BMC Cancer

SYSTEMATIC REVIEW

Open Access

Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: a systematic literature review

Hoda Naseri[1] and Ali A. Safaei[1,2*]

Abstract

Background Melanoma is a highly aggressive skin cancer, where early and accurate diagnosis is crucial to improve patient outcomes. Dermoscopy, a non-invasive imaging technique, aids in melanoma detection but can be limited by subjective interpretation. Recently, machine learning and deep learning techniques have shown promise in enhancing diagnostic precision by automating the analysis of dermoscopy images.

Methods This systematic review examines recent advancements in machine learning (ML) and deep learning (DL) applications for melanoma diagnosis and prognosis using dermoscopy images. We conducted a thorough search across multiple databases, ultimately reviewing 34 studies published between 2016 and 2024. The review covers a range of model architectures, including DenseNet and ResNet, and discusses datasets, methodologies, and evaluation metrics used to validate model performance.

Results Our results highlight that certain deep learning architectures, such as DenseNet and DCNN demonstrated outstanding performance, achieving over 95% accuracy on the HAM10000, ISIC and other datasets for melanoma detection from dermoscopy images. The review provides insights into the strengths, limitations, and future research directions of machine learning and deep learning methods in melanoma diagnosis and prognosis. It emphasizes the challenges related to data diversity, model interpretability, and computational resource requirements.

Conclusion This review underscores the potential of machine learning and deep learning methods to transform melanoma diagnosis through improved diagnostic accuracy and efficiency. Future research should focus on creating accessible, large datasets and enhancing model interpretability to increase clinical applicability. By addressing these areas, machine learning and deep learning models could play a central role in advancing melanoma diagnosis and patient care.

Keywords Melanoma, Dermoscopy, Machine learning, Deep learning, Convolutional neural networks (CNN), Diagnosis, Prognosis

*Correspondence:
Ali A. Safaei
aa.safaei@modares.ac.ir
[1] Department of Data Science, Faculty of Interdisciplinary Science and Technology, Tarbiat Modares University, Tehran, Iran
[2] Department of Medical Informatics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

Introduction

Today, skin cancer is an important issue that affects many people in the world. Skin cancer appears when the normal growth of skin cells is affected, causing a mutation in the DNA and eventually leading to skin cancer. Exposure to ultraviolet (UV) rays is considered to be the main

cause of skin cancer [55]. Melanoma, one of the most aggressive forms of skin cancer, has become a significant public health concern due to its increasing incidence and potential for rapid progression [3]. Although melanoma accounts for only 4% of all skin cancers, it is responsible for 75% of all skin cancer-related deaths, according to statistics from the American Cancer Society [1]. Survival rate results show that people typically live between 5 and 10 years after a melanoma diagnosis [17]. The occurrence of melanoma depends on various factors including family history, place of residence, skin color and other factors [17]. Sunlight also plays an important role in the development of melanoma. In fact, one of the most common causes is the over exposure of skin to ultraviolet radiations coming from the sun [2]. The dangers of UV rays are greater in fair-skinned people who are sensitive to the sun than in people with dark skin who are less sensitive to the sun [42]. This disease is more common in geographical areas such as Australia, where there is a lot of sunlight, and most people have fair skin. Studies have shown that most gene mutations are like those caused by UV light. (Supplementary File 1)

A common way to diagnose melanoma is to use the ABCDE (asymmetry, borders, color, diameter, and evolution) rule. These are warning signs to check for melanoma. High levels of asymmetry or border irregularities are the first warning sign, as well as strange color of the mole and diameter greater than 6 mm. All these signs are examined to analyze their evolution over time [43]. The more changes, the more likely the mole is malignant [1]. Early detection is critical for improving patient outcomes and increasing survival rates, as treatments are more effective in the early stages of the disease [5]. Melanoma starts with the formation of cells in the melanin pigment that gives the skin its color. It comes from skin cells called melanocytes [4]. It has the ability to travel to the lower layers of our skin, enter the circulation, and then spread to other regions of our bodies [5]. In the event that the indications are analyzed appropriately, this ailment can be treated in its beginning period and can be relieved; yet on the off chance that it is analyzed past the point of no return, it can become further into the skin, spread to different pieces of the body and can be dangerous, as it gets hard to treat [6]. Therefore, early detection and accurate prognosis are pivotal in improving patient outcomes and survival rates. In addition, a significant reduction in costs related to the diagnosis of critical conditions may help reduce the financial burden [57]. Traditional diagnostic methods, such as dermoscopy and pathology, play an essential role in melanoma detection and prognosis, but they rely heavily on the expertise and judgment of clinicians, which can lead to variability in diagnosis and errors due to the subjective nature of image interpretation [61].

