemble models and hybrid CNN-transformer architectures achieved higher diagnostic accuracy than individual models. Use of XAI also improves the model prediction accuracy. However, inconsistent evaluation metrics, heterogeneous datasets, and lack of widely acceptable benchmark datasets continue to be major limitations.

RQ3: Data Augmentation strategies for vitiligo AI research.

The reviewed study focused on traditional data augmentation techniques such as image flipping, rotation, cropping [45, 46], and photometric adjustments involving brightness and contrast modification. Researchers widely use these techniques to address the problem of limited datasets. MixUp-based [47] and oversampling fusion strategies [51] are further used to improve class balance and reduce overfitting across small samples. Traditional methods face several limitations in generating diverse and representative images. Therefore, researchers are increasingly exploring GAN-based augmentation and lesion-aware strategies to enhance training diversity and improve model robustness. Although these strategies have shown promising results, the individual contribution of each augmentation technique has yet to be investigated. A standardized vitiligo dataset benchmark is needed to evaluate their effectiveness.

RQ4: Dataset, resources, and evaluation metrics in vitiligo

research.

The review study also focused on publicly available dataset sources such as Kaggle, Mendeley, DermNet, and PUMCH databases. These datasets have a relatively small number of images and inconsistent annotation protocols, which directly constrain model generalizability. To improve the dataset size, augmentation will play a vital role. Classification studies most commonly use accuracy, sensitivity, F1-Score, and AUC-ROC, while segmentation prefers the Dice similarity coefficient and IoU as primary evaluation metrics. However, the use of different evaluation metrics across studies makes it difficult to compare results and reproduce the findings. Therefore, the lack of a unified vitiligo benchmark dataset and standardized evaluation framework remains a major challenge and is important to explore.

Transformer-based architectures, multimodal learning, and ensemble approaches offer promise for improving the robustness and generalizability of AI-based vitiligo image analysis. However, these techniques should be considered potential areas of research at this time rather than clinically proven solutions. Their clinical utility should be rigorously tested on independent external datasets, diverse patient populations and skin tones, standardized imaging conditions, and prospective clinical studies. For proof of their practical value, they also need to be compared with experienced dermatologists and assessed within real clinical workflows. Moreover, models intended for clinical use must meet relevant ethical, data-privacy, safety, and regulatory requirements, including appropriate validation and documentation prior to deployment.

11. CONCLUSION

Many researchers are working on vitiligo in a wide range of specifications all over the world. This review provides a comprehensive overview of existing approaches to automated vitiligo image analysis. As a study, it represents a comprehensive analysis of vitiligo types, segmentation, augmentation, and detection by examining the evolution of methodologies from traditional practice to AI models. The discussion underscores the constraints in all areas and suggests a suitable solution. Data augmentation plays a very important role in mitigating dataset scarcity and class imbalance for improving model generalization and performance.

Future studies should prioritize the creation of larger, diverse datasets encompassing various skin tones, ages, disease stages, and imaging conditions. Essential steps include establishing standardized annotation protocols and validated lesion boundaries for improved supervised learning reliability. Methodological advancements in self-supervised learning, Vision Transformers, multimodal learning, and foundation models are encouraged to enhance generalizability. Investigating XAI is vital for providing interpretable predictions and boosting dermatologist trust in AI diagnostics. Evaluation should extend to external, multi-center datasets for better clinical applicability.

Overall, this review highlights the role of DL in automated vitiligo assessment and outlines directions for future research.

CONFLICT OF INTEREST

Authors declared no conflict of Interest.

DATA AVAILABILITY

Data sharing is not applicable to this article. The study did not create or analyze any new data.

