University Of AL-Qadisiyah

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

An Analytical and Comparative Study of Modern Intelligent Models for Skin Cancer Diagnosis

Zahraa Saad Abdul Wahid[a], Zahraa Ch. Oleiwi[a]*, Rasha Falah kadhem[a]

[a]*College of computer science and Information Technology, University of Al-Qadisiyah, Diwanya, Iraq. Email: zahraasaadadbulwahid@gmail.com, zahraa.chaffat@qu.edu.iq, rasha.kadhem@qu.edu.iq*

ARTICLE INFO

Article history:
Received: 19 /01/2026
Rrevised form: 24 /02/2026
Accepted : 26 /02/2026
Available online: 30 /03/2026

Keywords:

Keywords: Skin cancer, Transfer learning, intra-class variability, CNN, HAM10000.

ABSTRACT

Early diagnosis of skin cancer especially melanoma is a significant issue because of the visual similarities of lesion types, intra class variability, and drawbacks of the manual diagnosis. Over the past years, there is a wide range of artificial intelligence (AI) methods suggested, but the results reported were quite different because of the variations in datasets, preprocessing pipelines, model architectures, and evaluation protocols. This is a systematic review article of 29 peer-reviewed studies published 2015–2025 that were identified in major scientific databases and divided into three key categories, namely traditional machine learning (ML), deep learning (DL), and transfer learning (TL) frameworks. Comparatively, the studies are assessed by aspects of characteristics of the dataset, feature representation strategies, model complexity, diagnostic performance and clinical applicability. The discussion has shown that deep and transfer learning models tend to perform better than traditional techniques of ML in terms of classification accuracy, yet various issues on the class imbalance, data heterogeneity, the ability to perform generalization as well as efficiency of computation still exist. Another critical research gap that has been found in the literature is the lack of a standardized analytical framework to compare the ML, DL, and TL models under standardized assessment procedures taking into account the real-life limitations in clinical deployment. This paper is a comparative perspective on the current intelligent skin cancer diagnostic systems in a structured way and it identifies the future research directions of stronger, scalable and clinically sound AI-based solutions.

MSC..

https://doi.org/10.29304/jqcsm.2026.18.12671

1. Introduction

According to the World Health Organization (WHO), skin cancer represents a major global health concern, affecting nearly 25% of the world's population. Among the different types of skin cancer, melanoma is considered the most aggressive and life-threatening form, with an annual increase rate of approximately 5%. Melanoma is primarily caused by excessive exposure to ultraviolet (UV) radiation, particularly in individuals with fair skin, and it originates from the uncontrolled proliferation of melanocyte cells following DNA damage. If not detected and treated at an early stage, melanoma can rapidly metastasize to other parts of the body, significantly increasing mortality rates [1].

*Corresponding author: Zahraa Ch. Oleiwi

Email addresses: zahraa.chaffat@qu.edu.iq

Communicated by 'sub etitor'

Globally, melanoma is one of the most common cancers affecting both men and women. In 2024, more than 100,640 new melanoma cases were reported worldwide, reflecting a 5.0% increase compared to previous years. In the same year, approximately 8,290 deaths were attributed to melanoma, accounting for about 1.4% of total cancer-related deaths. These alarming statistics highlight the critical importance of early detection and accurate diagnosis, particularly for aggressive skin cancer types such as melanoma [1][2].

Traditionally, dermatologists rely on visual inspection and dermoscopic analysis of skin lesions to differentiate between benign and malignant cases. However, manual diagnosis is highly dependent on clinical expertise and is often subject to inter-observer variability, fatigue, and human error. In addition, biopsy procedures are invasive, costly, time-consuming, and may not be suitable for elderly patients. These limitations emphasize the urgent need for automated, reliable, and efficient diagnostic systems to support clinical decision-making [3].

Despite recent advances, skin cancer diagnosis remains challenging due to several factors, including high intra-class variability, visual similarity between different lesion types, hair occlusion, low contrast images, and ambiguous lesion boundaries. Moreover, the rapidly increasing number of skin cancer cases places additional pressure on healthcare systems, further underscoring the necessity for automated diagnostic solutions [4].

In recent years, artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, has demonstrated significant potential in medical image analysis and skin lesion classification. Machine learning algorithms, such as k-nearest neighbors (KNN), artificial neural networks (ANN), decision trees, and random forests, have been widely applied to assist in skin disease diagnosis by learning patterns from handcrafted features. These approaches have shown promising results; however, their performance is often limited by the reliance on manual feature engineering, which restricts generalization across diverse datasets [5].

Deep learning, a subset of machine learning inspired by the human brain's neural structure, has achieved remarkable success in complex image-based tasks. Convolutional Neural Networks (CNNs), in particular, have proven highly effective in medical image classification, segmentation, and detection. Recent developments have enabled the application of advanced deep learning models to skin cancer diagnosis, achieving improved accuracy and robustness. Nevertheless, many existing approaches employ multi-stage pipelines, complex architectures, or extensive preprocessing strategies, which increase computational complexity and limit practical clinical deployment [6].

Although numerous studies have explored the use of AI for skin cancer diagnosis, the reported results vary considerably due to differences in dataset size, preprocessing techniques, model architectures, and evaluation protocols. Furthermore, previous studies often lack a unified analytical framework that systematically compares traditional machine learning methods with modern deep learning and hybrid approaches, particularly in terms of performance, robustness, and clinical practicality [7].

Thus, this paper includes a methodical analytical and comparison study of modern machine learning, deep learning, and transfer learning-based application to skin cancer classification. Instead of just reporting classification accuracy, the study gives a critical comparison on the model architectures, strategies of feature representation, computation complexity, and robustness to different states in datasets. Special focus is given to the evaluation of diagnostic performance, clinical applicability as well as implementation feasibility in the real-life healthcare setting. Besides, this piece of work also demarcates important methodological shortcomings that have been described in the literature, including imbalance of data, generalization issues, and computational constraints, and presents possible research directions to more credible, efficient, and clinically implementable automated skin cancer detection systems.

Though a number of review studies are already devoted to the use of artificial intelligence in skin cancer diagnostics, the vast majority of the currently existing surveys are devoted to either the deep-learning mechanisms or the descriptive accounts of the reported accuracies with no analytical comparison framework developed. Moreover, only in rare cases are reviews conducted systematically across categories comparing traditional machine learning techniques, deep learning methods, and transfer learning methods on the same assessment criteria. Conversely, this paper presents a systematic comparative approach which applies such three paradigms to standardized dimensions, such as feature representation strategy, computational complexity, data characteristics, generalization capacity, and whether the paradigm can be clinically deployed. This work presents the therapeutic viewpoint as more holistic and practical in response compared to the surveys published before it, as the quantitative performance comparison with the critical methodological analysis approach is integrated.

The remainder of this review paper is organized as follows. In section 2, we present some popular public datasets in the medical domain for classification tasks, especially skin cancer cases. Section 3 declares the review methodology. In Section 4, we discuss and explain several recent machine learning and deep learning algorithms and models commonly applied in skin lesion classification challenges. In section 5 and 6 we compared and evaluated some results of research used in this context. Conclusion is presented in section 7 and at last references.

2. Skin Cancer Image Datasets

The quality of artificial intelligence models, their size and variety of datasets have a significant effect on their performance and reliability when it comes to skin cancer diagnosis. The availability of skin lesion datasets that are publicly available is essential in enabling fair comparison of studies and also in propelling research forward. The following section provides a brief overview of the most popular datasets of skin cancer images and outlines their key features as well as the challenges.

- HAM10000 Dataset ("Human Against Machine with 10000 training images")

One of the most popular datasets to use as a benchmark to classify skin lesions is the HAM10000 (Human Against Machine with 10000 training images) database. It has 10,015 dermoscopic images that are classified into seven diagnostic types, which include melanoma and various types of benign lesions. The dataset will be gathered through various sources, which offer heterogeneity in the conditions of acquisition and lesion manifestation, which facilitates sound feature learning. HAM10000, although being popular, has a serious class imbalance as the data predominantly consists of benign tumors and the sample of malignant ones is insufficient. This imbalance can cause classification models to favor the majority classes and typically requires the application of resampling or class weighting, or data augmentation methods [8].

