Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Review Article

AI-Based Detection Techniques for Skin Diseases: A Review of Recent Methods, Datasets, Metrics, and Challenges

Oluwayemisi Jaiyeoba [1,*], Oluwaseyi Jaiyeoba [2], Emeka Ogbuju [1,3], and Francisca Oladipo [4]

[1] Department of Computer Science, Federal University Lokoja, Lokoja, Kogi State, Nigeria; e-mail: oluwayemisi.jaiyeoba@fulokoja.edu.ng
[2] Department of Computer Graphics Technology, Purdue University, West Lafayette, Indiana, United States; e-mail: ojaiyeob@purdue.edu
[3] Department of Computer Science, Miva Open University, Federal Capital Territory, Abuja, Nigeria; e-mail: emeka.ogbuju@fulokoja.edu.ng
[4] Department of Computer Science, Thomas Adewumi University, Oko, Kwara State, Nigeria; e-mail: francisca.oladipo@tau.edu.ng
* Corresponding Author: Oluwayemisi Jaiyeoba

Abstract: The identification and early treatment of skin diseases are crucial to mitigate serious health risks. The growing attention on researching skin disease analysis stems from the transformative impact of artificial intelligence (AI) in dermatology. In this systematic review, we adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to comprehensively assess recent approaches for skin disease detection. Our study addressed four key research questions exploring the methods for skin disease detection, the evaluation techniques employed to measure the effectiveness of skin disease detection models, the datasets utilized, and the challenges encountered in applying machine learning and deep learning techniques for skin disease detection. We screened studies from 2019 to 2023 from reputable databases, including IEEE Explore, Science Direct, and Google Scholar. Our findings revealed that the CNN model outperformed other deep learning models. Additionally, our analysis identified the ISIC public dataset as the most frequently used dataset. The studies reviewed employed evaluation metrics such as accuracy, recall, precision, sensitivity, and F1 score to evaluate model performance. We identified several limitations in the studies we reviewed, including the use of limited datasets, challenges in distinguishing between diseases with similar features, and other related limitations. Overall, we provided a comprehensive overview of the current state-of-the-art techniques in skin disease detection and highlighted the future directions.

Keywords: Artificial Intelligence; Convolutional Neural Network; Deep Learning; Detection; Machine Learning.

Received: October, 22nd 2024
Revised: December, 20th 2024
Accepted: December, 21st 2024
Published: December, 28th 2024
Curr. Ver.: December, 28th 2024

Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

The skin is the body's largest organ, serving as a vital protective barrier that not only provides a physical shield against external threats but also helps keep things in balance and controls how the body works [1]. Skin types are of various colors, ranging from dark brown to various lighter shades. These color differences mainly happen because of a pigment called melanin, made by special skin cells [2]. Figure 1 shows that the skin comprises three main layers: the epidermis, dermis, and hypodermis. The epidermis, the outermost layer of the skin, protects against the external environment and contains melanin produced by cells beneath it. The dermis consists of robust connective tissue, while the hypodermis is located beneath it and primarily contains fat tissue [3]. The skin is affected by various factors such as sun ultraviolet radiation, smoking, alcohol usage [4], environmental conditions, bacteria and viruses, a weak immune system, fungi, heat rash, and dark spots [5]. The skin is so noticeable and important it's prone to many different issues, such as skin diseases.

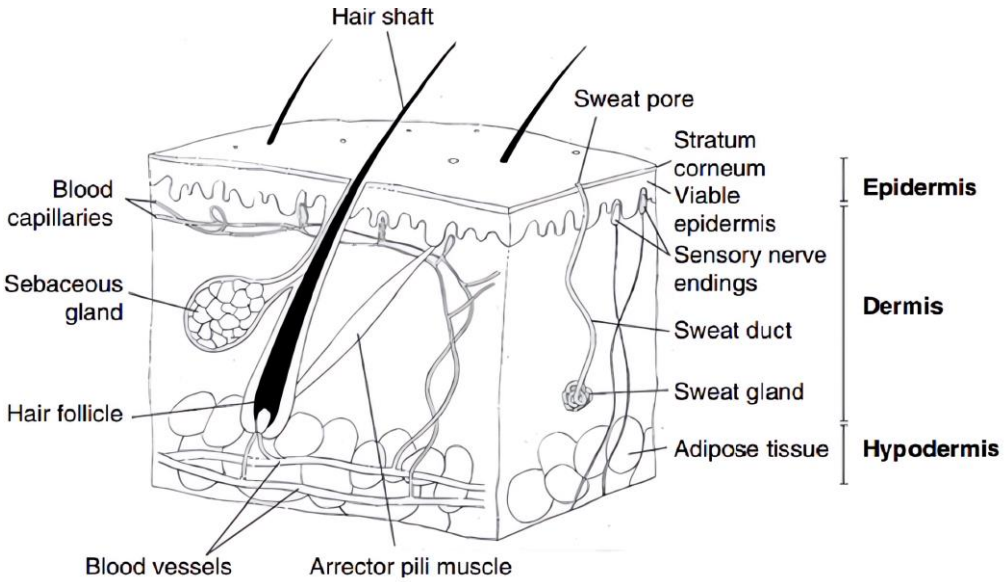

Figure 1. A diagrammatic representation of the structure of human skin in cross-section. Showing the Three Primary Skin Layers: Epidermis, Dermis, and Hypodermis [6]

Skin diseases occur commonly among humans, covering a wide range of problems that can affect how the skin looks, works, and feels [7]. Skin diseases are becoming increasingly common due to various factors, especially environmental pollution, including pollutants like polycyclic aromatic hydrocarbons and other harmful chemicals that can damage the skin barrier. This damage can lead to a range of problems, including skin cancer [8]. Conditions like acne, livedoid vasculopathy, and pemphigus vulgaris are often triggered by factors such as excessive sebum production, hormonal imbalances (particularly insulin-like growth factor 1 and androgens), diet (such as high dairy and meat intake), hyperkeratinization, inflammation, and environmental factors [9].

Skin diseases are usually common and can affect individuals belonging to different age groups, and it can also lead individuals to feel insecure as they may be perceived and judged by others based on their skin color. [10]. Skin diseases have diverse conditions, each with unique characteristics and symptoms. Some common skin diseases include dermatitis exfoliative generalisata, impetigo, pityriasis rosea, erisipelas, eczema, psoriasis, acne, warts, vitiligo, tinea corporis, scabies, hives, rosacea, shingles, boils, cold sores, corns, calluses, dyshidrotic dermatitis, molluscum contagiosum, neurofibromatosis, skin tags, melanoma(MEL), rashes, malignant melanoma – squamous cell carcinoma (SCC), basal cell carcinoma (BCC), genetic diseases - genetic skin disorders, sickle cell disease anemia, leprosy, viral infections, seborrheic dermatitis, lichen planus, pink pityriasis, chronic dermatitis, Pityriasis Rubra Pilaris, herpes, seborrheic keratosis (SK), bullae, splints, venous malformations, and scleroderma, among others. Approximately 75% of skin cancer patients die worldwide each year. Early skin disease detection is crucial as it enables prompt intervention to eradicate the illness from the body completely. Failure to detect it early can result in the cancer becoming a big problem, making it almost impossible to treat [11]. Skin diseases are frequently overlooked and left untreated by people, particularly in rural and economically disadvantaged populations. [12]. This is primarily attributable to factors, including limited awareness, financial constraints, and a lack of access to resources [13]. Currently, the diagnosis of skin diseases often involves invasive techniques like clinical examinations and histological procedures, where doctors use the biopsy method to detect skin diseases. This procedure involves taking a sample of the affected area for medical examination in the lab [14]. The procedure is painful and time-consuming. However, when computer-based technology is introduced, it will provide a comfortable and less time-consuming procedure for detecting skin diseases [15]. One of the most important developments has been the application of artificial intelligence (AI) and machine learning (ML) techniques to assist in the early diagnosis and management of skin diseases. AI can analyze extensive datasets to identify timely, pertinent, and knowledge-based information, allowing individuals to make informed choices regarding their health [16]. Working together,

healthcare and technology have created new ways for innovative approaches to skin disease detection, offering healthcare professionals, dermatologists, and patients' new tools and strategies to address these conditions more effectively. In this study, we thoroughly review recent advancements in detecting skin diseases from 2019 to 2023. We began with an overview of skin diseases and then went into reviewing the deep learning models, the dataset used, and the accuracies achieved. We analyzed the performance of each model and discussed the evaluation metrics commonly employed, particularly in imbalanced datasets, and metrics that are most suitable for skin disease detection. While acknowledging the limitations of the current approaches, we also explore potential future research directions that could further enhance the detection of skin diseases.

Previous studies, such as [17], conducted a systematic review of machine learning techniques for melanoma skin cancer classification from 2019–2023, focusing on analyzing the different algorithms, datasets, evaluation metrics, and challenges. Another review by [18] focused on reviewing studies that worked on detecting and classifying skin diseases using models like ANN, CNN, K-neural networks, and Radial Basis Function Networks from 2011–2021, examining their strengths and limitations. While these reviews provided valuable insights, our study goes further by comprehensively analyzing models, datasets, results, limitations, and evaluation metrics of recent advancements from 2019–2023, making our review more current and comprehensive.

1.1. Research Questions

- Research Question 1: What are the most commonly used methods for skin disease detection?
- Research Question 2: What evaluation metrics are commonly used to assess the performance of skin disease detection models?
- Research Question 3: What commonly available datasets focus on skin disease detection?
- Research Question 4: What are the challenges faced in applying machine learning and deep learning techniques to detect skin diseases?

1.2. Contribution

1. An overview of skin problems is provided, data collection techniques used, and the freely accessible datasets for skin disease detection.
2. Conducts a thorough review of research articles published between 2019 and 2023, ensuring the inclusion of the latest advancements in detecting skin diseases.
3. Explores how emerging technologies, such as machine learning and deep learning, are used in recent skin disease detection studies.
4. Analyzes the various datasets used in recent skin disease detection studies.