REFERENCES

[1] Bergqvist, C., Ezzedine, K. (2020). Vitiligo: A review. Dermatology, 236(6): 571-592. https://doi.org/10.1159/000506103

[2] Alikhan, A., Felsten, L.M., Daly, M., Petronic-Rosic, V. (2011). Vitiligo: A comprehensive overview—Part I. Journal of the American Academy of Dermatology, 65(3): 473-491. https://doi.org/10.1016/j.jaad.2010.11.061

[3] Yaghoobi, R., Omidian, M., Bagherani, N. (2011). Vitiligo: A review of the published work. Journal of Dermatology, 38(5): 419-431. https://doi.org/10.1111/j.1346-8138.2010.01139.x

[4] Lin, X.R., Meng, X.M., Lin, J.R. (2025). Segmental vitiligo: Autoimmune pathogenesis, neuronal mechanisms, and somatic mosaicism. International Journal of Dermatology, 64(3): 490-498. https://doi.org/10.1111/ijd.17627

[5] Speeckaert, R., Lambert, J., Bulat, V., Belpaire, A., Speeckaert, M., van Geel, N. (2020). Autoimmunity in segmental vitiligo. Frontiers in Immunology, 11: 568447. https://doi.org/10.3389/fimmu.2020.568447

[6] Huang, H.Q., Wang, C.Q., Gao, G., et al. (2024). Intelligent diagnosis of hypopigmented dermatoses and intelligent evaluation of vitiligo severity on the basis of deep learning. Dermatology and Therapy, 14(12): 3307-3320. https://doi.org/10.1007/s13555-024-01296-9

[7] Akl, J., Lee, S., Ju, H.J., et al. (2024). Estimating the burden of vitiligo: A systematic review and modelling study. Lancet Public Health, 9(6): e386-e396. https://doi.org/10.1016/S2468-2667(24)00026-4

[8] Ramot, Y., Rosenberg, V., Zhou, L.M., Harbers, S. (2024). Epidemiology and treatment patterns of patients with vitiligo: A real-world analysis. Advances in Therapy, 41(7): 2890-2906. https://doi.org/10.1007/s12325-024-02875-0

[9] Zhang, L., Mishra, S., Zhang, T.Y., et al. (2021). Design and assessment of convolutional neural network-based methods for vitiligo diagnosis. Frontiers in Medicine, 8: 754202. https://doi.org/10.3389/fmed.2021.754202

[10] Jain, S., Singhania, U., Tripathy, B., et al. (2021). Deep learning-based transfer learning for classification of skin cancer. Sensors, 21(23): 8142. https://doi.org/10.3390/s21238142

[11] Wang, G., Qian, P., Tian, R.R., Wang, J.Y. (2024). Lightweight convolutional neural network model for skin lesions classification. In Proceedings of SPIE 13185, International Conference on Communication, Information, and Digital Technologies, 13185. https://doi.org/10.1117/12.3033617

[12] Rashid, J., Ishfaq, M., Ali, G., et al. (2022). Skin cancer disease detection using transfer learning technique. Applied Sciences, 12: 5714. https://doi.org/10.3390/app12115714

[13] Mustafa, S., Jaffar, A., Rashid, M., Akram, S., Bhatti, S.M. (2025). Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet–random