- ISIC Challenge Datasets (2016–2020)

International Skin Imaging Collaboration (ISIC) is a set of publicly available dermoscopic image datasets, which are popular in benchmarking skin cancer classification and segmentation models in the research world. The ISIC Archive has a big and ever-growing collection of quality-controlled images of skin lesions, donated by numerous clinical centers, and labeled to facilitate a variety of diagnostic and analytical tasks. The ISIC datasets are published every year as international challenges and have been developed over time in terms of scale and complexity. Older versions were primarily binary classification and the more recent versions can handle more extensive multiclass and segmentation. To be more precise, in ISIC 2016, there are about 900 images, the main purpose of which is to classify binary melanoma. ISIC 2017 was further extended to approximately 2,000 training images and 600 testing images with multiclass classification and segmentation elements. The data of ISIC 2018 was also expanded to over 10,000 training images in seven diagnostic classes and allowed the stronger testing of classification and segmentation models. Follow up versions represented a big jump in scale and variety of the datasets[9]. ISIC 2019 has around 25331 training images of eight lesion categories, which are complemented by abundant clinical information, including the age, sex, and location of the lesion of the patient. The biggest release of the series, ISIC 2020, offers over 30000 labeled images and binary as well as multiclass classification problems. ISIC 2019 and ISIC 2020 have emerged as the most widely used standards of comparative analysis of modern machine learning and deep learning methods in skin cancer analysis due to their size and difference. There are various challenges and limitations of using the ISIC datasets even though they are widely adopted. Image quality, resolution and acquisition conditions have a significant amount of heterogeneity between the years and various contributing sources. In addition, differences in labelling procedures between different editions of challenges and a structural imbalance of performance in a persistent class (especially the malignant lesion classes) make it difficult to directly compare performances across studies. These aspects may influence the generalization of the models and create problems when it is necessary to draw the fair conclusions when comparing the results which were reported in the literature [9][10].

- PH2 Dataset

PH2 dataset is rather small but of high quality the collection of dermoscopic images of 200 images separated into three major diagnostic groups. Dermatology experts annotated it very carefully and it is commonly used to validate and assess performance. The main disadvantage of PH2 is that the size of this control type is limited, which limits its applicability in training deep learning models in a non-pretrained state. PH2 is therefore normally applied together with transfer learning methods or as an external test set to gauge the generalization of the model [9].

- Other Public Skin Lesion Datasets

Other publicly shared datasets, including MED-NODE, BCN20000 and XiangyaDerm, are usually also employed in skin cancer studies as supplementary data in order to enhance the variety of data. MED-NODE is a small data set with 170 images, which was mainly created to perform the binary classification of melanoma and benign lesions. The BCN200000 is generated on top of ISIC images and contains about 19,000-20,000 images of eight lesion types and is stored on Hugging Face. XiangyaDerm is a fairly larger collection with approximately 47,075 pictures, representing approximately 541 different skin conditions both malignant and non-malignant. These datasets have certain drawbacks in common with ISIC, even though they are useful. They either tend to be of smaller sample sizes, poorer coverage of classes, or lacking clinical annotations, which diminishes their own utility and general prevalence in benchmarking studies. However, they are useful as an addition to datasets, cross-dataset validation, and enhancing model generalization. [11].

Even though publicly available datasets have been very useful in automated detection of skin cancer, a number of issues remain:

- Extreme imbalance of classes, especially in case of malignant lesions.
- Image resolution and image acquisition devices.
- Less representation of a variety of skin colors and demographics.
- Limited access to the clinically validated metadata.

To overcome these shortcomings, it is necessary to create larger, more balanced, and more clinically diverse datasets that will enhance the strength and clinical relevance of AI-driven systems based on diagnosing skin cancer diseases.

3. Review Methodology

The review adheres to a systematic methodology in order to be transparent and reproducible. Major scientific databases such as IEEE Xplore, Scopus, and Web of Science were searched on the basis of relevant literature. The keywords were related to skin cancer classification, machine learning, deep learning, and transfer learning and identified the studies published within the period 2015 to 2025. The inclusion criteria were: (1) the studies that use ML, DL or transfer learning to diagnose skin cancer; (2) the studies that have well reported experimental outcomes; and (3) the studies that use the English language. The exclusion criteria were as follows: (1) the studies that were not concerning skin cancer classification, (2) the studies that did not provide performance evaluation, and (3) the articles which were not peer-reviewed.

The chosen studies were subsequently divided into three groups namely conventional machine learning, deep learning, and transfer learning methods, on which the comparative analysis in this review is based.

4. Artificial Intelligence for Skin Cancer Diagnosis

Initial AI-based systems to diagnose skin cancer used mainly the standard machine learning algorithms. Such methods usually consist of a series of processes, such as image preprocessing, lesion segmentation, handcrafted feature extraction, and classification. The color, texture, shape and asymmetry characteristics based on dermoscopic images are some of the most commonly extracted features. The use of classical classifiers, including Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), and Artificial Neural Networks (ANN) has been extensively used in this context and has shown promising results in tasks of melanoma and non-melanoma classification [12][13][14]. These approaches however greatly rely on the quality of the handcrafted features and the accuracy of segmentation.

CNNs have emerged as the primary paradigm of automated skin cancer diagnosis with the development of deep learning. The CNN-based models can also learn hierarchical representations of features directly using raw images as opposed to the traditional methods, which requires manual engineering of features. It has been demonstrated through many studies that deep CNN architectures are able to work as well as or even better than a trained dermatologist in skin lesion classification [15][16]. VGGNet, ResNet, Inception, DenseNet and EfficientNet are some of the popular CNN models studied widely to perform this task.

Transfer learning has also helped to hasten the process of using deep learning in diagnosis of skin cancer especially where there is a deficiency in labeled medical data. Through the way that CNNs are fine-tuned on pre-trained systems on large picture datasets like ImageNet, researchers can also make adaptations to learned representations to dermoscopic image analysis [17]. This method has demonstrated better accuracy of classification and less training and overfitting. As a result, transfer learning has become a common practice in most of the recent AI-based diagnostics.

Besides single-model architectures, a number of studies have investigated the hybrid as well as ensemble-based AI frameworks. All these methods merge several CNNs or also combine deep learning and optimization algorithms to improve the level of feature discrimination and classification strength. Genetic Algorithms, Particle Swarm Optimization, Artificial Bee Colony and Grey Wolf Optimization have been used as metaheuristic optimization methods to optimize the use of features, hyperparameters or network weights [18][19]. Even though these hybrid systems can be more accurate, adding to them are higher computational complexity and lower interpretability.

In spite of the significant advancements made by the AI based skin cancer diagnostic systems, there are still multiple challenges. The lack of data, imbalance in the classes, and variability in the conditions of image acquisition remain the limiting factors of model generalization. Also, most of the state-of-the-art solutions focus on precision without considering computational efficiency and real-time usability, which are necessary in clinical implementation. These issues emphasize the necessity of additional studies in the direction of strong, effective, and clinically viable AI solutions to skin cancer diagnosis.

4.1. 4.1Machine learning algorithms for skin cancer diagnosis

The initial progress of skin cancer diagnostic systems that are automated has relied heavily on machine learning algorithms. The methods are one of the oldest smart solutions, which were aimed at helping dermatologists to recognize the malignant skin lesions, especially melanoma, through dermoscopic images. The objective of systems that are based on machine learning is to aid in the early detection, decrease the number of diagnostic errors, and enhance the accuracy of decisions in clinical practice.

Conventional machine learning architectures of skin cancer diagnosis commonly consist of a series of ordered operations comprising of image preprocessing, lesion localization, manually devised feature extraction, and lesion classification. The use of preprocessing is used to amplify image quality, eliminate artifacts like hair and noise as well as normalize illumination. Lesion segmentation is a very essential process because it has a direct bearing on the accuracy of the features extracted and the subsequent classification accuracy [20].

The machine learning-based approaches rely on feature extraction in order to identify clinically meaningful features of skin lesion. Features commonly used are color descriptors, texture (Local Binary Patterns and Gray-Level Co-occurrence Matrices), shape-based features, and asymmetry-based features based on the ABCD dermatological rule. All these handmade characteristics will replicate the visual patterns adopted by dermatologists in the manual examination [21][22].