The remaining sections of the paper are organized as follows: Section 2 discusses the related work done in detecting skin diseases; Section 3 presents the methodology used in performing the systematic literature review; Section 4 presents the findings of the systematic literature review in a summary Table 1, and Section 5 presents a comprehensive discussion of the research questions, providing a thorough exploration and analysis of the key findings guiding the systematic literature review. Finally, Section 6 presents the conclusion and recommendations for future research in the systematic review paper.

2. Literature Review

Study [19] proposed an ensemble model that combines three different CNN architectures: Visual Geometry Group (VGG16), DenseNet, and Inception for detecting skin diseases such as Actinic Keratoses (AK), BCC, Benign Keratosis (BK), Dermatofibroma (DF), Melanoma, Melanocytic Nevi (NV), and Vascular Lesions (VASC) using the HAM10000 dataset. The ensemble model combined three CNN architectures, namely VGG16, DenseNet, and Inception, to leverage the strengths of each network. The individual CNN models achieved an accuracy of 80.3% using VGG16, 82.3% using DenseNet, and 80.4%. Using the Inception model and an ensemble of these models resulted in an impressive accuracy of 85.02%.

Study [20] conducted research by proposing the application of machine learning algorithms, including Random Forest (RF), Naive Bayes, logistic regression, kernel support vector

machine (kernel SVM), and CNN, for detecting skin diseases. They preprocessed a skin image to detect acne, lichen planus, and Stevens-Johnson Syndrome (SJS) TEN skin disease. The testing data was divided into three classes (class a, b, and c) corresponding to these diseases. Among the algorithms tested, CNN achieved the highest testing accuracy of 96% with a low error rate of 0.04.

Study [21] proposed a targeted ensemble machine classification model" (TEMCM) for skin disease detection. The evaluation process involved testing various data processing methods and aggregations, including training and cross-testing validations on seven skin conditions: AK, BCC, BK, DF, MEL, NV, and VASC. They used the HAM10000 dataset and focused on six well-known deep learning models, namely VGG16, Inception, Xception, MobileNet, ResNet50, and DenseNet161. The TEMCM model achieved an accuracy rate of approximately 98.48%.

Study [22] proposed an adaptive ensemble CNN using the International Skin Imaging Collaboration (ISIC) 2019 dataset to detect DF, VASC, SSC, and unknown skin conditions. The model was evaluated with an accuracy of 95.6% and precision and recall of 0.95 for the first four classes. However, its performance suffered when the model was extended to incorporate four new classes that were not part of its initial training. In the first experiment, the classification accuracy dropped after adapting to the new classes. In the second experiment, the proposed framework achieved an 89% classification accuracy at the cloud level but encountered increased loss after adapting to the new classes.

A study [23] proposed a mobile-based system for the early detection and classification of skin diseases, such as MEL, eczema, and impetigo, containing 408 images. The system used two different approaches, namely SVM and CNN, for the detection and classification tasks. The researchers employed 2D Wavelet Transform and Gray-Level Co-occurrence Matrix (GLCM) methods to extract distinctive features from skin lesions. The model's performance was evaluated, and the support vector machine (SVM) classifier achieved an accuracy of 90.7%, while the CNN classifier achieved an accuracy of 99.1%.

A study [24] proposed a model for detecting skin diseases using CNN implemented with the Keras Sequential API. To improve the accuracy, they employed transfer learning techniques using pre-trained models such as VGG11, RESNET50, and DENSENET121. The dataset used was the HAM10000 dataset, which consisted of 10015 images representing various skin conditions, including AK, BCC, BK-like lesions, DF, MEL, NV, and VASC. The data was normalized and augmented. The evaluation results showed CNN achieving an accuracy of 79%, ResNet50 and DenseNet121 achieving 90%, and VGG11 with Batch Normalization (VGG11 BN) achieving 85% accuracy.

A study [25] proposed an innovative MEL localization and segmentation approach using a combination of YOLOv4 and active contour segmentation on dermoscopic images. They used morphological operations to effectively remove artifacts like hairs, gel bubbles, and clinical marks from the images. The proposed method was evaluated on two datasets, ISIC2018 and ISIC2016, using dice score and Jaccard coefficient as evaluation metrics. They achieved an average dice score of 1 and a Jaccard coefficient of 0.989.

A study [26] proposed an SVM for detecting melanoma skin cancer at an early stage. Their study used a dataset containing 5341 images from the ISIC, consisting of both MEL and non-MEL cases. They extracted features related to asymmetric behavior, color, and border irregularity from the images to identify key indicators. The evaluation of their SVM-based model gave an accuracy of 96.9%, specificity of 90.2%, and sensitivity of 95.7%.

A study [27] proposed a super-resolution framework to address the challenge of limited access to high-resolution skin disease images for skin cancer detection. They designed a new loss function to enhance the generation of HR images. To compute the P-loss function, they used the VGG-19 pre-trained model without retraining. For training the SR models, such as Enhanced Super-Resolution Generative Adversarial Network (ESRGAN), Enhanced Deep Super-Resolution Generative Adversarial Network (EDSRGAN), and Super-Resolution Generative Adversarial Network (SRGAN), they used the HAM10000 dataset, which contains 10,015 dermoscopy images of various skin conditions, including NV, MEL, BK, BCC, AK, vascular, and DF images. During the preprocessing stage, the evaluation yielded a classification accuracy of 98.9958% when tested on image classification with a pre-trained EfficientNet-B0 model.

A study [28] proposed a CNN for detecting healthy skin and skin suffering from NV, MEL, BK-like lesions, BCC, AK, VASC, and DF. To assess the model's effectiveness, they

used the HAM10000 dataset, and the model demonstrated remarkable consistency in performance throughout a range of epochs, from Epoch 2 to Epoch 50, with an accuracy of 93.35% and a comparatively low loss of about 15.65% across both the testing and validation datasets.

A study [29] proposed a new framework for automated skin cancer detection involving two stages: a fully Convolutional Network (FCN) for lesion feature learning with skip connections and a conditional random field module for contour refinement. The second stage used a DenseNet-based FCN for lesion classification, with hyperparameter optimization for efficiency. The training dataset, HAM10000, contained 10,030 images of various skin diseases, including AKIEC, BCC, BKL, DF, MEL, NV, and VASC. The test data was obtained from two different datasets, ISBI 2018 and Pigmented Lesion (PH2) Dataset, consisting of 200 dermoscopic images of NV, including common nevi, atypical nevi, and MEL. Data augmentation techniques were applied to both the segmentation and classification datasets, resulting in the model achieving an impressive 98% accuracy, 98.5% recall, and 99% AUC score..

A study [30] proposed a deep CNN model to classify ML images as benign or malignant. The dataset used for training and testing was derived from the HAM10000 dataset, ISIC 2018 Challenge, HAM10000, Benign vs. Malignant, and PH2. The dataset was split into an 8:2 ratio for training and testing to ensure an unbiased evaluation. The images were preprocessed, which involved rescaling and proper labeling, with' 0' representing benign and' 1' representing malignant cases. They also developed a user-friendly web system with a graphical user interface (GUI) using the Website Frontend Bootstrap and Cascading Style Sheets (CSS) framework. The model achieved an accuracy of approximately 70%.

A study [31] proposed a deep learning-based CNN for classifying skin diseases, specifically focusing on detecting MEL skin cancer. They collected a dataset comprising 2102 dermatoscopic images, each accompanied by metadata. The dataset covered various skin conditions, including MEL, NV, BCC, BK-like lesions, VASC, AK, and DF. To train and validate their model, they used Kaggle's kernels. The optimization process during training was carried out using the Adam optimizer, running for 50 epochs, and the model achieved an accuracy of approximately 89.45% during evaluation.

Study [32] proposed a hybrid feature-based model that integrated local binary pattern (LBP), GLCM, and wavelet transform (DWT) features for early detection of skin lesions. The model employed both artificial neural network (ANN) and feed-forward neural network (FFNN) classifiers on two datasets, ISIC 2018 and PH2, containing a total of 1400 images with various skin lesion types, including AKs, BCC, BK lesions, DF, MEL, NV, and VASC. The evaluation of the proposed system yielded an accuracy of 95.24% for the ISIC 2018 dataset and 97.91% for the PH2 dataset using the hybrid feature-based approach with FFNN and ANN classifiers, and 90% for the ISIC 2018 dataset and 95.8% for the PH2 dataset using the CNN-based approach with the ResNet-50 model..

Study [33] proposed an innovative deep clustering approach to detect MEL in dermoscopic images. They conducted experiments using three datasets: ISIC2020, ISIC2019, and HAM10000, totaling 58,341 images. One of the significant challenges they tackled was the class imbalance in the data. Their method, SDCCOM Triplet, demonstrated superior performance in supervised settings compared to direct classifiers, achieving notable improvements of 1.21% on ISIC2019 and 0.75% on HAM10000 in terms of Area Under the Curve (AUC). Moreover, in unsupervised settings, their method, SDCCOM-Triplet, outperformed other approaches with an average improvement of 11.24% on ISIC2019 and 2.18% on HAM10000. These results suggest that their deep clustering approach has great potential in aiding the early detection of MEL from dermoscopic images, even in class imbalance.

Study [34] proposed a CNN model for classifying various skin diseases using input images of different skin conditions, such as dermatitis hand, eczema hand, eczema subacute, lichen simplex, stasis dermatitis, and ulcers. They collected a comprehensive dataset comprising over 500 images from Dermnet to develop and evaluate the model. The research process involved two distinct stages: the training stage, where normal picture features were extracted and used to train the CNN model, and the testing stage, where the previously extracted feature representations were utilized to classify the image features accurately. The dataset was enhanced through augmentation, image compression, and feature extraction techniques. The model achieved a precision rate of 73%.