forest for enhanced segmentation and classification. PLoS One, 20(1): e0315120. https://doi.org/10.1371/journal.pone.0315120
[14] Li, Y.J., Xu, C., Han, J.B., et al. (2022). MHAU-Net: Skin lesion segmentation based on multi-scale hybrid residual attention network. Sensors, 22: 8701. https://doi.org/10.3390/s22228701
[15] Luschi, A., Salman, A., Cartocci, A., et al. (2025). Enhancing dermatological diagnostics: A GAN-based approach for synthetic skin lesion image generation. In International Conference on Biomedical and Health Informatics (ICBHI 2024), IFMBE Proceedings, 118: 87-94. https://doi.org/10.1007/978-3-031-86323-3_12
[16] IEEE. (2024). 2894-2024-IEEE guide for an architectural framework for explainable artificial intelligence. In IEEE Std 2894-2024, pp. 1-55. https://doi.org/10.1109/IEEESTD.2024.10659410
[17] Volkov, E.N., Averkin, A.N. (2023). Explainable artificial intelligence in medical image analysis: State of the art and prospects. In 2023 XXVI International Conference on Soft Computing and Measurements (SCM), Saint Petersburg, Russian Federation, pp. 134-137. https://doi.org/10.1109/SCM58628.2023.10159033
[18] Alsayaydeh, B., Oakley, A. (2022). Vitiligo. DermNet NZ. https://dermnetnz.org/topics/vitiligo.
[19] Alavi, M., Kargari, M. (2022). A new contrast enhancement method for color dark and low-light images. In 2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Bam, Iran, Islamic Republic of, pp. 1-7. https://doi.org/10.1109/CFIS54774.2022.9756426
[20] Shobana, P., Deny, J. (2022). Image fusion-based removal of color artifacts for the enhancement of dermoscopy images. In 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, pp. 1-6. https://doi.org/10.1109/ICSTSN53084.2022.9761322
[21] Yu, S.J., Chen, Z.L., He, J.Y., Wang, H. (2025). Comparative study of dermatologists and deep learning model on diagnosing childhood vitiligo. Photodiagnosis and Photodynamic Therapy, 54: 104727. https://doi.org/10.1016/j.pdpdt.2025.104727
[22] Asheer, A., Shilpa. (2024). Vitiligo image categorization using convolution neural networks. Journal of Propulsion Technology, 45(1). https://www.propulsiontechjournal.com/index.php/journal/article/view/5020.
[23] Sharma, S., Guleria, K., Kumar, S., Tiwari, S. (2023). Deep learning based model for detection of vitiligo skin disease using pre-trained inception V3. International Journal of Mathematical, Engineering and Management Sciences, 8: 1024-1039. https://doi.org/10.33889/IJMEMS.2023.8.5.059
[24] Teixeira, B., Pinto, G., Filipe, V., Teixeira, A. (2024). Enhancing medical imaging through data augmentation: A review. In Computational Science and Its Applications-ICCSA 2024 Workshops, Cham, Switzerland: Springer, pp. 297-310. https://doi.org/10.1007/978-3-031-65223-3_23
[25] Al-Smadi, K., Imran, M., Leite-Silva, V.R., Mohammed, Y. (2023). Vitiligo: A review of aetiology, pathogenesis, treatment, and psychosocial impact. Cosmetics, 10(3): 84. https://doi.org/10.3390/cosmetics10030084