After extraction of features, the different classes of machine learning classifier are used to identify benign and malignant lesions. The reason Support Vector Machines (SVM) have gained widespread use is the fact that they are very robust in high-dimensional feature space and small data set. Most of the literature has indicated that the SVM based classifiers perform better than other conventional classifiers on melanoma detection tasks [23]. On the same note, the k-Nearest Neighbors (KNN) has been applied due to its ease of use and the ability to be used in pattern recognition but its effectiveness is susceptible to feature scaling and class imbalance.

Multilayer Perceptrons (MLP) and Artificial Neural Networks (ANN) have also been popularly studied in skin cancer classification. These models can learn nonlinear association among extracted features and diagnostic outcomes. Moreover, machine learning algorithms like Random Forest and AdaBoost, which are ensemble-based, have shown better accuracy in their classification because of the combination of a large number of weak learners and alleviation of overfitting [24].

A number of studies have also improved the performance of machine learning by incorporating feature reduction and optimization. Principal Component Analysis (PCA) is a dimensionality reduction technique usually used to get rid of the redundant features and enhance the efficiency of the classifiers. Furthermore, algorithms of optimization have been used to choose the most discriminating features with the resultant increase in diagnostic accuracy and low cost of computation [25][26].

The article by Natha and Pothuraju (2023) is a machine learning-based skin cancer detection framework that integrates handcrafted feature extraction, optimization, and classical classifiers. Their approach is based on the derivation of the texture and structure information of the dermoscopic images with the help of Contourlet Transform (CT) and Local Binary Pattern (LBP). Particle Swarm Optimization (PSO) was used to obtain the most discriminative features to reduce the dimensionality of the features and enhance the efficiency of the classification. Support Vector Machine (SVM), Random Forest (RF) and Neural Network (NN) classifiers were then used to classify the selected features. The results of the experiments showed that the SVM classifier had the highest level of performance with an overall recognition value of 86.9 and lesser computational complexity than the other classifiers. Although the pipeline is structured and feature optimization process is well applied, the performance reported is lower compared to the performance attained by recent deep learning-based pipelines. Also, the use of handcrafted features could restrict the capability of the model to generalize against a wide range of datasets with large intra-class variance. Moreover, the study does not discuss much on dataset size and class imbalance which are important considerations in medical image analysis. However, the paper demonstrates the opportunities of integrating feature engineering and optimization solutions in machine learning-aided skin cancer classification, as well as the necessity of more scalable and representation-oriented models [27].

Ozkan and Koklu (2017) used machine learning to classify dermoscopic images of skin lesions, based on the PH2 dataset of skin lesions (dermoscopic images). The primary aim of the research was to assist the dermatologists with the separation of melanoma and non-melanoma cases through the minimization of subjectivity of the manual diagnosis systems. The authors compared four popular machine learning classifiers, i.e. Artificial Neural Network (ANN), Support Vector machine (SVM), K-Nearest Neighbors (KNN) and Decision Tree (DT) using a collection of extracted lesion features. The effectiveness of these classifiers was measured on 10-fold cross-validation and conventional measures of evaluation like the accuracy, sensitivity and specificity. According to the experimental results, ANN classifier was the most successful with the highest accuracy of 92.5 percent and KNN had the lowest accuracy among the models that were tested [28]. Even though the results received are promising, the research is limited in several ways. The model is based on manual feature extraction that can negatively affect its strength when used on images of other origin. Moreover, the size of the data set used is relatively low and the samples of the lesion types are not evenly distributed especially the melanoma cases as they represent the larger fraction. Moreover, the experiments were performed on one data set that was not validated, which diminishes the validity of the findings in the actual clinical practice. In spite of these drawbacks, the paper still offers an informative comparison of classical machine learning classifiers, as well as a sound foundation of future studies regarding deep learning classifiers or hybrids.

Razia et al. (2025) established an optimized model of automated skin lesion classification with classical machine learning algorithms on three publicly available datasets of skin lesions, namely, DermNet, PH2, and ISIC. Some of the preprocessing procedures that were incorporated in the study to address both class imbalance and large feature space dimensions include, image normalization, Gaussian noise filter, data augmentation techniques (Random Oversampling, SMOTE) and dimensionality reduction procedures (principal component analysis, PCA). The datasets were trained and evaluated using four conventional classifiers namely, Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbors (k-NN), and Logistic Regression. The findings indicated that the highest accuracy of the Random Forest model is 99.3 percent on the ISIC dataset whereas the SVM model worked well with a 93.1 percent and 94.2 percent accuracy on DermNet and PH2 respectively. Though it was a study about machine learning, instead of deep learning, the authors emphasised that the future advancements should consider the use of Convolutional Neural Networks (CNNs) and non-visual aspects of data to enhance the performance of the classification further [29].

Natha & RakaRajeswari (2024) suggested an ensemble learning model, which is designed to enhance the prediction of skin cancer types based on dermoscopic images datasets, including the ISIC 2018 and HAM10000. This study used a hybrid approach to maximize the power of multiple machine learning models, namely AdaBoost, CatBoost, Random Forest, Gradient Boosting, and Extra Trees, and employed a max Voting approach to ensemble, in which each model provides its own contribution to the overall outcome of the final prediction. The best way of creating feature vectors was through a Genetic Algorithm (GA) and then classification which provided better representation and stability. The proposed Max Voting ensemble had a high accuracy of 95.80 percent and better F1 -measure, recall, and precision than the individual ensemble models. This indicates the ability of ensemble techniques in enhancing the reliability of skin cancer lesion classification through the exploitation of strengths in the models and the mitigation of weaknesses in the models respectively. Limitations of diversity to datasets and computational resource requirements were also emphasized by the authors and future work that may investigate hybrid integration with deep learning models to enhance further generalization and performance was proposed [30].

Although machine learning-based methods are effective, they have a number of limitations. The quality of features that are handcrafted and the correct segmentation of the lesion are the most important factors in their performance. Moreover, such approaches tend to be unable to process large and highly heterogeneous data, and so have less ability to extrapolate. Consequently, in spite of machine learning algorithms forming the basis of intelligent skin cancer detection, it has been gradually supplanted and, in most instances, overtaken by deep learning-based methods. Table 1 presents the example machine learning-based methods of automated skin cancer diagnosis, which indicates the methods used, data sets, and their performance.

Table 1 gives comparative details on classical machine learning-based methods used in the diagnosis of skin cancer. In general, the literature reviewed shows that the traditional machine learning models, especially the Support Vector Machines and ensemble-based classifiers can be used to reach the competitive classification performance when supplemented with thoughtful handcrafted features. Methods based on the use of texture, color and shape descriptors driven by dermatological criteria like the ABCD rule reported that the accuracy levels are very acceptable, particularly on small and medium sized datasets.

Table 1 - Summary and Critical Analysis of Classical Machine Learning-Based Methods Specifically Applied to Skin Cancer Diagnosis.

Ref	Authors (Year)	Method / Approach	Dataset(s)	Key Results	Critical Remarks
[27]	Natha & Pothuraju (2023)	Handcrafted features (CT + LBP) + PSO feature selection + SVM, RF, NN	Dermoscopic images	SVM achieved 86.9% accuracy with lower computational complexity	Limited generalization due to handcrafted features; dataset size and class imbalance not sufficiently discussed
[28]	Ozkan & Koklu (2017)	Manual feature extraction + ANN, SVM, KNN, DT	PH2	ANN achieved 92.5% accuracy; KNN showed lowest performance	Small dataset; class imbalance; evaluation limited to a single dataset
[29]	Razia et al. (2025)	Preprocessing + PCA + SVM, RF, KNN, Logistic Regression	DermNet, PH2, ISIC	RF achieved 99.3% (ISIC); SVM achieved 93.1% (DermNet) and 94.2% (PH2)	Relies on classical ML; authors suggest CNNs for future improvement
[30]	Natha & RajaRajeswari (2024)	Ensemble learning (AdaBoost, CatBoost, RF, GB, Extra Trees) + GA + Max Voting	ISIC 2018, HAM10000	Ensemble achieved 95.80% accuracy, improved F1, recall, precision	High computational cost; limited dataset diversity
[24]	Natha et al. (2025)	Ensemble-based ML classifiers	Dermoscopic images	Improved diagnostic accuracy over individual classifiers	Performance dependent on feature quality and dataset characteristics
[21]	Soundarya & Poongodi (2025)	Handcrafted feature fusion + ML/Hybrid models	Skin lesion images	Enhanced feature representation	Generalization limited by handcrafted features
[22]	Atim & Ibrahim (2025)	Rule-based expert system (Backward chaining)	Symptom-based skin diseases	Logical interpretability	Not image-based; limited scalability