A study [19] developed a deep learning model for skin disease detection using CNN and the Keras Sequential API. They employed transfer learning with pre-trained models like VGG11, ResNet50, and DenseNet121 on the HAM10000 dataset of 10,015 images. After

resizing images to 100 x 75 x 3 dimensions, they split the data into 80% for training and 20% for testing, ensuring normalization. Data augmentation techniques were applied for enhanced generalization. The CNN model achieved 79% validation and approximately 76% test accuracy. Pre-trained models (VGG11, ResNet50, DenseNet121) trained on ImageNet data significantly improved results, achieving 90% accuracy in training with minimal loss.

A study [35] proposed a skin cancer detection model using a modified EfficientNetV2-M deep neural network. They used transfer learning and fine-tuning techniques; data augmentation methods were used to improve model performance. The study used three datasets, HAM1000, ISIC 2019, and ISIC 2020, comprising 58,032 dermoscopic images of various skin conditions. Preprocessing and augmentation techniques, such as blurring, distortion, resizing, and normalization, were applied to enhance data quality and distribution. The EfficientNetV2-M model achieved a training accuracy of 94.10% and a minimal validation loss of 0.34 by the 11th epoch. Also, EfficientNet-B4 attained a validation loss of 0.41 by the 16th epoch, with a training accuracy of 89.92%. EfficientNetV2-an accuracy of 97.62%, 95.49%, and 94.8% on the ISIC 2020, ISIC 2019, and HAM10000 datasets, respectively. Micro average F1 scores were achieved from 0.95 to 0.98. For binary-level classification, EfficientNetV2-M achieved peak accuracy and AUC scores of 99.23%, 97.06%, and 95.95% and 0.98, 0.99, and 0.98 AUC scores on the ISIC 2020, ISIC 2019, and HAM10000 datasets, respectively.

A study [36] proposed a pre-trained CNN to extract distinctive features from color images depicting affected skin regions. These extracted features were inputs for a multi-class SVM model, enabling the accurate classification of diverse skin ailments such as eczema, MEL, psoriasis, and healthy skin. The dataset was compiled by collecting images from various websites specific to skin diseases. They achieved a 100% detection rate for three different skin diseases using their MATLAB-based system, which uses pre-trained CNNs for feature extraction and SVM classification on a dataset of 100 skin images.

A study [37] proposed an automated skin disease prediction and classification approach by combining image processing techniques and machine learning classifiers. Which involves digital hair removal using morphological filtering, Gaussian filtering for image enhancement, and automatic GrabCut segmentation to isolate affected areas. They also used the GLCM and statistical features techniques to extract meaningful patterns from the skin images. The classification system used Decision Tree (DT), SVM, and K-Nearest Neighbor (KNN) algorithms and achieved an accuracy of 95%, 94%, and 93% on the ISIC 2019 dataset and 97%, 95%, and 95% on the HAM10000 dataset, respectively.

A study [38] proposed a CNN model for detecting MEL, an aggressive type of skin cancer, employing a user-friendly camera for image acquisition. The data was preprocessed to adjust intensity levels and remove noise, as well as segmentation and feature extraction. They used the DermIS and DermQuest datasets to validate their approach, each comprising 22 images of benign and malignant skin lesions. The preprocessing stage encompassed filtering and edge detection procedures, while convolution was employed to mitigate noise before extracting highly perceptive features like color, irregularity, boundaries, shape, and diameter. The model was evaluated and achieved an efficiency of 95% with a sensitivity of 93.3%.

Study [3] proposed a neural network approach that combined an ANN classifier with hybrid texture features to effectively identify and categorize MEL lesions as benign or malignant, using a dataset comprising 1940 benign and 1448 malignant lesion images collected from the ISIC archive and the PH2 dermoscopic image database. They integrated the Maximum Gradient Intensity algorithm to eliminate hairs and enhance the images, and they performed segmentation through the Otsu Thresholding algorithm. The evaluation employed metrics such as accuracy, sensitivity, specificity, and precision, revealing a classification accuracy of 97.7%, 97.4% sensitivity, 97.9% specificity, and 97.2% precision for the image dataset. They used the mean square error and ROC curves to assess the system's overall performance.

Study [39] proposed a novel dynamic graph cut algorithm for precise skin lesion segmentation and employed the Naïve Bayes classifier for skin disease classification. Their method was evaluated using the ISIC 2017 dataset, which encompasses a variety of skin conditions like eczema, psoriasis, vitiligo, MEL, carcinoma, and blue nevus, providing a comprehensive testing platform. Image segmentation was employed, Gaussian filtering was employed to suppress noise, and HSV color space is preferred for skin detection due to its perceptual uniformity also. Asymmetry, color, and texture features were extracted from segmented skin regions. They achieved an accuracy of 94.3% for benign cases, 91.2% for MEL, and 92.9% for keratosis.

A study [40] proposed an inventive method to identify conjunctival MEL using a small set of ocular images. They used low-shot deep learning models, including GoogleNet, InceptionV3, Neural Architecture Search Network, ResNet50, and MobileNetV2. They used GANs to create synthetic images, broadening their training dataset with images from conjunctival MEL, nevus, pterygium, and normal conjunctiva categories. 3D MEL models were also used for external validation, captured through smartphone cameras. With just 136 conjunctival MEL images, 93 nevus images, 75 pterygium images, and 94 normal conjunctival images, MobileNetV2 with GAN augmentation achieved a high accuracy of 87.5% for the four-class classification and an impressive 97.2% for binary conjunctival MEL detection. External validation with 3D MEL phantom images yielded 94.0% accuracy..

Study [4] proposes a multi-class, multi-level (MCML) classification algorithm to address the challenges of accurately classifying multiple skin diseases. They classified skin diseases such as MEL, healthy, eczema, and erythematous-squamous diseases from the ISIC Dermoscopic Archive and the PH2 dataset of a total of 3672 images. The algorithm involves various preprocessing steps, including hair removal, black frame removal, circle removal, and segmentation. Features are then extracted from the segmented images, and a feed-forward neural network is used for classification. The model was evaluated, and the MCML achieved an accuracy of 96.47%.

A study [41] proposed a multi-class support vector machine (MSVM) for detecting and classifying various skin cancers, including MEL, BCC, and SCC. The images underwent different processing steps, including removing unwanted hair particles using the dull razor method, Gaussian filtering for image smoothing, and median filtering for noise reduction while preserving lesion edges. The project utilizes the ABCD (Asymmetry, Border, Color, Diameter) method and GLCM for statistical and texture feature extraction on the ISIC 2019 Challenge dataset to extract features for analysis. The model was evaluated and archived with an accuracy of 96.25%. and 96. % precision.

A study [42] proposed a melanoma skin cancer detection model combining CNN, KNN, and SVM prediction methods. They trained these models using features describing skin lesion borders, texture, and color. They also used a hair removal approach using color enhancement via the blue channel of red, blue, and green color (RGB). Hair detection involved techniques like 2-D derivatives of Gaussian on the blue channel, followed by thresholding for hair separation. Their dataset included over 23,000 melanoma images from ISIC, with 640 chosen for analysis. Images were collected through dermatoscopy, 512 for training, and the rest for testing. The CNN was trained for 10 epochs, achieving an accuracy of 57.3% with KNN, 71.8% with SVM, and 85.5% with CNN

Study [43] proposed a model for MEL detection and classification. Their model was a CNN, used for image preprocessing, incorporating various filters. They used clustering algorithms for skin lesion segmentation, specifically K-means and fuzzy C-means. Also, feature extraction was performed, and the skin lesions were classified using another CNN. This study used the ISIC archives dataset, which comprised approximately 23,906 images across various classes, including benign and MEL categories. The model's performance was evaluated with an accuracy of 98.43%, an impressive specificity of 98.77%, and an outstanding sensitivity of 99.83%.

Study [44] proposed a deep learning architecture called" FcResNet" for skin lesion classification using the ISIC-2018 challenge dataset, which consists of seven distinct classes, including AK, BCC, BK, DF, MEL, NV, and VASC, with 327, 514, 1099, 115, 6705, and 142 number of images respectively. The authors employed upsampling and data augmentation methods to enhance dataset diversity. The model achieved an accuracy of 92.34% on the HAM10000 dataset, while an enhanced version featuring transition layers yielded a higher accuracy of 96.07%.

A study [45] proposed a CNN for skin cancer detection. They sourced their dataset, originally from the ISIC, and split it into training and testing sets. Their model's task was to categorize images as benign or malignant skin cancer. The dataset comprised 1,906 training images and 816 testing images, making up 3,000. Their CNN achieved an impressive accuracy rate of 91% in identifying the disease.

A study [46] proposed an automated melanoma classification framework utilizing a deep convolutional neural network to classify malignant and benign melanoma skin lesions. The study utilized dermoscopic images from the International Skin Imaging Collaboration datasets, including ISIC 2016, ISIC 2017, and ISIC 2020. The model was evaluated and achieved

an accuracy of 81.41%, 88.23%, and 90.42% on the ISIC 2016, 2017, and 2020 datasets, respectively.

Study [47] proposed using deep learning algorithms, specifically CNN and RNN, for the classification and image recognition of various skin diseases. They started by preparing the data, converting images into the RGB format, and then cleaning and segmenting them to isolate the important regions. To streamline the analysis, they also converted the images to grayscale. They employed a feature selection process to pick out the most relevant data for their model carefully, and they finally integrated RNNs into their framework to effectively process sequential or textual data, resulting in improved accuracy when detecting skin diseases.

A study [48] proposed a novel method called Random Forest Deep Convolutional Neural Networks (RF-DCNN) Classifier for effectively categorizing skin disorders based on dermatoscopic images. Their research employed the HAM10000 dataset, which consists of images representing seven distinct types of skin conditions. They successfully solved the challenge of imbalanced class distribution by implementing data augmentation techniques. The model consists of various stages, including preprocessing, feature extraction, and classification. By combining the RF-DCNN, they achieved an accuracy rate of 96.1%.