[26] Pawar, P., Hamid Khan, R.A. (2024). Enriched vitiligo detection using ML and DL approaches. In 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, pp. 1486-1490. https://doi.org/10.1109/ICSES63445.2024.10763103
[27] Pawar, P., Kulkarni, A., Pawar, P., Bhende, M., Bhavana, P. (2024). Enhancing inflammatory vitiligo detection with transfer learning techniques. In Emerging Technologies in AI, Computation, Communication, and Cybersecurity, CRC Press, p. 6. https://doi.org/10.1201/9781003739791-58
[28] Shaikh, S. (2025). Vitiligo images dataset. Kaggle. https://www.kaggle.com/datasets/shaikhshahid/vitiligo-images.
[29] Mendeley Data. (2025). Skin disease classification dataset. https://data.mendeley.com/datasets/3hckgznc67/1.
[30] Kaggle. (2025). Vitiligo. https://www.kaggle.com/datasets/shinynose/vitiligo.
[31] DermNet NZ. (2025). Vitiligo image gallery. https://dermnetnz.org/images/vitiligo-images.
[32] Kaggle. (2025). Vitiligo dataset. https://www.kaggle.com/datasets/zienabesam/vitiligo-dataset.
[33] Wang, J.C., Zhang, Y.L., Xie, F.Y., Liu, J. (2025). Enhancing diagnosis of psoriasis and inflammatory skin diseases: A spatially aligned multimodal model integrating clinical and dermoscopic images. Journal of Investigative Dermatology, 145(11): 2736-2744.e6. https://doi.org/10.1016/j.jid.2025.03.034
[34] Kaggle. (2024). Skin Disease Dataset. https://www.kaggle.com/datasets/pacificrm/skindisease dataset.
[35] Jia, H.M., Su, Y.Y., Rao, H.H., et al. (2025). Improved artificial rabbits algorithm for global optimization and multi-level thresholding color image segmentation. Artificial Intelligence Review, 58: 55. https://doi.org/10.1007/s10462-024-11035-3
[36] Bradley, D., Roth, G. (2007). Adaptive thresholding using the integral image. Journal of Graphics Tools, 12(2): 13-21. https://doi.org/10.1080/2151237X.2007.10129236
[37] Guo, Y.M., Wang, Y., Meng, K., Zhu, Z.N. (2023). Otsu multi-threshold image segmentation based on adaptive double-mutation differential evolution. Biomimetics, 8(5): 418. https://doi.org/10.3390/biomimetics8050418
[38] Liu, S.W., Zhou, B.Y., Lin, Y.H., Wang, P.X. (2024). Efficient and real-time skin lesion image segmentation using spatial-frequency information and channel convolutional networks. Journal of Real-Time Image Processing, 21: 165. https://doi.org/10.1007/s11554-024-01542-5
[39] Wan, Q., Su, H.B., Liu, X.Y., Yu, Y., Lin, Z.Z. (2025). SAM-enhanced cross-domain framework for semantic segmentation: Addressing edge detection and minor class recognition. Processes, 13(3): 736. https://doi.org/10.3390/pr13030736
[40] Mazzetto, R., Sernicola, A., Tartaglia, J., Ciolfi, C., Alaibac, M. (2025) Potential of automated image analysis for the measurement of vitiligo lesions. Frontiers in Medicine, 12: 1623408. https://doi.org/10.3389/fmed.2025.1623408
[41] Chen, Y.L., Lukic, T., Chen, A.S., et al. (2026). Automated measurement of depigmentation extent with a new AI tool applied to the example of vitiligo.