Nonetheless, in spite of the encouraging outcomes, some limitations are always witnessed in these studies. Machine learning-based systems are very sensitive to changes in the conditions of image acquisition and the appearance of lesions, and the quality of the handcrafted features and lesion segmentation affects their performance significantly. Furthermore, a number of studies make use of relatively small or unbalanced data and this raises the concern about overfitting and a lack of generalization to clinical practice. It has been demonstrated that ensemble learning and feature optimization methods (e.g., genetic algorithms, particle swarm optimization and dimensionality reduction algorithms e.g., PCA) can enhance the accuracy and robustness of classification. However, these advantages usually require higher levels of computation and lower scalability. Besides, the use of manual feature engineering limits the visual heterogeneous skin lesion data to detect complex and high-level visual patterns by the machine learning models. According to the results summarized in Table 1, it is clear that although classical machine learning solutions provided the basis of automated skin cancer diagnosis, their nature has driven a gradual shift toward models based on deep learning. Deep learning architectures provide end-to-end feature learning abilities, better generalization and scalability to large and heterogeneous datasets, as needed for trusted clinical implementation.

Further studies are needed to minimize the reliance on handcrafted features through the use of the data-driven representation learning techniques. Moreover, standardized assessment procedures, increased multi-center data, and powerful management of class imbalance should be considered to guarantee the fair comparison and clinical reliability. Having hybrid frameworks with machine learning interpretability and deep learning representation power could be a promising way forward to creating accurate and reliable skin cancer diagnostic frameworks. It should be mentioned that the apparent temporal difference between 2017 and 2023 in Table 1 is a shift in the methodology used in the research community as opposed to underdevelopment. This shift in research focus caused a relative drop in independent studies of classical machine learning in the skin cancer classification task to deep learning and transfer learning architectures. Thus, the table exemplifies current and recent optimized studies on the ML to demonstrate the development of the methodology.

4.2. Deep learning algorithms for skin cancer diagnosis

Over the last few years, deep learning systems have been incredibly successful in the analysis of medical images, especially skin cancer. In contrast to the traditional machine learning methods, which use hand-crafted features, deep learning models discover hierarchical features representations directly out of raw images. This is possible and enables them to be more useful in managing complicated visual patterns, including color, texture, and shape changes that are popular on the skin lesions.

A number of researches have been dedicated to the design of custom CNNs to diagnose skin cancer. An example is the framework suggested by Dorj et al. (2018), which is built based on CNN to classify both melanoma and non-melanoma lesions, demonstrating encouraging accuracy on publicly available data. Their findings legitimized that deep learning models are capable of extracting discriminative features without necessarily segmenting them or hand-engineering features [31].

Other scholars have delved in more extensive architectures in order to improve the performance of classification. Yu et al. (2019) used a deep CNN model, which consists of several convolutional layers, and identified a higher accuracy with this type of network than with shallow networks tested on the same data. Nevertheless, they also said that more complex models need larger data and increased computing capabilities that can limit their use in a real-time clinical setting [32].

Mohammed & Inik (2024) suggested a convolutional neural network (CNN)-based framework to classify skin cancer in multiple classes based on the ISIC dermoscopic image data. The authors, besides, applied a custom CNN architecture, as well as tested four popular deep CNNs (ResNet, GoogleNet, AlexNet, and VGG16) on image resolutions (64x64, 100x100, 128x128, and 224x224 pixels). The experiment showed that the proposed CNN had the best classification rate of 86.76 per cent on 128x128 images and this implied that the proposed CNN was competitive with other tested models and proved that deep learning models could be used to differentiate between various lesions. Data augmentation was also used to balance data in the study and extensive preprocessing was used to increase generalization. The contribution of this work to the field is that it provides an assessment of both custom and existing CNN architecture on a standardized skin cancer dataset but its performance is moderate compared to other recent and more advanced deep learning architectures. Secondly, although the study examines various resolutions of images, more gains could be achieved by using transfer learning with bigger pretrained backbones or by adding attention to help identify finer variations of lesions [33].

Badawi et al. (2024) examined the deep learning-based methods of classifying and segmenting skin cancer based on the dermoscopic image data, such as HAM10000 and PH2. Their methodology includes a lot of preprocessing and data augmentation to enhance model generalization, a convolutional neural network (CNN) to classify lesions and a U-Net segmentation model to precisely draw the lesion boundaries. The CNN model was tested by accuracy, classification reports, and confusion matrices, in addition to Dice coefficient and Intersection over Union (IoU) that indicated the high efficiency of CNN-based deep learning methods to improve diagnostic accuracy in skin cancer assignments [34].

Ozdemir and Pacal (2025) suggested a powerful hybrid deep learning system to multiclassify skin cancer based on the combination of sophisticated convolutional blocks and attention to maximize the features representation and the diagnosis. They combine ConvNeXtV2 for a fine-grained local features capture, separable self-attention layers to learn global contextual information. This model was trained and tested on the ISIC 2019 dataset of eight different lesion types and achieves a high performance of 93.48% accuracy, 93.24% precision, 90.70% recall and 91.82% F1-score, outperforming more than 20 other state-of-the-art deep learning models under similar conditions. The design is also computationally efficient but has only a few parameters, namely, the number of parameters is approximately 21.9 million, which is acceptable in terms of high performance and feasible deployment capabilities [35].

In the article, Hussein et al. (2025) introduced a new hybrid deep learning model based on a hybrid quantum convolutional neural network (HQCNN), bidirectional long short-term memory network (BiLSTM), and MobileNetV2 to sort skin cancer images. The model was tested on a clinically relevant dermoscopic dataset and obtained a training accuracy of 97.7 and test accuracy of 89.3 and high F1-score and recalls on malignant lesion detection. This shows that the combination of hierarchical feature extraction, sequential contextual learning and high-performing convolutional backbones can generate better results when handling intricate medical image classification tasks [36].

Nawaz et al. (2025) proposed a deep learning-based system to detect the presence of skin cancer with dermoscopic images and their variations of subtle lesions that conventionally restrict machine learning methods. Their work suggested a new FCDS-CNN, the architecture that is specifically designed to analyze the dermoscopic images and consists of data augmentation and class weighting methods to reduce the number of the majority classes and enhance the presence of the under-represented types of lesions. It was trained on a big dataset of 10,015 dermoscopic images that represented seven types of lesions in a publicly available Kaggle repository. The FCDS-CNN was found to be the most accurate, achieving a total accuracy of around 96, compared to a range of common pretrained models (e.g., ResNet, EfficientNet, Inception, MobileNet) on multiple important performance metrics such as precision, recall, F1-score, and area under the ROC curve. This paper has shown that custom deep learning architecture design, which explicitly addresses the imbalance of data and feature complexity of skin lesion datasets, can be effective. Nevertheless, even with impressive performance, the paper continues to point at typical issues that are associated with deep learning-based skin cancer diagnosis, in particular, the requirement of large, heterogeneous training data and attending preprocessing to address intra-class variance. The results support the tendency to create more sturdy and flexible CNN models to be used in real-world clinical applications, particularly in the context of early melanoma detection when the specificity and sensitivity are of high importance [3].

Mohamed et al. (2025) have created a new deep learning-based melanoma classification system that involves Convolutional Neural Networks (CNNs) and Aquila Optimizer (AO) to reduce the dimensionality of the features. This hybrid model is designed to improve computational efficiency and classification accuracy, which can help in solving dilemmas that are common to the diagnosis of skin cancer, including high dimensionality of features and limited resources. The proposed approach was tested using three popular dermoscopic image datasets, i.e., ISIC 2019, ISBI 2016, and ISBI 2017 and is shown to be exceptionally high in accuracy: 98.42% (ISIC 2019), 97.46% (ISBI 2016), and 98.89 (ISBI 2017), sensitivity, 97.91 precision, 97.68 F1-score, and 99 Other comparable good outcomes were found in the ISBI datasets, with accuracy values of close to 98, with high sensitivity and specificity scores. Also, in conjunction with the Aquila Optimizer, the framework substantially decreased computational complexity (by up to 37.5%), which made it very well appropriate to be deployed in resource-constrained settings, e.g. mobile or edge computing devices. This paper presents a benefit of using a combination of deep learning feature extraction and metaheuristic optimization that will enhance the accuracy and efficiency of detecting melanoma [37].