A study [49] developed a computer-assisted system aimed at diagnosing skin cancer, particularly distinguishing between malignant and non-malignant melanomas. The system applied digital image processing techniques and ANNs for its analysis. The research involved the collection of skin cancer images, followed by thorough preprocessing steps that included noise reduction through a median filter. To segment the images effectively, they employed the maximum entropy thresholding method. Feature extraction was carried out using 2-D wavelet transformation. The paper reported varying accuracies, with the highest achieved accuracy of 90% using a specific backpropagation neural network configuration.

Study [50] proposed using the K-nearest neighbor (KNN) classifier and Gray Level Co-occurrence Matrix (GLCM) to classify benign and malignant skin cancer types using the ISIC dataset. The dataset contained variations of 3297, 1649, 825, and 210 images. The study shows that 480 experiments were conducted, testing different KNN parameters, including the number of neighbors (k) and distance (d) for angles 0, 45, 90, and 135.

This section must contain a state-of-the-art explanation. This can be achieved in several ways. First, discussing various related papers, including their objects, methods, and results, while highlighting gaps or differences compared to the current research. Another approach is integrating theory with related literature and elaborating on each theory in separate sub-chapters.

Table 1. Summary of Related Works

Ref	Methods	Dataset	Accuracy
[19]	VGG16, DenseNet, and Inception	HAM10000 dataset	VGG16: 80.3%, DenseNet: 82.3%, Inception:80.4%
[20]	RF, NB, LR, Kernel SVM, and CNN		CNN: 96%
[21]	TEMCM model	HAM10000 dataset	98.48%
[22]	Ensemble CNN	ISIC 2019 dataset	95.6%,
[23]	SVM and CNN		SVM: 90.7%, CNN: 99.1%.
[24]	CNN & pre-trained models: VGG11, RESNET50, and DENSENET121	The HAM10000 dataset	CNN: 79%, ResNet50 and DenseNet121: 90%, VGG11: 85%
[25]	YOLOv4 + active contour segmentation	ISIC2018 and ISIC2016	dice score: 1 and a Jaccard coefficient: 0.989.
[26]	SVM	ISIC Dataset	Accuracy 96.9%, specificity: 90.2% & sensitivity: 95.7%.
[27]	Enhanced Deep Super-Resolution Generative Adversarial Network (EDSR-GAN).	The HAM10000 dataset	98.9%
[28]	CNN	The HAM10000 dataset	93.35% & low loss: 15.65%

Ref	Methods	Dataset	Accuracy
[29]	First stage: FCN & Conditional Random field Second stage: DenseNet-based FCN & and hyper-parameter optimization	The HAM10000 dataset, ISBI 2018 & PH2	98% recall: 98.5%, & AUC 99%
[30]	Deep CNN	The HAM10000 dataset, ISIC 2018, &ISIC 2018	70%
[31]	deep learning-based CNN	Kaggle's dataset	89.45%
[32]	ANN & FFNN, CNN + ResNet-50	ISIC 2018 & PH2 dataset	95.24% & .91% for FFNN& ANN, & 90% & 95.8% for CNN& ResNet-50
[33]	SDCCOM-Triplet	ISIC2020, ISIC2019, and HAM10000	
[34]	CNN	Dermnet	Precision rate: 73%.
[24]	CNN & Keras Sequential API like VGG11, pre-trained model (ResNet50, and DenseNet121)	The HAM10000 dataset	CNN: 76% & pre-trained model: 90%
[35]	EfficientNetV2-M deep neural network	HAM1000, ISIC 2019, and ISIC 2020	Multi-class Skin Lesion Classification: ISIC 2020: 97.62% ISIC 2019: 95.49% HAM10000: 94.80% Binary Classification (Skin Lesion): ISIC 2020: 99.23% ISIC 2019: 97.06% HAM10000: 95.95%
[36]	SVM	websites that are specific to skin diseases	Accuracy: 100%
[37]	DT, SVM & KNN	ISIC 2019 dataset & The HAM10000 dataset	95%, 94%, and 93% on the ISIC 2019 dataset, and 97%, 95%, and 95% on the HAM10000 dataset
[38]	CNN	DermIS and DermQuest datasets	Identification Efficiency (IE) of 95% with a sensitivity of 93.3%.
[3]	Neural Network	ISIC archive & PH2	ACC: 97.7%, sensitivity: 97.4%, specificity: 97.9% & precision: 97.2%
[39]	Naïve Bayes	ISIC 2017 dataset	94.3%
[40]	GoogleNet, InceptionV3, Neural Architecture Search Network, ResNet50, and MobileNetV2.	A publicly available dataset	97.2%
[4]	Multi-Class Multi-Level	ISIC Archive and PH2 dataset	96.47%.
[41]	Multi-class Support Vector Machine	ISIC 2019 Challenge dataset	Accuracy: 96.25%, precision:96.32%.
[42]	CNN, KNN, SVM	ISIC dataset	KNN: 57.3% SVM: 71.8% CNN: 85.5%
[43]	CNN	ISIC archives dataset	Accuracy: 98.43%, specificity:98.77%, sensitivity: 99.83%.
[50]	KNN	ISIC dataset	76%
[51]	FcResNet	ISIC-2018 Challenge dataset	96.07%
[45]	CNN	ISIC dataset	91%

Ref	Methods	Dataset	Accuracy
[42]	DCNN	ISIC 2016, ISIC2017, and ISIC 2020	ISIC2016: 81.41%, ISIC 2017: 88.23%, and ISIC 2020: 90.42%
[47]	CNN & RNN	The HAM10000 datasets	
[48]	Random Forest Deep Convolutional Neural Networks (RF-DCNN)	The HAM10000 dataset	96.1%
[49]	Digital Image Processing (DIP) techniques and Artificial Neural Networks (ANNs)	The HAM10000 dataset	90.2%
[52]	Support Vector Machine (SVM)	ISIC Dataset	83%
[53]	Random Forest, AdaBoost, Majority voting, Boosted SVM, Boosted GMM classifiers, and Ensemble CNN models.		Ensemble CNN outperforms other ensemble models with an accuracy of 98.67%.

The studies in Table 1 focused on classifying skin diseases with machine learning and deep learning. Most of the study focused on using accuracy as the evaluation metric. In contrast, some used other metrics such as precision, recall, sensitivity, and specificity, which provide a more comprehensive assessment of model performance. Many studies archived an accuracy above 90% with precision, sensitivity, recall, and specificity. In contrast, others did not, which can sometimes indicate overfitting due to an imbalanced dataset. Other metrics like precision, recall, and F1 score may show poor performance in the minority class, even though accuracy appears high. Therefore, it is important to consider multiple evaluation metrics to ensure the model performs well across all classes, not just the majority class, especially when working with imbalanced datasets. In such cases, accuracy alone may be misleading, as it does not account for false positives or false negatives, which are important, especially in medical image classification. Therefore, using a combination of evaluation metrics would offer a more accurate and reliable measure of model performance, particularly in scenarios where precision and recall are more important than overall accuracy.

3. Research Methodology

The methods and results of this study were structured following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PRISMA is a widely recognized and respected framework for conducting systematic reviews, ensuring transparency, and enhancing the quality of research synthesis.

3.1. Search Strategy

The search for relevant literature was conducted across reputable academic databases, including Google Scholar, Science Direct, and IEEE Explore, focusing on journal articles and conference papers. The articles reviewed were from 2019 to 2023, with the latest research findings. The search terms were "skin disease detection" AND (machine learning OR AI OR "deep learning"), aimed at identifying studies in the field of skin disease detection using machine learning techniques.

3.2. Study selection: inclusion and exclusion criteria

We evaluated the titles and abstracts of each publication to determine the relevant studies. After the initial screening, we accessed full-text articles for further assessment. Studies were considered eligible for inclusion if they met the following criteria. A publication providing relevant information on skin disease detection, a publication written in English, a paper publication from 2019-2023, research and review articles, and open-access papers. Studies were excluded if they did not align with the criteria of the research questions, were published before 2019, were written in a language other than English, or were in the form of books, book chapters, or papers that did not include their result. The selection process has been illustrated in the PRISMA flow diagram in Figure 2. The initial search identified 1716 literature sources satisfying the search criteria. After the removal of duplicate records, a total of 1701

records were left, and after applying inclusion criteria, 36 papers were retrieved, which were analyzed, and their results were discussed in the study.

3.3. Data Extraction Process

We extracted the studies from 36 research papers closely aligned with our research question. Applying a structured data extraction methodology, consisting of parameters such as titles, authors' names, types of skin diseases studied, datasets used, models employed, and achieved accuracy. These extracted parameters facilitated a comprehensive analysis of skin disease classification through deep learning models and extracting evidence-based conclusions about their performance and accuracy.

3.4. Quality Assessment

The quality of the studies included in this review was carefully assessed to ensure the inclusion of only high-quality and relevant research. Each study was independently evaluated by two reviewers to ensure accuracy and reliability. The assessment focused on key criteria, such as the clarity of the research objectives, the methodology, the dataset, the validity of the results obtained, and the evaluation metrics applied. We prioritized studies that provided detailed information about the datasets used, the models implemented, the achieved accuracy, and the evaluation metrics used. To maintain consistency and minimize bias, any studies that failed to meet these criteria were excluded from the review. This evaluation process ensured that only quality research contributed to the findings of this study.