Dermatology and Therapy (Heidelberg), 16: 3139-3149. https://doi.org/10.1007/s13555-026-01736-8
[42] Wang, Y.H., Zheng, Y.H., Yue, C., et al. (2026). GloW-VSNet: A scribble-based weakly supervised framework for global-view vitiligo lesion segmentation. Medical Image Analysis, 109: 103920. https://doi.org/10.1016/j.media.2025.103920
[43] Li, Y.L., Thng, S.T.G., Kong, A.W.K. (2024). Bridging the gap between vitiligo segmentation and clinical scores. IEEE Journal of Biomedical and Health Informatics, 28(3): 1623-1634. https://doi.org/10.1109/JBHI.2023.3342069
[44] Li, Y.L., Kong, A.W., Thng, S. (2021). Segmenting vitiligo on clinical face images using CNN trained on synthetic and internet images. IEEE Journal of Biomedical and Health Informatics, 25(8): 3082-3093. https://doi.org/10.1109/JBHI.2021.3055213
[45] Park, S., Kim, J., Wang, S., Kim, J. (2025). Effectiveness of image augmentation techniques on non-protective personal equipment detection using YOLOv8. Applied Sciences, 15(5): 2631. https://doi.org/10.3390/app15052631
[46] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6: 60. https://doi.org/10.1186/s40537-019-0197-0
[47] Lama, N., Stanley, R.J., Lama, B., et al. (2024). LAMA: Lesion-aware mixup augmentation for skin lesion segmentation. Journal of Imaging Informatics in Medicine, 37(4): 1812-1823. https://doi.org/10.1007/s10278-024-01000-5
[48] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321: 321-331. https://doi.org/10.1016/j.neucom.2018.09.013
[49] Aleem, S., Kumar, T., Little, S., Bendechache, M., Brennan, R., McGuinness, K. (2022). Random data augmentation-based enhancement: A generalized enhancement approach for medical datasets. arXiv preprint arXiv:2210.00824. https://arxiv.org/abs/2210.00824
[50] Goceri, E. (2024). GAN based augmentation using a hybrid loss function for dermoscopy images. Artificial Intelligence Review, 57: 234. https://doi.org/10.1007/s10462-024-10897-x
[51] Wu, L., et al. (2022). Data augmentation based on multiple oversampling fusion for medical image segmentation. PLoS One, 17(10): e0274522. https://doi.org/10.1371/journal.pone.0274522
[52] Liu, X.Q., Ono, K., Bise, R. (2024). A data augmentation approach that ensures the reliability of foregrounds in medical image segmentation. Image and Vision Computing, 147: 105056. https://doi.org/10.1016/j.imavis.2024.105056
[53] Qi, X.Y., Zhang, Z.Y., Gang, C.X., et al. (2026). MediAug: Exploring visual augmentation in medical imaging. In Medical Image Understanding and Analysis, Cham, Switzerland: Springer, pp. 218-232. https://doi.org/10.1007/978-3-031-98688-8_16
[54] Zhong, F., He, K.Q., Ji, M.Q., et al. (2024). Optimizing vitiligo diagnosis with ResNet and Swin transformer deep learning models: A study on performance and interpretability. Scientific Reports, 14: 9127. https://doi.org/10.1038/s41598-024-59436-2
[55] He, K.Q., Xu, T.L., Feng, Y.N., et al. (2026). A two-stage workflow for vitiligo diagnosis: Clinical characteristic classification and large language model (LLM)-based report generation. Frontiers in Immunology, 17: 1853327. https://doi.org/10.3389/fimmu.2026.1853327
[56] Fan, Z.Z., Wang, C.Q. (2024). Deep learning-based segmentation of lesions from wide-field vitiligo images. International Journal of Pattern Recognition and Artificial Intelligence, 38(10): 2457007. https://doi.org/10.1142/S0218001424570076
[57] Gholijani, N., Taheri, Z., Bazgiri, M., Darya, G., Dehghan, Z. (2025). Transforming vitiligo diagnosis and treatment through artificial intelligence: A review. Scandinavian Journal of Immunology, 102(6): e70076. https://doi.org/10.1111/sji.70076
[58] UdriŞtoiu, A.L., Stanca, A.E., Ghenea, A.E., et al. (2020). Skin diseases classification using deep learning methods. Current Health Sciences Journal, 46(2): 136-140. https://doi.org/10.12865/CHSJ.46.02.06
[59] Luo, W., Liu, J., Huang, Y.D., Zhao, N. (2023). An effective vitiligo intelligent classification system. Journal of Ambient Intelligence and Humanized Computing, 14: 5479-5488. https://doi.org/10.1007/s12652-020-02357-5
[60] Hillmer, D., Merhi, R., Boniface, K., et al. (2023). Evaluation of facial vitiligo severity with a mixed clinical and artificial intelligence approach. Journal of Investigative Dermatology, 144: 351-357.e4. https://doi.org/10.1016/j.jid.2023.07.014
[61] Wang, Z., Hu, H., Ou, Y.Y., et al. (2025). Computer-aided assessment of repigmentation rates in vitiligo patients: Implications for treatment efficacy-A retrospective study. Journal of Investigative Dermatology, 145(1): 56-64.e5. https://doi.org/10.1016/j.jid.2024.05.016
[62] Brinker, T.J., Hekler, A., Enk, A.H., et al. (2019). Deep neural networks are superior to dermatologists in melanoma image classification. European Journal of Cancer, 119: 11-17. https://doi.org/10.1016/j.ejca.2019.05.023
[63] Biswas, R., Al Sohan, M.F.A., Rudro, R.A.M., Parvin, S. (2025). VGG19-ResSE: An optimized hybrid model for accurate segmentation and classification of vitiligo lesions. Neural Computing and Applications, 37(32): 27227-27248. https://doi.org/10.1007/s00521-025-11670-z
[64] Kantoria, V., Sharma, S., Bhushan, S., Saini, H., Nijhawan, R. (2020). Implication of convolutional neural network in the classification of vitiligo. International Research Journal of Engineering and Technology, 7(3): 5170-5176.
[65] Usman, M., Iqbal, M.Y., Zafar, K., Basharat, S. (2024). A novel approach to vitiligo diagnosis using artificial neural networks and dermatological image analysis. Journal of Computing & Biomedical Informatics, 8(1). https://jcbi.org/index.php/Main/article/view/736.
[66] Liu, Y., Jain, A., Eng, C., et al. (2020). A deep learning system for differential diagnosis of skin diseases. Nature Medicine, 26: 900-908. https://doi.org/10.1038/s41591-020-0842-3
[67] Vakalopoulou, M., Christodoulidis, S., Burgos, N., Colliot, O., Lepetit, V. (2023). Deep learning: Basics and