Vincent, Darian & Surantha (2025) compared the effectiveness and realistic functionality of a Convolutional Neural Network (CNN)-based model of skin cancer detection when used on a resource-constrained platform (e.g., edge computing device, e.g. Raspberry Pi and NVIDIA Jetson Nano). In contrast to most studies which only consider the accuracy of classification in high-performance computing clusters, the paper trained a CNN to learn with a large dataset of dermoscopic images (more than 10,000 of the Harvard Skin Lesion dataset) with a variety of lesion types, and then tested the model on low-power hardware to emulate actual performance in the field in distant regions or underserved areas. It has been demonstrated that the model was able to reach a maximum 98.25 percent classification accuracy, and that the Raspberry Pi 5 only needed a short detection time (~0.01 s) without using much energy or consuming a lot of memory[38].

However, even with deep learning model performance, there are a number of challenges. Most studies use small or unbalanced data sets, and this could result in overfitting and diminished generalization. Furthermore, deep learning models in many aspects consume a lot of computation and take long to train. These constraints lead to the necessity of lightweight and efficient architectures that can be both diagnostic and deployment friendly. Table 2 provides a summary of recent deep learning-based methods of skin cancer diagnosis.

Table 2 - Summery of the latest deep learning-based skin cancer diagnosis systems, including the architecture used and the performance.

Ref	Study	Dataset	Model / Method	Task	Key Results	Critical Remarks
[31]	Dorj et al., 2018	Public dermoscopic dataset	Custom CNN	Melanoma vs non-melanoma	Encouraging accuracy; effective feature learning without segmentation	Effective feature learning is demonstrated; however, the binary setting and limited data reduce generalizability.
[32]	Yu et al., 2016	Dermoscopy images	Deep Residual CNN	Melanoma classification	Higher accuracy than shallow CNNs; higher computational cost	Deep residual networks improve accuracy but suffer from high computational complexity.
[33]	Mohammed & İnik, 2024	ISIC	Custom CNN + ResNet, GoogLeNet, AlexNet, VGG16	Multi-class classification	Best accuracy 86.76% (128×128); competitive but moderate	Provides a useful comparison, yet performance remains moderate without advanced architectures.
[34]	Badawi et al., 2024	HAM10000, PH2	CNN + U-Net	Classification + Segmentation	High accuracy, Dice & IoU indicate strong segmentation performance	Joint classification and segmentation improve reliability but increase computational overhead.
[35]	Ozdemir & Pacal, 2025	ISIC 2019	ConvNeXtV2 + Self-Attention	Multi-class classification (8 classes)	Accuracy 93.48%, F1-score 91.82%, ~21.9M parameters	Achieves strong performance with attention mechanisms, though validation is limited to one dataset.
[36]	Hussein et al., 2025	Dermoscopic dataset	HQCNN + BiLSTM + MobileNetV2	Skin cancer classification	Test accuracy 89.3%, strong malignant detection	Hybrid quantum–deep model is promising but complex for practical clinical deployment.
[37]	Mohamed et al., 2025	ISIC 2019, ISBI 2016/2017	CNN + Aquila Optimizer	Melanoma classification	Accuracy up to 98.89%, reduced complexity by 37.5%	Optimization-enhanced CNN improves accuracy and efficiency,

						requiring broader validation.
[38]	Darian & Surantha, 2025	Harvard Skin Lesion	CNN on Edge Devices	Skin cancer detection	Accuracy 98.25%, fast inference (~0.01 s)	Demonstrates feasibility on edge devices but lacks robustness analysis under varied conditions.

Table 2 reveals that the models of deep learning have proved highly effective in diagnosing skin cancer, especially in terms of their ability to detect the intricate visual patterns without the use of handcrafted features. The initial research presented on custom CNN architectures [31] and then later research implemented more elaborate and advanced models to enhance the accuracy of classification [32]. Much more recent work focuses on attention-based and hybrid architectures, including ConvNeXtV2 with self-attention [35] and CNN-optimizer hybrids [37], which attain over 93% accuracy on large-scale data.

Although high performance is reported, the results differ significantly with the size of the datasets, the balance of classes and the evaluation protocols. Incorporating segmentation [34] or optimization strategies [37] into studies also lead to improved accuracy of the diagnostic, and literature concentrating on edge deployment [38] emphasizes that computational efficiency is essential to implement diagnostic in the real world. However, overfitting, data imbalance, and high computational cost still persist, and the necessity to find lightweight but strong deep learning models emerges.

4.3. Skin Cancer Diagnosis using Transfer Learning Models

Advanced machine learning algorithms have also been applied to the automated detection of skin cancer, especially with the transfer learning models which have been developed incredibly fast [39][40]. The reason behind this growing popularity of this method in dermatological image analysis is its capacity to realize high classification performance when annotated medical datasets are small enough [41]. Transfer learning can be used to exploit knowledge gained using large-scale visual data to transfer the convolutional neural networks to new tasks, which involve skin lesion classification [42].

A number of studies have proved the usefulness of transfer learning models in the diagnosis of melanoma and skin lesions. VGG16, ResNet, Inception, DenseNet, MobileNet, Xception, and EfficientNet are among the most popular architectures that have been used widely because of their good feature extraction properties and converse to the stable convergence [43][44][45]. Such models are usually modified using the end classification layers so that the model can fit the category of skin cancer of interest, enabling it to be trained effectively with less training time and lower computational cost [46][47].

According to recent studies, lightweight models, such as MobileNetV2 and EfficientNet variants, offer a good trade-off between accuracy and computational efficiency and are thus applicable in real-time and resource-intensive clinical practice [48]. Alternatively, an in-depth architecture including ResNet and DenseNet has proven to be more superior in capturing fine-scale visual patterns related to the malignant lesion especially in complex heterogeneous datasets [49][50]. The literature on classification accuracy has reported reported classification accuracy of 85% or higher with certain variations based on the size of the dataset, class balance, and evaluation protocol [51].

Searching refers to the approaches based on transfer learning, although they have achieved success, there are several issues associated with this methodology [38]. Numerous researches work with small or unbalanced datasets, which may result in overfitting and decreased extrapolation to unknown groups of people [17]. Moreover, more complicated or more complex architectures are frequently implemented to improve performance, raising the cost of computation and practical usefulness. Moreover, the differences in preprocessing methods, dataset formation as well as evaluation measures render it challenging to compare studies one-on-one [52][53].

Recently, Alotaibi and AlSaeed (2025) performed research on transfer learning using deep attention mechanisms through Xception architecture binary classification of skin lesions (benign and malignant). The authors tested the impact of including various attention modules of self-attention, hard attention and soft attention to the Xception model on the use of HAM10000 dermoscopic image dataset. Findings indicated that the inclusion of attention mechanisms always produced better results than the baseline Xception model with accuracy ranging between 91.05

(base Xception) and 94.11 (self-attention) and higher than 92.91 (other types of attention), and the recall metrics which are important in medical diagnostics. This paper underscores the possibility of integrating transfer learning and attention system to increase the accuracy of diagnosis of skin cancer further [54].

To enhance automatic classification of skin cancer using dermoscopic images, Aboulmira et al. (2025) suggested a hybrid deep learning system, which combines the use of wavelet decomposition and an EfficientNet backbone, which is an effective model in image classification. This hybrid model uses wavelet transforms to break down images into multi-scale frequency features after running them through an EfficientNet convolutional network, which improves the representation of features and resistance. The proposed method was tested on publicly available datasets of HAM10000 and ISIC2017 and produced high accuracy, 94.7% on HAM10000 and 92.2% on ISIC2017, showing that transfer-learning based CNN models with traditional signal processing methods such as wavelets can be used with a high level of accuracy, as well as cope with such challenges as variability of data and imbalance in classes[55].