Dermoscopy is a non-invasive imaging technique that improves the visualization of skin lesions, enhancing the ability to distinguish between benign and malignant conditions. Before the development of dermoscopy pictures, most skilled dermatologists had a rate of success of only 60% in diagnosing skin cancer, but dermoscopy images raised success rates to between 75% and 84% [7]. However, despite its benefits, interpreting dermoscopy images remains challenging due to the complexity of skin lesion patterns, and manual analysis is often time-consuming and prone to human error [44]. While pathology provides a more definitive diagnosis through biopsy and histological examination, it is invasive, costly, and time-consuming. As a result, there is an increasing demand for automated diagnostic systems that can assist clinicians by providing more objective, efficient, and accurate interpretations of dermoscopy images [45]. Recent advances in image processing methods and artificial intelligence can help melanoma diagnosis and prognosis, and early detection can increase survival rates. Moreover, computer-aided diagnostic (CAD) tools save time and effort compared with existing clinical approaches [9].

In recent years, the rapid advancements in machine learning and deep learning have provided new opportunities to enhance the diagnosis and prognosis of melanoma [45]. These technologies have demonstrated the ability to analyze large datasets of dermoscopy images, extracting critical features such as texture, shape, and color to classify lesions with high accuracy [46]. Machine learning is a remarkable field in health care, its contribution to various fields of diagnosis and detection holds high importance. Methods like decision trees and artificial neural networks in cancer prediction have been widely used for years; however, melanoma prediction is a relatively younger field of study [10]. Deep learning has been applied to resolve very complex classification and segmentation tasks without the use of any image preprocessing method. The architecture of deep networks is generally based on convolution layers that extract the important features of images to learn different lesions [1]. Convolutional neural networks (CNNs), are capable of learning from complex data, identifying patterns that may be difficult for human observers to detect, and improving diagnostic accuracy beyond traditional methods. These techniques have the potential to not only assist in early diagnosis but also provide insights into the prognosis of melanoma, aiding in clinical decision-making and personalized treatment plans [11]. Indeed, various federated learning (FL) and transfer learning (TL) techniques have been widely used to help dermatologists classify melanoma lesions [55].

Figure 1 illustrates the steps and techniques for melanoma detection. In the first step, dermoscopy images of

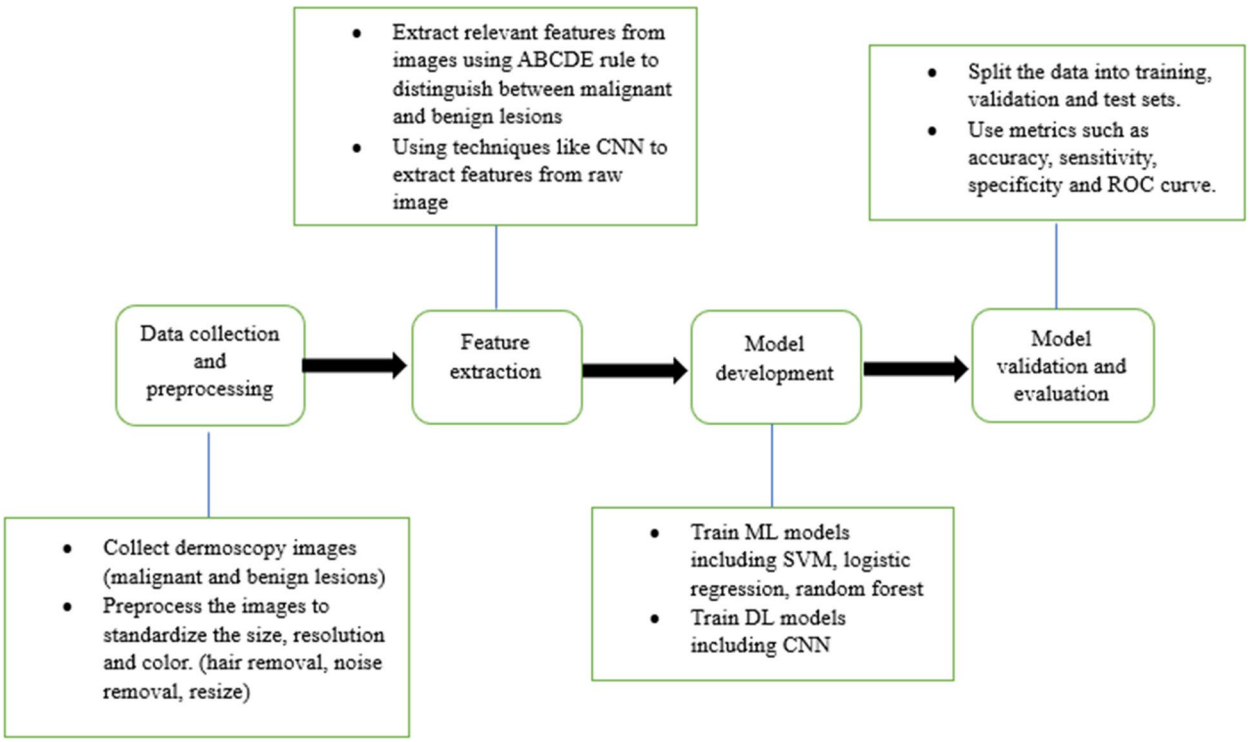

Fig. 1 Melanoma diagnosis process, from collecting dermoscopy images to choosing the optimal model

skin lesions, including melanoma and benign lesions, are collected and preprocessed to standardize their size, resolution, and color balance. Relevant features such as texture, shape and color are then automatically extracted from the images using techniques like convolutional neural networks (CNNs). Subsequently, machine learning (ML) and deep learning (DL) models are developed using the extracted features or raw image data. Machine learning (ML) models like support vector machines (SVM) and DL models like CNNs are commonly used for this purpose due to their ability to learn from complex data. The developed models are then validated and evaluated using metrics such as accuracy, sensitivity, and specificity on separate training, validation, and test datasets.