convolutional neural networks (CNNs). In Machine Learning for Brain Disorders, New York, NY, USA: Humana, pp. 77-115. https://doi.org/10.1007/978-1-0716-3195-9_3

[68] Montesinos López, O.A., Montesinos López, A., Crossa, J. (2022). Convolutional neural networks. In Multivariate Statistical Machine Learning Methods for Genomic Prediction, Cham, Switzerland: Springer, pp. 533-577. https://doi.org/10.1007/978-3-030-89010-0_13

[69] Gulzar, Y., Agarwal, S., Soomro, S., et al. (2025). Next-generation approach to skin disorder prediction employing hybrid deep transfer learning. Frontiers in Big Data, 8: 1503883. https://doi.org/10.3389/fdata.2025.1503883

[70] Abou Baker, N., Zengeler, N., Handmann, U. (2022). A transfer learning evaluation of deep neural networks for image classification. Machine Learning and Knowledge Extraction, 4(1): 22-41. https://doi.org/10.3390/make4010002

[71] Krichen, M. (2023). Convolutional neural networks: A survey. Computers, 12(8): 151. https://doi.org/10.3390/computers12080151

[72] Nijhawan, R., Mendirtta, N., Verma, M., Bohra, R., Kumar, S. (2024). Vitiligo detection using machine learning algorithms. In Intelligent Sustainable Systems, Singapore: Springer, pp. 33-42. https://doi.org/10.1007/978-981-99-8031-4_4

[73] Chauhan, A., Kumar, I. (2025). Deep feature extraction and optimized VGG16–SVM architecture for breast cancer characterization. Discover Computing, 28: 208. https://doi.org/10.1007/s10791-025-09736-6

[74] Sharma, S., Guleria, K. (2025). VGG16 and SqueezeNet-based ensembled models integrated with hybrid trio stacked classifiers for early clinical decision making. Journal of Transformative Technologies and Sustainable Development, 9: 16. https://doi.org/10.1007/s41314-025-00082-5

[75] Gwyn, T., Roy, K., Atay, M. (2021). Face recognition using popular deep net architectures: A brief comparative study. Future Internet, 13(7): 164. https://doi.org/10.3390/fi13070164

[76] He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

[77] Hasan, N., Bao, Y.K., Shawon, A., Huang, Y.M. (2021). DenseNet convolutional neural networks application for predicting COVID-19 using CT image. SN Computer Science, 2: 389. https://doi.org/10.1007/s42979-021-00782-7

[78] Zhou, T., Ye, X.Y., Lu, H.L., Zheng, X., Qiu, S., Liu, Y.C. (2022). Dense convolutional network and its application in medical image analysis. BioMed Research International, 2022: 2384830. https://doi.org/10.1155/2022/2384830

[79] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 2261-2269. https://doi.org/10.1109/CVPR.2017.243

[80] Howard, A.G., Zhu, M.L., Chen, B. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://doi.org/10.48550/arXiv.1704.04861

[81] Rybczak, M., Kozakiewicz, K. (2024). Deep machine learning of MobileNet, efficient, and inception models. Algorithms, 17(3): 96. https://doi.org/10.3390/a17030096

[82] Howard, A., Sandler, M., Chu, G., et al. (2019). Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, pp. 1314-1324.