Khan et al. (2025) have developed a complex computer-aided diagnostic model of automatic classification of the melanoma and non-melanoma skin cancer using a two-stage transfer learning model utilizing a fine-tuning deep convolutional neural network (CNNs). The paper overcomes key issues in training deep models with small dermoscopic data sets including overfitting and high computation cost through an adaptive learning approach that refines the existing CNN layers through the skin lesion regions of interest (ROIs). The proposed approach initially re-initiates and optimizes the bottom-up layers to learn the lesion-relevant visual features and finally replaces the fully connected layers with a principal component analysis (PCA)-based layer. The integration assists in the extraction of discriminative global features, and will serve to overcome overfitting, and increase model generalization. The system was tested on benchmark sets of dermoscopy images and demonstrated good potential of detecting skin cancer early and efficiently. This study points out that adaptive fine-tuning of transfer learning in conjunction with the unsupervised feature extraction models such as PCA can large-scale enhance the skin cancer classification through CNN-based learning, which can be used effectively to guide clinicians to make decisions on time and enhance patient outcomes [56].

Guler and Agraz (2025) performed an in-depth experiment comparing the performance of some transfer learning models on binary (benign vs malignant) and multiclass classification of skin cancer using the ISIC 2019 and ISIC 2020 dermoscopic databases. The authors examined the effect of the use of preprocessing methods (e.g., DullRazor, histogram equalization, gamma correction) and data augmentation, and hyperparameter optimization on the performance of the classifier, and compared six popular transfer learning architectures including VGG16, ResNet50, InceptionV3, Xception, DenseNet121, and MobileNetV2.

Findings revealed that transfer learning models can be optimally preprocessed and optimized in terms of hyperparameters. In binary classification, VGG16 model recorded the best accuracy of 90.17 post tuning and DenseNet121 and ResNet50 also performed well. VGG16 was once again the best in the multiclass environment with 92.92 percent accuracy over other models. The results of all trained transfer learning models played out better accuracy post-optimization than in their default settings, which points to the importance of image preprocessing and parameter modification in skin cancer diagnosis tasks. The paper notes that with properly design transfer learning architectures, coupled with an appropriate preprocessing scheme coupled with tuning method, high classification can be attained in a wide range of lesion classes. Also, the authors observe that class imbalances continue to pose problems in multiclass classification, both in terms of larger and more balanced datasets, as well as more advanced means of augmentation, to continue to improve model robustness [57].

On the whole, it is possible to state that transfer learning models have become a potent and feasible choice when it comes to skin cancer diagnostics. Still, systematic assessment, lightweight model development, and effective validation on various datasets are still required to make sure that the model can be introduced into a real clinical environment. Table 3 has identified some of the studies that have use of transfer learning architectures to classify skin cancer using various datasets and test conditions. Transfer learning has emerged to be one of the most popular approaches to skin cancer diagnosis, especially when an annotated medical data is scarce, as shown in Table 3.

Table 3 - Summery of the latest deep learning-based skin cancer diagnosis systems, including the architecture used and the performance.

Ref	Study	Dataset	Transfer Learning Model	Task	Key Results	Critical Remarks
[39]	Karboua et al., 2025	Dermoscopic images	Pretrained CNN-based TL	Skin cancer classification	Improved accuracy with reduced training data	Demonstrates effective use of transfer learning, but validation is limited to a single dataset.
[40]	Remya et al., 2024	Multimodal skin lesion data	Transfer learning framework	Lesion classification	Robust performance across modalities	Multimodal fusion improves robustness; however, increased complexity may hinder deployment.
[43]	Gairola et al., 2022	Skin lesion dataset	VGG16, ResNet, Inception	Binary classification	Strong feature extraction and stable convergence	Classical pretrained models perform well, yet lack adaptation to recent architectures.
[47]	Spolaor et al., 2024	Small clinical datasets	Fine-tuned pretrained CNNs	Medical image classification	Effective learning with limited data	Effective for small datasets, though scalability to large cohorts is not addressed.
[48]	Al-Azad & Ahmmed, 2025	Dermoscopic images	MobileNetV2 + Attention	Melanoma detection	High accuracy with low computational cost	Achieves high accuracy with low cost, but evaluation is restricted to melanoma detection only.
[50]	Eliwa, 2025	Dermoscopic datasets	Two-phase TL models	Skin cancer classification	Improved generalization and fine-grained detection	Two-phase transfer learning improves generalization; however, training overhead increases.
[54]	Alotaibi & AlSaeed, 2025	HAM10000	Xception + Attention	Benign vs malignant	Accuracy up to 94.11%	Attention mechanism enhances discrimination, but tested on binary classification only.
[55]	Aboulmira et al., 2025	HAM10000, ISIC2017	EfficientNet + Wavelet	Multi-class classification	94.7% (HAM10000), 92.2% (ISIC2017)	Hybrid wavelet–DL approach improves accuracy, though preprocessing complexity is high.
[56]	Khan et al., 2025	Dermoscopy benchmarks	Fine-tuned CNN + PCA	Melanoma / non-melanoma	Improved generalization, reduced overfitting	Feature reduction reduces overfitting; generalizability across datasets needs validation.
[57]	Güler & Ağraz, 2025	ISIC 2019 & 2020	VGG16, ResNet50, DenseNet121	Binary & Multiclass	Up to 92.92% accuracy after optimization	Comprehensive comparison provided, yet performance varies across datasets and tasks.

The common examples of the popular pretrained architectures, like VGG16, ResNet, DenseNet, Xception, and EfficientNet, exhibit the high feature extraction performance and convergence stability. MobileNetV2 and EfficientNet are lightweight models that offer a reasonable tradeoff between classification accuracy and computational efficiency and are therefore applicable in real-time and resource-intensive clinical settings.

Recent literature also can provide additional improvements in the performance of transfer learning through the incorporation of attention mechanisms [54], wavelet-based feature decompositions [55], or adaptive fine-tuning plans and dimensionality reduction methods [56]. This type of hybrid methods typically give an accuracy in classification above 90 percent, with other models obtaining above 94 percent of the accuracy on several benchmark datasets like HAM10000 and ISIC.

Although these are encouraging findings, Table 3 also demonstrates that there are significant drawbacks in the literature. Most studies use small or unbalanced datasets which predisposes overfitting and reduces extrapolation to unknown populations. Also, differences in preprocessing pipelines, data augmentation schemes and evaluation protocols make it difficult to directly compare models. In turn, standardized evaluation frameworks, lightweight architecture design, and validation across a wide range of heterogeneous datasets should be given priority in the future research to support the reliable clinical implementation.

5. A Comparative Analysis: AI Algorithms For Skin Cancer Diagnosis

Skin cancer diagnosis artificially intelligence methods may be divided into classical machine learning (ML), deep learning (DL), and transfer learning (TL)-based classification methods. These paradigms are not just different in terms of accuracy but different in terms of feature representation strategy, computational demands, scalability, and whether they can be deployed in clinical settings or not.

Classical ML approaches are based on manually designed feature extractors and regular classification pipelines like SVM, RF, KNN and ensemble models [27] [28] [29] [30]. Although these methods perform well when optimized (i.e. RF with as many as 99.3% on ISIC) [29] [30], feature engineering quality and homogeneity of the dataset are crucial in their success [27] [28]. Their key benefit is that they are interpretable and less demanding.

Hierarchical representations of dermoscopic images are automatically extracted using deep learning architectures, especially, CNN-based networks, including ResNet, VGG16, EfficientNet, and hybrid attention-based networks [31] [32] [33] [35] [37]. These models, on the whole, are stronger with complicated visual patterns and it is reported that the accuracy is as high as 98.42 percent with optimization strategies [35] [37]. Nonetheless, they take a lot of annotated data and a lot of computing power and are sensitive to imbalance in data as well as preprocessing changes[3] [35].

Transfer learning uses the pretrained networks such as VGG16, ResNet, DenseNet, EfficientNet, and MobileNet to identify skin lesions [38] [48] [52][53] [54] [55][57]. Hybrid enhancement and fine-tuning algorithms have the important benefit of enhancing the classification performance, usually by a range of 90%-94% in multi-class classification [54] [55][57]. The lightweight architectures provide a better feasibility to be deployed to the resource-constrained clinical settings [48]. However, overfitting can be found in small or skewed data sets [38] [52].

Generally, the main difference between these paradigms is not just in performance measures but also in data dependency, model complexity as well as practicality in deployment that should be taken into account when designing clinically applicable diagnostic systems.

6. Performance Analysis of AI Models in Skin Cancer Classification

Comparison of studies performance evaluation indicates that significantly different performance metrics have been reported with accuracy varying around 86 to 99 percent with some variation depending on the characteristics of the dataset and assessment conditions [3] [29] [30] [35] [37]. Variations in preprocessing pipelines, strategy of augmentation, and validating protocols have a great impact on the reported results.