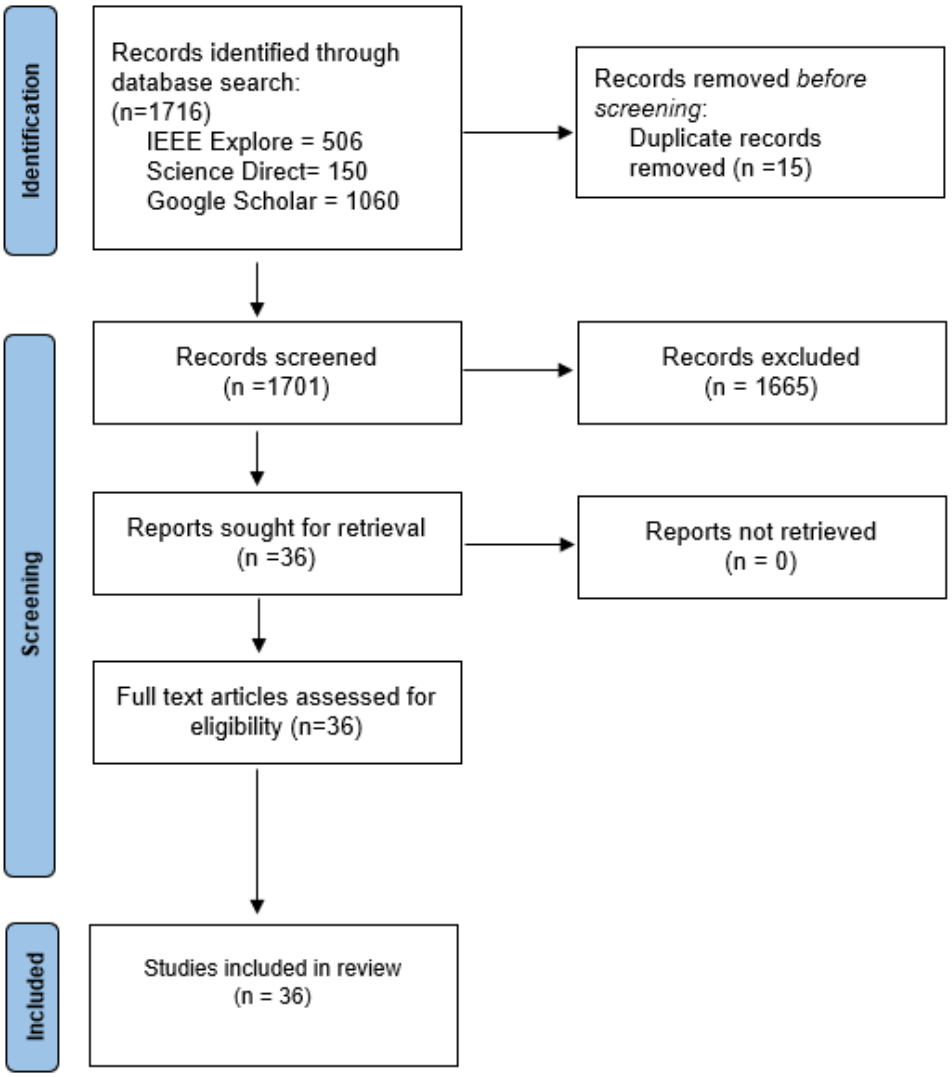

Figure 2. The selection process for relevant articles in skin disease detection

4. Results

This systematic review aims to thoroughly examine existing studies to identify the methods, datasets, results, and limitations used in detecting skin diseases. The illustration in figure 3 shows that the most commonly used method for detecting skin diseases is the CNN model. Among the datasets, publicly available datasets like ISIC and Ham10000 were the most frequently used for skin disease detection. The reported accuracies varied depending on the dataset and method, with a range from 75% to 95%, and the highest accuracy was achieved using the CNN method on the ISIC dataset. The review also pointed out significant gaps, including a limited focus on diverse skin tones in skin disease datasets, difficulty in distinguishing different skin diseases due to similar features, and limited datasets. Table 1 shows the reference of the publication, the method used, the dataset used, and the accuracy achieved in each study. The table is used to address the research questions (RQ1, RQ2, RQ3, RQ4) of the systematic review. The data from the article are extracted and analyzed systematically to address the four (4) research questions, and conclusions are drawn based on the answered research questions. In the Discussion section below, we will discuss each research question in detail, evaluate the correct state of the research question, and determine the direction for future development.

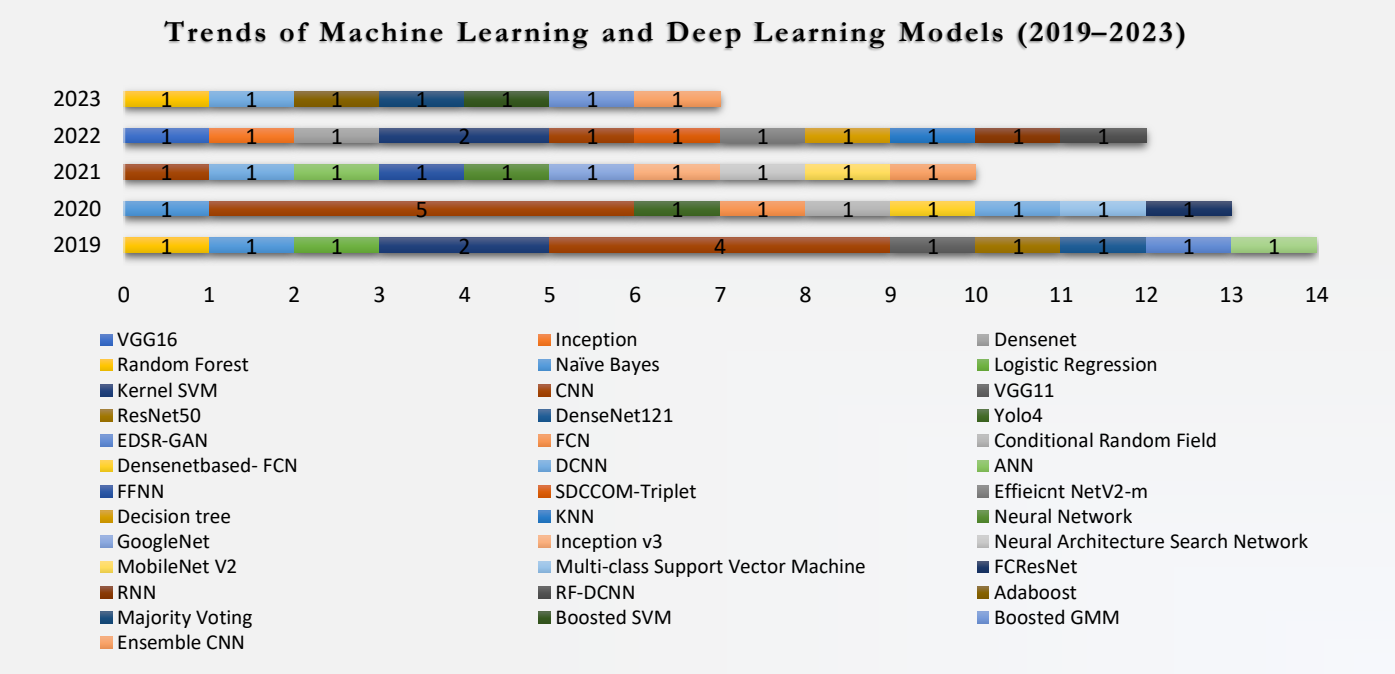

Figure 3. Illustrates the trend in methods employed

5. Discussion

The systematic reviews examine the methods for detecting skin diseases using machine learning. Our systematic review focused on studies from the past 4 years that answer the research question. The purpose of the review is to provide a comprehensive overview of the existing literature on skin disease detection using machine learning, as the result is used for prominent issues in the future because of the requirement for personalized treatment. This section will thoroughly address the research question and provide future findings and analysis related to the research questions.

5.1. Research Question 1: What are the most commonly used methods for skin disease detection?

Machine learning and deep learning have emerged as powerful tools in dermatology, allowing skin disease detection, which has improved how we detect skin problems, making it more efficient and quicker. Table 1 shows the models used in the detection of skin diseases, which are CNN, pre-trained models such as VGG16, DenseNet, Inception, ResNet50, VGG11, MobileNetV2, GoogleNet, InceptionV3, Neural Architecture Search Network,

EfficientNetV2-M, machine learning models such as RF, NB, LR, SVM, DT, KNN, LVQ, and other models such as TEMCM, FCN, Conditional Random Field, SDCCOM-Triplet, YOLOv4, Neural Network, RF-DCNN, the stacking method, and ANN. The analysis shows that the CNN model is the most commonly used among the various models used in the previous study. Multiple studies have adopted the CNN model, including studies like [20], [22]–[24], [28], [42], [43], [45], [47]. Some studies, such as [24], [32], [35], [40], [51], [54] have taken a different route by using pre-trained models. Also, study such as [20], [23], [26], [36], [37], [39], [41], [42], [50] used machine learning models. Finally, studies such as [3], [21], [25], [29], [33], [48], [49] have adopted other models.

5.2. Research Question 2: What evaluation metrics are commonly used to assess the performance of skin disease detection models?

Researchers use several evaluation metrics to assess the performance of models, and these metrics provide quantitative measures that help gauge the effectiveness of the models in correctly identifying and classifying skin conditions.