[83] Alruwaili, M., Mohamed, M. (2025). An integrated deep learning model with EfficientNet and ResNet for accurate multi-class skin disease classification. Diagnostics, 15(5): 551. https://doi.org/10.3390/diagnostics15050551

[84] Tan, M.X., Le, Q. (2021). EfficientNetV2: Smaller models and faster training. In Proceedings of the 38th International Conference on Machine Learning, pp. 10096-10106. https://proceedings.mlr.press/v139/tan21a.html.

[85] Kallipolitis, A., Revelos, K., Maglogiannis, I. (2021). Ensembling EfficientNets for the classification and interpretation of histopathology images. Algorithms, 14(10): 278. https://doi.org/10.3390/a14100278

[86] Kandel, I., Castelli, M. (2020). How deeply to fine-tune a convolutional neural network: A case study using a histopathology dataset. Applied Sciences, 10(10): 3359. https://doi.org/10.3390/app10103359

[87] Becherer, N., Pecarina, J., Nykl, S., Hopkinson, K. (2019). Improving optimization of convolutional neural networks through parameter fine-tuning. Neural Computing and Applications, 31: 3469-3479. https://doi.org/10.1007/s00521-017-3285-0

[88] Sarhan, N., Lauri, M., Frintrop, S. (2022). Multi-phase fine-tuning: A new fine-tuning approach for sign language recognition. KI-Künstliche Intelligenz, 36: 91-98. https://doi.org/10.1007/s13218-021-00746-2

[89] Angriawan, M. (2023). Transfer learning strategies for fine-tuning pretrained convolutional neural networks in medical imaging. Research Journal of Computer Systems and Engineering, 4(2): 73-88.

[90] Iman, M., Arabnia, H.R., Rasheed, K. (2023). A review of deep transfer learning and recent advancements. Technologies, 11(2): 40. https://doi.org/10.3390/technologies11020040

[91] Kim, H.E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M.E., Gansland, T. (2022). Transfer learning for medical image classification: A literature review. BMC Medical Imaging, 22: 69. https://doi.org/10.1186/s12880-022-00793-7

[92] Duhaim, A.M., Baqer, N.S., Fadhel, M.A. (2026). Developing a trustworthy and explainable framework for classifying skin lesions through transfer learning and attention mechanisms. Computational Biology and Chemistry, 122: 108914. https://doi.org/10.1016/j.compbiolchem.2026.108914

[93] Nguyen, A.T.P., Shak, M.S., Al-Imran, M. (2024). Advancing early skin cancer detection: A comparative analysis of machine learning algorithms for melanoma diagnosis using dermoscopic images. International Journal of Medical Science and Public Health Research, 5(12): 119-133. https://doi.org/10.37547/ijmsphr/volume05issue12-10

[94] Gajera, H.K., Nayak, D.R., Zaveri, M.A. (2023). A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. Biomedical

Signal Processing and Control, 79(2): 104186. https://doi.org/10.1016/j.bspc.2022.104186

[95] Pasumarthy, R., Gondi, L. (2024). An extensive examination of methods for detecting and classifying malignant and benign skin cancers using deep learning techniques. In 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, pp. 1-7. https://doi.org/10.1109/ICDCECE60827.2024.10548394

[96] Varghese, A., Jain, A., Rahman, M.I., et al. (2025). Robust skin cancer detection through CNN-Transformer-GRU fusion and generative adversarial network-based data augmentation. CMES-Computer Modeling in Engineering and Sciences, 144(2): 1767-1791. https://doi.org/10.32604/cmes.2025.067999