Competitive performance of machine learning models can be achieved along with feature optimization and dimensionality reduction methods like PCA, PSO and GA [29] [30]. Their extrapolation ability can however, weaken when using heterogeneous or skewed data [29] [30].

Deep learning models exhibit a high ability to learn difficult lesion patterns, especially when complemented with mechanisms of attention and a combination of optimization methods [35] [36] [37]. However, their performance is still very conditional on the size of datasets, augmentation strategies, and computing resources [3] [35].

The models of transfer learning offer consistent performance in the moderate or small data conditions, particularly with the application of fine-tuning and preprocessing approaches [54] [55] [57]. Nevertheless, there are overfitting threats and inter-study evaluation methods that make it difficult to directly compare performance [38] [52].

Notably, the absence of standardized benchmarking models in datasets including ISIC, HAM10000, PH2 and DermNet also restricts objective comparison of cross-studies [29][30] [54]. The next direction of future studies must focus on systematic validation procedures, lightweight designs and repeatable evaluation pipelines to increase clinical trustworthiness.

7. Conclusion and Recommendations

This comparative and analytical paper has explored a vast number of machine learning, deep learning, and transfer learning-based solutions to skin cancer diagnosis and made conclusions through quantitative performance comparison and methodology analysis. Through the comparative tables, there is an apparent development in the diagnostic ability, where classical machine learning models, in spite of being computationally effective and understandable, are heavily inhibited by manual feature dependency and the lack of generalization in heterogeneous data sets. Deep learning models are able to perform better in representation learning and higher classification accuracy through automatic extraction of hierarchical features of dermoscopic images. Nevertheless, the review shows that deep architectures that are not coupled typically require large annotated datasets in order to perform effectively. and computationally demanding, and thus, are not always scalable and cannot be applied to real-life clinical scenarios. Transfer learning models always have an adequate ratio between accuracy and efficiency especially in situations where there is scant medical data. The analyzed literature demonstrates that pretrained CNN backbones, particularly, EfficientNet, ResNet, DenseNet, and MobileNet variants, can be used more effectively than regular machine learning and regular CNN models in case of the appropriate fine-tuning and preprocessing approaches. However, the differences in datasets, evaluation procedures, and preprocessing pipelines do not allow comparing the performance of studies and recreating them. Moreover, comparative analysis shows that the hybrid and ensemble strategies, which also involve attention mechanisms, optimization of features, and fusion of models add to the enhanced robustness and diagnostic stability. Although this has been achieved, there are still many challenges that cannot be ignored like class imbalance, overfitting and more advanced computational complexity. All in all, it can be concluded that there is no best model architecture that is all-purpose in skin cancer diagnosis. Rather, future studies ought to focus on standardized assessment platforms, lightweight but effective architectures, and cross-dataset validation to have credible, clinically implementable skin cancer detection systems that can facilitate real-life dermatological decision-making. According to the comparative analysis, the transfer learning models are suggested to be used in cases of data limits as they are highly accurate with smaller training. Lightweight models such as MobileNet and EfficientNet are applicable to real-time or resource-constrained based systems, whereas deeper networks may be applied where there is no resource constraint. The challenge of solving the dataset including class imbalance and standardized preprocessing and augmentation help to increase robustness. The coherent evaluation guidelines are a necessity to compare and deploy the protocols to practice.

References

[1] R. Hamsalekha, G. D. George, and T. Y. Satheesha, “A Novel Deep Learning Approach for Automated Melanoma Classification using Hybrid CNN and Vision Transformer Model,” *Fusion Pract. Appl.*, vol. 19, no. 2, pp. 92–101, 2025, doi: 10.54216/FPA.190207.

[2] K. Mridha, M. M. Uddin, J. Shin, S. Khadka, and M. F. Mridha, “An interpretable skin cancer classification using optimized convolutional neural network for a smart healthcare system,” *IEEE Access*, vol. 11, pp. 41003–41018, 2023.

[3] K. Nawaz *et al.*, “Skin cancer detection using dermoscopic images with convolutional neural network,” *Sci. Rep.*, vol. 15, no. 1, p. 7252, 2025.

[4] V. Dillshad, M. A. Khan, M. Nazir, O. Saidani, N. Alturki, and S. Kadry, “D2LFS2Net: Multi-class skin lesion diagnosis using deep learning and variance-controlled Marine Predator optimisation: An application for precision medicine,” *CAAI Trans. Intell. Technol.*, vol. 10, no. 1, pp. 207–222, 2025.

[5] K. Razzaq and M. Shah, “Machine learning and deep learning paradigms: From techniques to practical applications and research frontiers,”

Computers, vol. 14, no. 3, p. 93, 2025.

[6] Z. C. Oleiwi, E. N. AlShemmary, and S. Al-Augby, "Multi-label Classification Technique of Chest X-Rays Image Based Cardiomegaly Disease Prediction," in *2023 International Conference on Information Technology, Applied Mathematics and Statistics (ICITAMS)*, IEEE, 2023, pp. 134–139.

[7] U. K. Lilhore *et al.*, "SkinEHDLF a hybrid deep learning approach for accurate skin cancer classification in complex systems," *Sci. Rep.*, vol. 15, no. 1, p. 14913, 2025.

[8] P. Tschandl, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Harvard Dataverse*. doi: doi:10.7910/DVN/DBW86T.

[9] M. Shaheer, T. Arooj, H. Adeel, T. Saleem, and I. R. Rao, "AI-based skin cancer detection algorithms: opportunities, challenges and a way forward," *Asian J. Sci. Eng. Technol.*, vol. 4, no. 1, pp. 87–109, 2025.

[10] V. Rotemberg *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Sci. data*, vol. 8, no. 1, p. 34, 2021.

[11] J. Amin, M. Azhar, H. Arshad, A. Zafar, and S.-H. Kim, "Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and transformer neural networks," *Front. Med.*, vol. 12, p. 1524146, 2025.

[12] S. A. Amiri, M. Nasrolahzadeh, Z. Mohammadpoory, and A. H. Z. Kordkheili, "Skin Lesion Classification via ensemble method on deep learning," *Multimed. Tools Appl.*, vol. 84, no. 18, pp. 19379–19397, 2025.

[13] H. Eghtesaddoust, M. Valizadeh, and M. C. Amirani, "Fusion of Deep and Time–Frequency Local Features for Melanoma Skin Cancer Detection," *Appl. Comput. Intell. Soft Comput.*, vol. 2025, no. 1, p. 4767052, 2025.

[14] A. T. P. Nguyen, R. M. Jewel, and A. Akter, "Comparative analysis of machine learning models for automated skin cancer detection: Advancements in diagnostic accuracy and ai integration," *Am. J. Med. Sci. Pharm. Res.*, vol. 7, no. 01, pp. 15–26, 2025.

[15] H. Vega-Huerta *et al.*, "Classification Model of Skin Cancer Using Convolutional Neural Network," *Ing. des Syst. d'Information*, vol. 30, no. 2, p. 387, 2025.

[16] A. Raza, A. Ali, S. Ullah, Y. N. Anjum, and B. Rehman, "Optimizing skin cancer screening with convolutional neural networks in smart healthcare systems," *PLoS One*, vol. 20, no. 3, p. e0317181, 2025.

[17] M. Gholizade, H. Soltanizadeh, M. Rahmanimanesh, and S. S. Sana, "A review of recent advances and strategies in transfer learning," *Int. J. Syst. Assur. Eng. Manag.*, pp. 1–40, 2025.

[18] S. Mavaddati, "Skin cancer classification based on a hybrid deep model and long short-term memory," *Biomed. Signal Process. Control*, vol. 100, p. 107109, 2025.

[19] A. G. Diab, E.-S. M. El-Kenawy, N. F. F. Areed, H. M. Amer, and M. El-Seddek, "A metaheuristic optimization-based approach for accurate prediction and classification of knee osteoarthritis," *Sci. Rep.*, vol. 15, no. 1, p. 16815, 2025.

[20] M. Juneja, N. Aggarwal, S. K. Saini, S. Pathak, M. Kaur, and M. Jaiswal, "A comprehensive review on artificial intelligence-driven preprocessing, segmentation, and classification techniques for precision furcation analysis in radiographic images," *Multimed. Tools Appl.*, vol. 84, no. 19, pp. 21467–21520, 2025.