Table 2. Evaluation Metrics and Formulas

Evaluation Metrics	Definition	Formula
Accuracy (%)	The proportion of correctly classified cases (true positives and negatives) among all cases. It provides an overall measure of classification performance.	Accuracy = $\frac{TN+TP}{TN+TP+FN+FP} X\ 100$
Precision	Precision is also known as positive predictive value; It is the proportion of true positives among the instances that were predicted as positive. It focuses on the accuracy of positive predictions.	Precision=$\frac{TP}{TP+FP}$×100%
Sensitivity (TPR)	Also known as recall or true positive rate, sensitivity is the proportion of actual positive cases the model correctly identified. It measures the ability of the model to capture positive instances	Sensitivity=$\frac{TP}{TP+TN}$
Receiver Operating Characteristic - Area Under the Curve (ROC-AUC)	ROC-AUC is a metric that. quantifies the overall performance of a binary classification model by computing the area under the ROC curve.	ROC-AUC = $\sum_{i=1}^{n} \frac{(TPR+TPR).(FPR-FPR)}{2}$
Recall	Recall, also known as sensitivity, is defined as the ratio of correctly identified positive samples to the total number of actual positive samples	Recall = $\frac{TP}{FP+FN}$x100
F1-Score	The F1 score is the weighted average that returns a single value by combining precision and recall	F1- Score = 2 x $\frac{Precison\ x\ Recall}{Precision+Recall}$
Specificity (TNR)	Specificity measures the proportion of actual negative cases correctly identified by the model, it ensures that the model accurately classifies and minimizes the risk of false positives.	Specificity = $\frac{TN}{TN+FP}$x100

Table 2 presents the evaluation metrics used to assess a model's accuracy. The choice of these metrics depends on the dataset's objectives and characteristics. When working with imbalanced datasets, it is important to select the right performance metrics to evaluate the model's effectiveness accurately. Relying solely on accuracy, which measures the proportion of correctly predicted instances (both positive and negative) among all instances, can be misleading in the case of imbalanced data. The majority class can heavily influence accuracy. Precision measures the proportion of correctly predicted positive cases (true positives) out of all instances predicted as positive, including false positives. While precision helps reduce false positives, it does not account for cases where the model fails to identify true positives, which is important in imbalanced datasets. Recall, also known as sensitivity or the true positive rate, calculates the proportion of actual positive cases (true positives) correctly identified by the model, relative to the total number of actual positive cases (true positives + false negatives). Recall focuses on capturing as many positive instances as possible, which may lead to an increase in false positives. The F1 score balances precision and recall, aiming to minimize false positives while maximizing the identification of true positives. However, it may not fully reflect the model's performance in the case of highly imbalanced datasets. Specificity, or the true negative rate, measures the proportion of actual negative instances correctly classified as negative. Specificity should be considered alongside recall or the F1 score to ensure that the model effectively addresses both the majority and minority classes. Finally, sensitivity, which prioritizes identifying positive instances, should be combined with precision to minimize false positives. The selection of metrics to evaluate a model's performance depends on the specific goals of the problem, ensuring that both the majority and minority classes are considered, especially when dealing with imbalanced data.

5.3. Research Question 3: What commonly available datasets focus on skin disease detection?

Several datasets that address skin disease detection are available. These datasets serve as a resource for developing and evaluating skin detection algorithms. The datasets used in the reviewed studies are the HAM10000 dataset, ISIC 2019 dataset, ISIC 2018, ISIC 2016, ISBI 2018, PH2 dataset, Kaggle's dataset, ISIC 2020, Dermnet, ISIC Archive, DermIS dataset, DermQuest dataset, ISIC 2017, and The ISIC 2019 Challenge dataset.

1. The HAM10000 dataset

The ISIC assembled the HAM10000 dataset. It consists of 10,015 dermoscopic images representing seven different diseases found in pigmented skin lesions. These diseases include nematode disease and intraepithelial carcinoma, BCC, benign keratoid lesions, cutaneous fibroma, MEL, NV, and VASC, including hemangiomas, purulent granulomas, and subcutaneous hemorrhage [55], [56]. Many authors have used The Ham10000 dataset due to its diversity of skin lesions. For example, authors like [54], [21], [24], [27], [30], [33], [35], [37], [48].

2. ISIC Dataset

The ISIC is a publicly available skin disease dataset and has become a driving force for research into melanoma classification. The ISIC organizes yearly skin lesion challenges to attract more involvement among researchers and improve the diagnosis of computer-aided diagnosis algorithms. The ISIC 2016 dataset collected [57] contains 900 training images and 379 test images, 1279 images. The dataset is for training and testing to identify the specific types of malignancy and benignancy. Examples of authors that used the ISIC2016 dataset [25]. The ISIC2017 dataset collected by [58] contains 2000 training and 600 test images, a total of 2600 images. The dataset is for both training and testing, and it contains images including four classes: (1) MEL, (2) NV, (3) SK, or (4) MEL or NV. Authors like [39] used the ISIC2017 dataset.

The ISIC 2018 dataset, as released by [55], includes an expanded dataset that included 10,015 training images and 1,512 test images for a combined total of 11,527 images of MEL, NV, BCC, AK, BK, DF, and VASC. Examples of authors that used the ISIC2018 dataset [25], [29], [30], [32], [51]. The ISIC 2019 dataset by [59] was released. This dataset contains 25,331 training images and 8,238 test images, totaling 33,569. The dataset is only for the training set. The testing metadata indicates the patient's approximate age, anatomical site, and gender. The images include MEL, NV, BCC, AK, BK, DF, VASC, and SCC. Examples of authors that use the ISIC2019 dataset [41], [37], [35], [33], [22]. The ISIC 2020 dataset, as published by [60], included 33,126 training images and 10,982 test images, comprising 44,108 images for only training. The test set also includes patient metadata indicating patient ID,

patient approximate age, anatomical site, and gender. An example of a dataset that uses the ISIC2020 dataset is [33], [35].

3. ISBI 2018

The ISBI 2018 dataset was published by [55]. The dataset comprised more than 20,000 images from leading clinical centers internationally covering multiple skin conditions, including MEL, SK, and NV.

4. PH2 dataset

Study [61] curated the PH2 dataset to facilitate research in developing and evaluating classification and segmentation methods. The dataset was built from a collaborative research effort involving the Universidade do Porto, T'ecnico Lisboa, and the Dermatology Service of Hospital Pedro Hispano in Matosinhos, Portugal. Comprising 200 RGB color images with a 768 × 560-pixel resolution, the dataset includes 80 common nevi, 80 atypical nevi, and 40 MEL. The dermoscopic images were acquired consistently under the same conditions using the Tuebinger Mole Analyzer system with a magnification of 20×. Some studies that leveraged the PH2 dataset, such as [29],[32], [3], [4].

5. Kaggle's dataset

Kaggle is an online data science platform hosting thousands of datasets covering various topics and industries. Anthony Goldbloom and Ben Hamner founded it in April 2010, with over 400,000 public notebooks and 50,000 public datasets. Studies like [62] used the Kaggle dataset.

6. DermNet

DermNet, the world's leading free dermatology resource, was created for healthcare professionals to serve as a comprehensive database of skin conditions. It consists of over 20,000 skin images separated into 23 classes. DermNet was launched in 1996 by dermatologists Dr. Amanda Oakley, Dr. Mark Duffill, and Dr. Marius Rademaker. Studies conducted by [34] detected skin disease using the DermNet dataset.

7. DermIS dataset

DermIS is a free dataset website built by the University of Heidelberg, Germany, and consists of 7,172 images divided into 735 categories. Each image in this dataset has regular disease diagnosis labels and textual descriptions detailing lesion location, race, and age. Studies conducted by [38] carried out skin disease detection using the DermIS dataset.

8. DermQuest dataset

DermQuest is a dataset consisting of 137 images divided into two classes, MEL and NV, of 76 and 61 images, respectively. The dataset was deactivated on December 31, 2019. Studies done by [38] used the dataset to detect and categorize skin lesions.

In Section 5.3, various datasets for skin disease detection have been discussed, each varying in size and the types of skin conditions covered. Datasets like HAM10000, ISIC 2016-2020, ISBI 2018, DermNet, PH2 dataset, DermQuest dataset, DermIS dataset, and Kaggle's dataset are commonly used in studies, with each dataset containing images of various skin conditions, which can be found in Table 1. Handling these datasets, including preprocessing steps like augmentation, normalization, and segmentation, is essential, as they are closely correlated with enhancing model performance by ensuring better data representation, consistency, and robustness. The ISIC dataset, in particular, stands out as the most widely used and publicly accessible dataset, offering a large number of images and accompanying metadata. When detecting skin diseases, distinguishing between certain conditions can be challenging due to significant visual similarities and overlapping features among different conditions. For instance, MEL and NV can appear as pigmented lesions, with differences in color within a skin lesion that are hard to notice. Similarly, psoriasis and seborrheic dermatitis share redness and scaling, while AK and BCC can both present as scaly, crusted patches. Conditions like eczema and contact dermatitis also exhibit similar patterns of inflammation and itching. Therefore, there is a need to use diverse datasets and advanced algorithms to enhance model performance and improve diagnostic accuracy.

5.4. Research Question 4: What are the challenges faced in applying machine learning and deep learning techniques to detect skin diseases?

While different limitations from different authors are outlined in Table 3, we also present a summary of these limitations. Skin disease detection through machine learning and deep learning techniques presents various challenges researchers face. These challenges include:

1. Limited and Biased Datasets

A limited dataset causes challenges in training machine learning models, resulting in overfitting, where the model becomes too tailored to the specific training data and struggles to generalize to new, unseen data. Obtaining a large dataset from training is essential as it captures a broader range of patterns and variations in the data, improving its ability to predict new and diverse datasets accurately. Data-augmentation techniques can be used to increase the size of the dataset [4], [20], [36], [39], [40], [49], [54], [62].

2. Limited to Detecting a Single Disease

Detecting only a single disease using machine learning can be challenging due to the model's limitations in capturing the variability of skin diseases. The lack of diverse datasets during model training can contribute to reduced sensitivity and specificity, affecting the accuracy of disease detection. Studies like [25], [26], [54] detected single skin diseases.

3. Lacks variability in detecting different skin types

Having different variations of skin types is crucial as it enables the model to classify any skin type more accurately. The diversity in skin tones and textures within the dataset ensures that the machine-learning model becomes robust and adaptable. Through training on different skin variations, the model acquires the ability to recognize the features, improving its capacity to classify diverse skin types with good accuracy and precision [26], [39], [52].