[97] Meswal, H., Kumar, D., Gupta, A., Roy, S. (2024). A weighted ensemble transfer learning approach for melanoma classification from skin lesion images. Multimedia Tools and Applications, 83: 33615-33637. https://doi.org/10.1007/s11042-023-16783-y

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

[99] Bhimavarapu, U., Battineni, G. (2022). Skin lesion analysis for melanoma detection using the novel deep learning model fuzzy GC-SCNN. Healthcare, 10(5): 962. https://doi.org/10.3390/healthcare10050962

[100] Li, M., Jiang, Y.P., Cao, G., Xu, T., Guo, R.Q. (2026). HyperFusionNet combines vision transformer for early melanoma detection and precise lesion segmentation. Scientific Reports, 16: 505. https://doi.org/10.1038/s41598-025-30184-1

[101] Yao, C.D. (2023). A comprehensive evaluation study on risk level classification of melanoma by computer vision on ISIC 2016–2020 datasets. arXiv preprint arXiv:2302.09528. https://arxiv.org/abs/2302.09528

[102] Lee, S., Heo, Y.W., Lee, S., Park, S.B. (2024). Development of a deep learning-based automated analysis system for facial vitiligo treatment evaluation. Journal of Biomedical Engineering Research, 45: 95-100. https://koreascience.kr/article/JAKO202419757662512.page.

[103] MurarrySetty, T., Deepa, C., Babi, K., Bhargavi, B., Kumar, E.A. (2024). Precise detection of vitiligo using deep learning methodology. In 2024 International Conference on Recent Innovation in Smart and Sustainable Technology (ICRISST), Bengaluru, India, pp. 1-6. https://doi.org/10.1109/ICRISST59181.2024.10921796

[104] Wang, J.C., Ding, X.L., Xiao, J. (2023). Poisson-based image editing for semi-supervised vitiligo lesion segmentation with limited annotations. Computers in Biology and Medicine, 165: 107320. https://doi.org/10.1016/j.compbiomed.2023.107320

[105] Bian, Z.X., Xia, S.Y., Xia, C., Shao, M. (2019). Weakly supervised vitiligo segmentation in skin image through saliency propagation. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, pp. 931-934. https://doi.org/10.1109/BIBM47256.2019.8983145

[106] Low, M., Huang, V., Raina, P. (2020). Automating vitiligo skin lesion segmentation using convolutional neural networks. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, pp. 1-4. https://doi.org/10.1109/ISBI45749.2020.9098682

[107] Gopalakrishnan, B., Swamynathan, S., Padmashree, A., Siamala, S., Lathajothi, V., Sathishkannan, R. (2025). Vitiligo detection and progression analysis using vision transformers and self-attention networks. In 2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, pp. 1-8. https://doi.org/10.1109/NMITCON65824.2025.11188357

[108] Guo, L.F., Yang, Y., Ding, H., et al. (2022). A deep learning-based hybrid artificial intelligence model for the detection and severity assessment of vitiligo lesions. Annals of Translational Medicine, 10(10): 590. https://doi.org/10.21037/atm-22-1738

[109] Parikh, M., Fang, G., Poon, F., et al. (2025). Technological advances in vitiligo management: Perspectives on AI, mobile tools, and clinical utility. Frontiers in Medicine, 12: 1661554. https://doi.org/10.3389/fmed.2025.1661554

[110] Khatibi, T., Rezaei, N., Ataei Fashtami, L., Totonchi, M. (2021). Proposing a novel unsupervised stack ensemble of deep and conventional image segmentation (SEDCIS) method for localizing vitiligo lesions in skin images. Skin Research and Technology, 27(2): 126-137

[111] Tanvir, S., Syed, S.A., Hussain, S., Zia, R., Rashid, M., Zahid, H. (2024). Detection of vitiligo through machine learning and computer-aided techniques: A systematic review. BioMed Research International, 2024: 3277546. https://doi.org/10.1155/bmri/3277546