[21] B. Soundarya and C. Poongodi, "A novel hybrid feature fusion approach using handcrafted features with transfer learning model for enhanced skin cancer classification," *Comput. Biol. Med.*, vol. 190, p. 110104, 2025.

[22] S. B. Atim and M. Y. I. I. Ibrahim, "Rule-Based Expert System Model with Backward Chaining Algorithm for Symptom-Based Skin Disease Diagnosis," *J. Teknol. dan Open Source*, vol. 8, no. 1, pp. 288–294, 2025.

[23] K. Niwa, K. Yane, Y. Saito, K. Hashikawa, and T. Nozaki, "Development of a Support Vector Machine-based Automated Facial Le Fort I Osteotomy Robot," 2025.

[24] P. Natha, S. P. Tera, R. Chinthaginjala, S. O. Rab, C. V. Narasimhulu, and T. H. Kim, "Boosting skin cancer diagnosis accuracy with ensemble approach," *Sci. Rep.*, vol. 15, no. 1, p. 1290, 2025.

[25] O. N. Oyelade, E. F. Aminu, H. Wang, and K. Rafferty, "An adaptation of hybrid binary optimization algorithms for medical image feature selection in neural network for classification of breast cancer," *Neurocomputing*, vol. 617, p. 129018, 2025.

[26] Z. I. Kalantan, L. S. Alharbi, M. H. Al-Zahrani, and S. M. S. Binhimd, "Robust Dimensionality Reduction: A Bootstrap-Based Evaluation of PCA with Applications in Nutritional and Environmental Sciences," *Contemp. Math.*, pp. 923–942, 2025.

[27] P. NATHA and R. POTHURAJU, "Skin cancer detection using machine learning classification models," 2023.

[28] I. A. Ozkan and M. Koklu, "Skin lesion classification using machine learning algorithms," *Int. J. Intell. Syst. Appl. Eng.*, vol. 5, no. 4, pp. 285–289, 2017.

[29] S. Razia, S. Balakrishnan, M. A. Hussain, and P. Chakrabarti, "An Optimized Classification Framework for Skin Lesion Detection Using Machine Learning," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 6, pp. 29605–29609, 2025.

[30] P. Natha and P. RajaRajeswari, "Advancing skin cancer prediction using ensemble models," *Computers*, vol. 13, no. 7, p. 157, 2024.

[31] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimed. Tools Appl.*, vol. 77, no. 8, pp. 9909–9924, 2018.

[32] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imaging*, vol. 36, no. 4, pp. 994–1004, 2016.

[33] B. Mohammed and Ö. İnik, "Using Deep Learning Architectures For Skin Cancer Classification," *Celal Bayar Üniversitesi Fen Bilim. Derg.*, vol. 20, no. 4, pp. 82–91, 2024.

[34] M. Badawi *et al.*, "Skin cancer classification and segmentation using deep learning," *Int. J. Telecommun.*, vol. 4, no. 01, pp. 1–23, 2024.

[35] B. Ozdemir and I. Pacal, "A robust deep learning framework for multiclass skin cancer classification," *Sci. Rep.*, vol. 15, no. 1, p. 4938, 2025.

[36] A. A. Hussein, A. M. Montaser, and H. A. Elsayed, "Skin cancer image classification using hybrid quantum deep learning model with BiLSTM and MobileNetV2," *Quantum Mach. Intell.*, vol. 7, no. 2, p. 66, 2025.

[37] J. Mohamed, N. S. Tezel, J. Rahebi, and R. Ghadami, "Melanoma Skin Cancer Recognition with a Convolutional Neural Network and Feature Dimensions Reduction with Aquila Optimizer," *Diagnostics*, vol. 15, no. 6, p. 761, 2025.

[38] G. Darian and N. Surantha, "Performance Evaluation of Convolutional Neural Network (CNN) for Skin Cancer Detection on Edge Computing Devices," *Appl. Sci.*, vol. 15, no. 6, p. 3077, 2025.

[39] S. Karboua, F. Harrag, S. Karboua, A. Booutadjine, and M. Deriche, "A Deep Learning and Transfer Learning-Based Application for Skin Cancer Classification," in *2025 IEEE 22nd International Multi-Conference on Systems, Signals & Devices (SSD)*, IEEE, 2025, pp. 745–750.

[40] S. Remya, T. Anjali, and V. Sugumaran, "A novel transfer learning framework for multimodal skin lesion analysis," *IEEE Access*, vol. 12, pp. 50738–50754, 2024.

[41] P. Singh and J. Cirrone, "A data-efficient deep learning framework for segmentation and classification of histopathology images," in *European Conference on Computer Vision*, Springer, 2022, pp. 385–405.

[42] A. W. Salehi *et al.*, "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023.

[43] A. K. Gairola, V. Kumar, and A. K. Sahoo, "Exploring the strengths of pre-trained CNN models with machine learning techniques for skin cancer diagnosis," in *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, IEEE, 2022, pp. 1–6.

[44] V. A. O. Nancy, P. Prabhavathy, M. S. Arya, and B. S. Ahamed, "Comparative study and analysis on skin cancer detection using machine learning and deep learning algorithms," *Multimed. Tools Appl.*, vol. 82, no. 29, pp. 45913–45957, 2023.

[45] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep convolutional neural networks in medical image analysis: A review," *Information*, vol. 16, no. 3, p. 195, 2025.

[46] F. Olayah, E. M. Senan, I. A. Ahmed, and B. Awaji, "AI techniques of dermoscopy image analysis for the early detection of skin lesions based on combined CNN features," *Diagnostics*, vol. 13, no. 7, p. 1314, 2023.

[47] N. Spolaor *et al.*, "Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets," *Multimed. Tools Appl.*, vol. 83, no. 9, pp. 27305–27329, 2024.

[48] S. Al-Azad and S. Ahmmed, "Melanoma Detection with MobileNetV2 and Channel Attention: Addressing the Diverse Skin Tone Challenge," in *2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)*, IEEE, 2025, pp. 1269–1274.

[49] S. S. Zareen, G. Sun, M. Kundi, S. F. Qadri, and S. Qadri, "Enhancing Skin Cancer Diagnosis with Deep Learning: A Hybrid CNN-RNN Approach.," *Comput. Mater. Contin.*, vol. 79, no. 1, 2024.

[50] E. H. I. Eliwa, "Enhancing Skin Cancer Diagnosis Through Fine-Tuning of Pretrained Models: A Two-Phase Transfer Learning Approach," *Int. J. Breast Cancer*, vol. 2025, no. 1, p. 4362941, 2025.

[51] S. Mustafa, A. Jaffar, M. Rashid, S. Akram, and S. M. Bhatti, "Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet-Random Forest for enhanced segmentation and classification," *PLoS One*, vol. 20, no. 1, p. e0315120, 2025.

[52] A. R. Khan, M. Mujahid, F. S. Alamri, T. Saba, and N. Ayesha, "Early-Stage Melanoma Cancer Diagnosis Framework for Imbalanced Data From Dermoscopic Images," *Microsc. Res. Tech.*, vol. 88, no. 3, pp. 797–809, 2025.

[53] A. K. W. Lee *et al.*, "Artificial intelligence application in diagnosing, classifying, localizing, detecting and estimation the severity of skin condition in aesthetic medicine: a review," *Dermatological Rev.*, vol. 6, no. 1, p. e70015, 2025.

[54] A. Alotaibi and D. AlSaeed, "Skin cancer detection using transfer learning and deep attention mechanisms," *Diagnostics*, vol. 15, no. 1, p. 99, 2025.

[55] A. Aboulmira *et al.*, "Hybrid model with wavelet decomposition and EfficientNet for accurate skin cancer classification," *J. Cancer*, vol. 16, no. 2, p. 506, 2025.

[56] M. A. Khan *et al.*, "Automatic melanoma and non-melanoma skin cancer diagnosis using advanced adaptive fine-tuned convolution neural networks," *Discov. Oncol.*, vol. 16, no. 1, p. 645, 2025.

[57] F. Güler and M. Ağraz, "Investigation of Binary and Multiclass Classification Performance of Skin Cancer Images Using Transfer Learning Methods," *J. Clin. Pract. Res.*, vol. 47, no. 3, p. 235, 2025.