4. Challenging to identify diverse skin diseases

Deep learning models sometimes struggle to accurately differentiate between different types of skin diseases. This is a detection challenge caused by a limited training dataset and can be addressed by incorporating diverse datasets, fine-tuning model parameters, and implementing advanced architectures and algorithms [32], [37].

6. Conclusions and Future Work

The systematic review was conducted to provide a comprehensive overview through the analysis of existing literature on the detection of skin diseases using machine learning, focusing on the past four years, with a specific focus on the methods, evaluation metrics, datasets, and challenges. Researchers have used traditional machine learning algorithms and deep learning algorithms to create models for detecting skin diseases, with more emphasis on deep learning models, particularly CNN and pre-trained networks. The evaluation metrics, including accuracy, precision, sensitivity, ROC curve, recall, F1-score, and specificity, were applied to evaluate the performance of skin disease detection models. In this study, CNN models consistently outperformed traditional machine learning models, achieving higher accuracy rates, where large, high-quality datasets such as ISIC, DermNet, and HAM10000 were used. Despite significant progress in employing machine learning and deep learning techniques, challenges persist in accurately detecting skin diseases, such as datasets' scarcity and bias, disease coverage restrictions, and difficulties in classifying diverse skin types. Researchers are continuously exploring innovative solutions to overcome these challenges. Compared to other studies on the detection of skin diseases, which focused on specific skin conditions and deep learning models from 2011 to 2021, our study stands out by outperforming theirs. We conducted a comprehensive systematic review covering recent years from 2019 to 2023, analyzing various deep-learning models and exploring multiple skin diseases.

Future studies should focus on obtaining more diverse and larger datasets to enhance the training of machine learning models, advanced methods to improve the classification performance of models, especially when faced with distinguishing similar features, and also improve the model's capability to identify various skin types. Also, a thorough analysis of more recent studies is necessary to ensure that the latest advancements in AI are incorporated. It is also important to consider the ethical implications of AI in dermatology. This review has shown how AI technologies have the potential to impact the field of dermatology significantly.

Abbreviations

ABCD	Asymmetric Border Color Diameter
AI	Artificial Intelligence
AK	Actinic Keratosis
ANN	Artificial Neural Network
AUC	Area Under Curve

BCC	Basal Cell Carcinoma
BK	Benign Keratosis
CSS	Cascading Style Sheets
DF	Dermatofibroma
DT	Decision Tree
DWT	Discrete Wavelet Transform
EDSRGAN	Enhanced Deep Super-Resolution Generative Adversarial Network
ESRGAN	Enhanced Super-Resolution Generative Adversarial Network
ESD	Erythemato-Squamous Diseases
FCN	Fully Convolutional Network
FFNN	Feed-forward Neural Network
GLCM	Gray-Level Co-occurrence Matrix
GUI	Graphical User Interface
ISIC	International Skin Imaging Collaboration
Kernel	SVM Kernel Support Vector Machine
KNN	K-Nearest Neighbor
LBP	Local Binary Pattern
MCML	Multi-class Support Vector Machine
ML	Melanoma
MSVM	Multi-class Support Vector Machine
NV	Melanocytes Nevus
PH2	Pigmented Lesion Dataset
RBG	Red Blue Green
RF	Random Forest
RF-DCNN	Random Forecast Deep Convolutional Neural Network
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic curve
SCC	Squamous Cell Carcinoma
SK	Seborrheic Keratosis
SRGAN	Super-Resolution Generative Adversarial Network
SVM	Support Vector Machine
UCI	University of California, Irvine
VASC	Vascular Lesions
VGLL BN	Batch Normalization with VGLL.

Author Contributions: Conceptualization: O.J.; Methodology: O.J, E.O, F.O; Formal analysis: O.J; Investigation: O.J., O.J., E.O.; Resources: O.J and O.J.; Writing—O.J., O.J., E.O.; Writing—review and editing: O.J., O.J., E.O., F.O; Supervision: O.J and E.O.

Funding: This research received no external funding

Conflicts of Interest: The authors have no conflict of interest related to this publication

References

[1] O. Jaiyeoba, E. Ogbuju, O. T. Yomi, and F. Oladipo, "Development of a Model to Classify Skin Diseases using Stacking Ensemble Machine Learning Techniques," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 22–38, May 2024, doi: 10.62411/jcta.10488.

[2] F. Solano, "Photoprotection and Skin Pigmentation: Melanin-Related Molecules and Some Other New Agents Obtained from Natural Sources," *Molecules*, vol. 25, no. 7, p. 1537, Mar. 2020, doi: 10.3390/molecules25071537.

[3] P. P. Tumpa and M. A. Kabir, "An artificial neural network based detection and classification of melanoma skin cancer using hybrid texture features," *Sensors Int.*, vol. 2, p. 100128, 2021, doi: 10.1016/j.sintl.2021.100128.

[4] 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," *Expert Syst. Appl.*, vol. 141, p. 112961, 2020, doi: 10.1016/j.eswa.2019.112961.

[5] S. Rasool, A. H. Khan, Q. Rasool, S. Abbas, and ..., "A Systematic Analysis for the Detection of Skin Disease Using Deep Learning Methodologies," *J. Comput. …*, vol. 04, no. 02, 2023.

[6] K. W. Ng and W. M. Lau, "Skin deep: The basics of human skin structure and drug penetration," *Percutaneous Penetration Enhanc. Chem. Methods Penetration Enhanc. Drug Manip. Strateg. Veh. Eff.*, pp. 3–11, Jan. 2015, doi: 10.1007/978-3-662-45013-0_1.

[7] V. Pugazhenthi *et al.*, "Skin Disease Detection And Classification," *Int. J. Adv. Eng. Res. Sci.*, vol. 6, no. 5, pp. 396–400, 2019, doi: 10.22161/ijaers.6.5.53ï.

[8] G. Bocheva, R. M. Slominski, and A. T. Slominski, "Environmental Air Pollutants Affecting Skin Functions with Systemic Implications," *Int. J. Mol. Sci.*, vol. 24, no. 13, p. 10502, Jun. 2023, doi: 10.3390/ijms241310502.

[9] L. Blicharz, J. Czuwara, L. Rudnicka, and A. Torrelo, "Autoinflammatory Keratinization Diseases—The Concept, Pathophysiology, and Clinical Implications," *Clin. Rev. Allergy Immunol.*, vol. 65, no. 3, pp. 377–402, Dec. 2023, doi: 10.1007/s12016-023-08971-3.

[10] A. K. Sah, S. Bhusal, S. Amatya, M. Mainali, and S. Shakya, "Dermatological Diseases Classification using Image Processing and Deep Neural Network," *Proc. - 2019 Int. Conf. Comput. Commun. Intell. Syst. ICCCIS 2019*, vol. 2019-Janua, pp. 381–386, Oct. 2019, doi: 10.1109/ICCCIS48478.2019.8974487.

[11] M. Al Mamun and M. S. Uddin, "A Survey on a Skin Disease Detection System," *Int. J. Healthc. Inf. Syst. Informatics*, vol. 16, no. 4, pp. 1–17, Oct. 2021, doi: 10.4018/IJHISI.20211001.oa35.

[12] F. A. Mphande, "The Structure and Function of the Skin," in *Skin Disorders in Vulnerable Populations*, Singapore: Springer Singapore, 2020, pp. 13–22. doi: 10.1007/978-981-15-3879-7_2.

[13] N. V. Kumar, P. V. Kumar, K. Pramodh, and Y. Karuna, "Classification of Skin diseases using Image processing and SVM," in *Proceedings - International Conference on Vision Towards Emerging Trends in Communication and Networking, ViTECoN 2019*, Mar. 2019. doi: 10.1109/ViTECoN.2019.8899449.

[14] P. R. Kshirsagar, H. Manoharan, S. Shitharth, A. M. Alshareef, N. Albishry, and P. K. Balachandran, "Deep Learning Approaches for Prognosis of Automated Skin Disease," *Life*, vol. 12, no. 3, p. 426, Mar. 2022, doi: 10.3390/life12030426.

[15] T. Goswami, V. K. Dabhi, and H. B. Prajapati, "Skin Disease Classification from Image - A Survey," pp. 599–605, 2020, doi: 10.1109/ICACCS48705.2020.9074232.

[16] D. Bordoloi *et al.*, "Classification and Detection of Skin Disease Based on Machine Learning and Image Processing Evolutionary Models," *Comput. Assist. Methods Eng. Sci.*, vol. 30, no. 2, pp. 247–256, Apr. 2023, doi: 10.24423/CAMES.479.

[17] H. Bhatt, V. Shah, K. Shah, R. Shah, and M. Shah, "State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review," *Intell. Med.*, vol. 3, no. 3, pp. 180–190, 2023, doi: 10.1016/j.imed.2022.08.004.

[18] M. Dildar *et al.*, "Skin Cancer Detection: A Review Using Deep Learning Techniques," *Int. J. Environ. Res. Public Health*, vol. 18, no. 10, p. 5479, May 2021, doi: 10.3390/ijerph18105479.

[19] T. Kameswara Rao, P. Chamanthi, N. T. Kumar, R. L. Amulya, M. Uday Sagar, and U. G. Students, "Skin Disease Detection Using Machine Learning," *Int. J. Food Nutr. Sci.*, vol. 11, no. 12, Apr. 2023, doi: 10.48047/ijfans/v11/i12/171.

[20] S. Bhadula*, S. Sharma, P. Juyal, and C. Kulshrestha, "Machine Learning Algorithms based Skin Disease Detection," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 2, pp. 4044–4049, Dec. 2019, doi: 10.35940/ijitee.B7686.129219.

[21] H. Q. Yu and S. Reiff-Marganiec, "Targeted Ensemble Machine Classification Approach for Supporting IoT Enabled Skin Disease Detection," *IEEE Access*, vol. 9, pp. 50244–50252, 2021, doi: 10.1109/ACCESS.2021.3069024.

[22] 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," *Appl. Sci.*, vol. 11, no. 5, p. 2145, Feb. 2021, doi: 10.3390/app11052145.

[23] M. J. Hajgude, A. Bhavsar, H. Achara, and N. Khubchandani, "Skin Disease Detection Using Image Processing with Data Mining and Deep Learning," *Int. Res. J. Eng. Technol.*, vol. 6, no. 4, pp. 4363–4366, 2019.

[24] M. M. I. Rahi, F. T. Khan, M. T. Mahtab, A. K. M. Amanat Ullah, M. G. R. Alam, and M. A. Alam, "Detection Of Skin Cancer Using Deep Neural Networks," in *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Dec. 2019, pp. 1–7. doi: 10.1109/CSDE48274.2019.9162400.

[25] S. Albahli, N. Nida, A. Irtaza, M. H. Yousaf, and M. T. Mahmood, "Melanoma Lesion Detection and Segmentation Using YOLOv4-DarkNet and Active Contour," *IEEE Access*, vol. 8, pp. 198403–198414, 2020, doi: 10.1109/ACCESS.2020.3035345.

[26] P. Banasode, M. Patil, and N. Ammanagi, "A Melanoma Skin Cancer Detection Using Machine Learning Technique: Support Vector Machine," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1065, no. 1, pp. 1–5, 2021, doi: 10.1088/1757-899X/1065/1/012039.

[27] I. Abd El-Fattah, A. M. Ali, W. El-Shafai, T. E. Taha, and F. E. Abd El-Samie, "Deep-learning-based super-resolution and classification framework for skin disease detection applications," *Opt. Quantum Electron.*, vol. 55, no. 5, p. 427, May 2023, doi: 10.1007/s11082-022-04432-x.

[28] K. S. Rao, P. S. Yelkar, O. N. Pise, and S. Borde, "Skin Disease Detection Using Machine Learning Techniques," *J. Xidian Univ.*, vol. 15, no. 4, pp. 64–68, Apr. 2021, doi: 10.37896/jxu15.4/022.

[29] A. A. Adegun and S. Viriri, "FCN-Based DenseNet Framework for Automated Detection and Classification of Skin Lesions in Dermoscopy Images," *IEEE Access*, vol. 8, pp. 150377–150396, 2020, doi: 10.1109/ACCESS.2020.3016651.

[30] M. Neema, A. S. Nair, A. Joy, and A. Pradeep Menon, "Skin Lesion/Cancer Detection Using Deep Learning," *Int. J. Appl. Eng. Res.*, vol. 15, no. 1, pp. 11–17, 2020.

[31] S. Ahmed *et al.*, "Human Skin Diseases Detection and Classification using CNN," in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2023, pp. 1–6. doi: 10.1109/ECCE57851.2023.10101636.

[32] I. Abunadi and E. M. Senan, "Deep Learning and Machine Learning Techniques of Diagnosis Dermoscopy Images for Early Detection of Skin Diseases," *Electronics*, vol. 10, no. 24, p. 3158, Dec. 2021, doi: 10.3390/electronics10243158.

[33] S. Ozturk and T. Cukur, "Deep Clustering via Center-Oriented Margin Free-Triplet Loss for Skin Lesion Detection in Highly Imbalanced Datasets," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 9, pp. 4679–4690, Sep. 2022, doi: 10.1109/JBHI.2022.3187215.

[34] T. A. Rimi, N. Sultana, and M. F. Ahmed Foysal, "Derm-NN: Skin Diseases Detection Using Convolutional Neural Network," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 1205–1209, 2020, doi: 10.1109/ICICCS48265.2020.9120925.

[35] V. Venugopal, N. I. Raj, M. K. Nath, and N. Stephen, "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images," *Decis. Anal. J.*, vol. 8, no. December 2022, p. 100278, Sep. 2023, doi: 10.1016/j.dajour.2023.100278.

[36] N. S. A. ALEnezi, "A Method of Skin Disease Detection Using Image Processing and Machine Learning," *Procedia Comput. Sci.*, vol. 163, pp. 85–92, 2019, doi: 10.1016/j.procs.2019.12.090.

[37] M. Ahammed, M. Al Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthc. Anal.*, vol. 2, no. October, p. 100122, 2022, doi: 10.1016/j.health.2022.100122.

[38] V. Srividhya, K. Sujatha, R. S. Ponmagal, G. Durgadevi, L. Madheshwaran, and V., "Vision based Detection and Categorization of Skin lesions using Deep Learning Neural networks," *Procedia Comput. Sci.*, vol. 171, no. 2019, pp. 1726–1735, 2020, doi: 10.1016/j.procs.2020.04.185.

[39] V. R. Balaji, S. T. Suganthi, R. Rajadevi, V. Krishna Kumar, B. Saravana Balaji, and S. Pandiyan, "Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier," *Measurement*, vol. 163, p. 107922, Oct. 2020, doi: 10.1016/j.measurement.2020.107922.

[40] T. K. Yoo, J. Y. Choi, H. K. Kim, I. H. Ryu, and J. K. Kim, "Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images," *Comput. Methods Programs Biomed.*, vol. 205, p. 106086, 2021, doi: 10.1016/j.cmpb.2021.106086.

[41] M. K. Monika, N. A. Vignesh, C. U. Kumari, M. N. V. S. S. Kumar, and E. Lydia, L, "Skin cancer detection and classification using machine learning," *Mater. Today Proc.*, vol. 33, pp. 4266–4270, 2020, doi: 10.1016/j.matpr.2020.07.366.

[42] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," *2020 Int. Conf. Adv. Technol. Signal Image Process. ATSIP 2020*, Sep. 2020, doi: 10.1109/ATSIP49331.2020.9231544.

[43] U. Aishwarya, I. J. Daniel, and R. Raghul, "Convolutional Neural Network based Skin Lesion Classification and Identification," *2020 Int. Conf. Inven. Comput. Technol.*, pp. 264–270, Feb. 2020, doi: 10.1109/ICICT48043.2020.9112485.

[44] T. T. Nguyen *et al.*, "Artificial Intelligence in the Battle against Coronavirus (COVID-19): A Survey and Future Research Directions," 2020.

[45] R. S. Sanketh, M. Madhu Bala, P. V. Narendra Reddy, and G. V. S. Phani Kumar, "Melanoma Disease Detection Using Convolutional Neural Networks," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2020, pp. 1031–1037. doi: 10.1109/ICICCS48265.2020.9121075.

[46] R. Kaur, H. Gholamhosseini, and R. Sinha, "Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images," pp. 1–15, 2022.

[47] Y. Salve, V. Khile, N. Dinkwar, A. Patil, P. Dhore, and S. Angadi, "A Study on Skin Disease Detection and Hospital Recommendation System," *J. Pharm. Negat. Results*, vol. 13, pp. 7539–7548, Dec. 2022, doi: 10.47750/pnr.2022.13.S07.909.

[48] A. Kalaivani and S. Karpagavalli, "Detection and classification of skin diseases with ensembles of deep learning networks in medical imaging," *Int. J. Health Sci. (Qassim).*, vol. 6, no. April, pp. 13624–13637, Jun. 2022, doi: 10.53730/ijhs.v6nS1.8402.

[49] A. Swaroop and A. K. Srivastava, "Artificial intelligence based approach for melanoma skin cancer detection," *Dogo Rangsang Res. J.*, vol. 9, no. 01, 2022.

[50] M. A. Araaf, K. Nugroho, and D. R. I. M. Setiadi, "Comprehensive Analysis and Classification of Skin Diseases based on Image Texture Features using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 31–40, Sep. 2023, doi: 10.33633/jcta.v1i1.9185.

[51] I. Razzak, G. Shoukat, S. Naz, and T. M. Khan, "Skin Lesion Analysis Toward Accurate Detection of Melanoma using Multistage Fully Connected Residual Network," *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9206881.

[52] A. M. A. Thaajwer and P. U. A. Ishanka, "Melanoma skin cancer detection using image processing and machine learning techniques," *ICAC 2020 - 2nd Int. Conf. Adv. Comput. Proc.*, pp. 363–368, Dec. 2020, doi: 10.1109/ICAC51239.2020.9357309.

[53] S. A. Alshawi and G. F. K. A. Musawi, "Skin Cancer Image Detection and Classification by CNN based Ensemble Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 5, p. 2023, 2023, doi: 10.14569/IJACSA.2023.0140575.

[54] T. Kameswara Rao, P. Chamanthi, N. T. Kumar, R. L. Amulya, M. Uday Sagar, and U. G. Students, "Skin Disease Detection Using Machine Learning," *IJFANS Int. J. Food Nutr. Sci.*, vol. 11, no. 1593–1604, 2022, doi: 10.48047/IJFANS/V11/I12/171.

[55] N. Codella *et al.*, "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)," 2018.

[56] P. Tschandl, C. Rosendahl, and H. Kittler, "Data Descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions Background & Summary," *Nat. Publ. Gr.*, 2018, doi: 10.1038/sdata.2018.161.

[57] D. Gutman *et al.*, "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)," 2016.

[58] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Apr. 2018, pp. 168–172. doi: 10.1109/ISBI.2018.8363547.

[59] M. Combalia *et al.*, "BCN20000: Dermoscopic Lesions in the Wild," Aug. 2019.

[60] V. Rotemberg *et al.*, "a patient-centric dataset of images and metadata for identifying melanomas using clinical context," 2021, doi: 10.1038/s41597-021-00815-z.

[61] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2 - A dermoscopic image database for research and benchmarking," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 5437–5440, 2013, doi: 10.1109/EMBC.2013.6610779.

[62] A. Ahmed, H. Ahmad, M. Khurshid, and K. Abid, "Classification of Skin Disease using Machine Learning," *VFAST Trans. Softw. Eng.*, vol. 11, no. 1, pp. 109–122, 2023.