The objective of this systematic review is to provide a comprehensive analysis of the recent advancements in machine learning and deep learning methodologies applied to the diagnosis and prognosis of melanoma using dermoscopy images. By evaluating a wide range of studies, our review synthesizes findings from various model architectures, including CNNs, transfer learning, and ensemble methods, offering a clearer picture of their performance in terms of accuracy and reliability, and identify the challenges and limitations encountered in this field. While previous studies have primarily focused on the success of machine learning and deep learning models, this review systematically identifies the challenges, such as the lack of large, well-labeled datasets, model interpretability, and computational limitations. Moreover, we offer recommendations for addressing these challenges, including the use of transfer learning and improved data-sharing practices. By conducting this comprehensive systematic review, we aim to offer a clear understanding of how machine learning and deep learning are being leveraged to advance melanoma diagnosis and prognosis. Furthermore, we provide actionable insights and recommendations for future research, addressing the current limitations of AI-based diagnostic systems and offering pathways for improving their efficacy in clinical practice.

Objectives

The primary objectives of this article are as follows:

- To evaluate and summarize machine learning and deep learning techniques for melanoma diagnosis.
- To analysis the role of machine learning and deep learning in the diagnostic analysis and prognosis of melanoma.
- To identify challenges and limitations in recent research.
- To explore the state-of-the-art research trends, opportunities, and challenges for other researchers in diagnosing melanoma.

Methods

This study adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA) [49, 50]. In this section, as the routine of the systematic review research, the decisions for this study are described in detail first:

Review questions

In the framework of this article, a series of questions are presented to deepen the topic. Comprehensive answers to these questions are provided in the following sections. We reviewed the articles from 2016 to 2024, this research aimed to answer the following questions:

1) *What machine learning and deep learning techniques were employed in the paper for the diagnosis and prognosis of melanoma from dermoscopy images?*
2) *How were the dermoscopy images preprocessed and feature-engineered to extract relevant information for melanoma detection and prognosis?*
3) *What datasets were used in the study, and how were they curated or collected to ensure diversity and representativeness?*
4) *What kind of study is it? (Classification, Diagnosis, Analysis, Prediction, Identification)*
5) *What performance metrics were used to evaluate the effectiveness of the proposed methods in melanoma diagnosis and prognosis?*
6) *Is the prognosis of melanoma worth it?*

Information sources

This research is a systematic review study searching for published sources in English in PubMed, Science Direct, Springer, frontiers, IEEE, MDPI since 2016 to 2024. We searched the articles for various combinations of relevant keywords such as melanoma, dermoscopy, machine learning, deep learning and obtained 12,657 articles.

Search strategy

The criterion for search is the following phrase:

(Melanoma) AND (Deep learning OR Machine learning) AND (Dermoscopy images) AND (Diagnosis OR Prognosis) NOT review.

In the first step, according to our search criteria, 12,657 articles were identified. After removing duplicate records, the number of articles reached 11,528 records. After applying inclusion criteria, 170 full-text articles were identified, which were further inspected by applying the exclusion criteria. Finally, 34 articles were selected and analyzed and their results were discussed in this research. The above conditions were applied and the results of Table 1 were obtained.

Table 1 The number of founded studies and selected ones in six databases

Database	Founded studies	Selected studies
PubMed	96	1
Science direct	16	5
Springer	35	10
Frontiers	91	1
IEEE	56	5
MDPI	32	12

Eligibility criteria and selection process

As the first essential points for selecting the most relevant papers, we reviewed the chosen papers considering the keywords, the year of the publication, the subject, and their journal. The selection index for the finalized studies was based on:

- Published between 2016 and 2024.
- Related to melanoma, dermoscopy, machine learning, deep learning.

Exclusion criteria

The deletion index for removing articles is as follows:

- Not English.
- Not indexed in ISI.
- Not deep learning or machine learning.
- Not Melanoma.
- Not dermoscopy images.

Search results

The PRISMA flow diagram in Fig. 2 illustrates the selection procedure [50].

Figure 3 shows the rate of each of the 6 databases separately among the 34 selected articles. MDPI publishes 35% of the articles, Springer publishes 29%, and Science Direct (Elsevier) published 15%, and these three account for the largest share (Supplementary File 2).

The highest frequency of studies is related to the years 2020 and 2022, which indicates more attention to this issue over time. The number of selected studies published in this review between 2018 and 2024 is shown in Fig. 4.

Table 2 shows the databases in which the selected articles have been published, the name of the article, the year of publication, and the authors (Supplementary File 3).

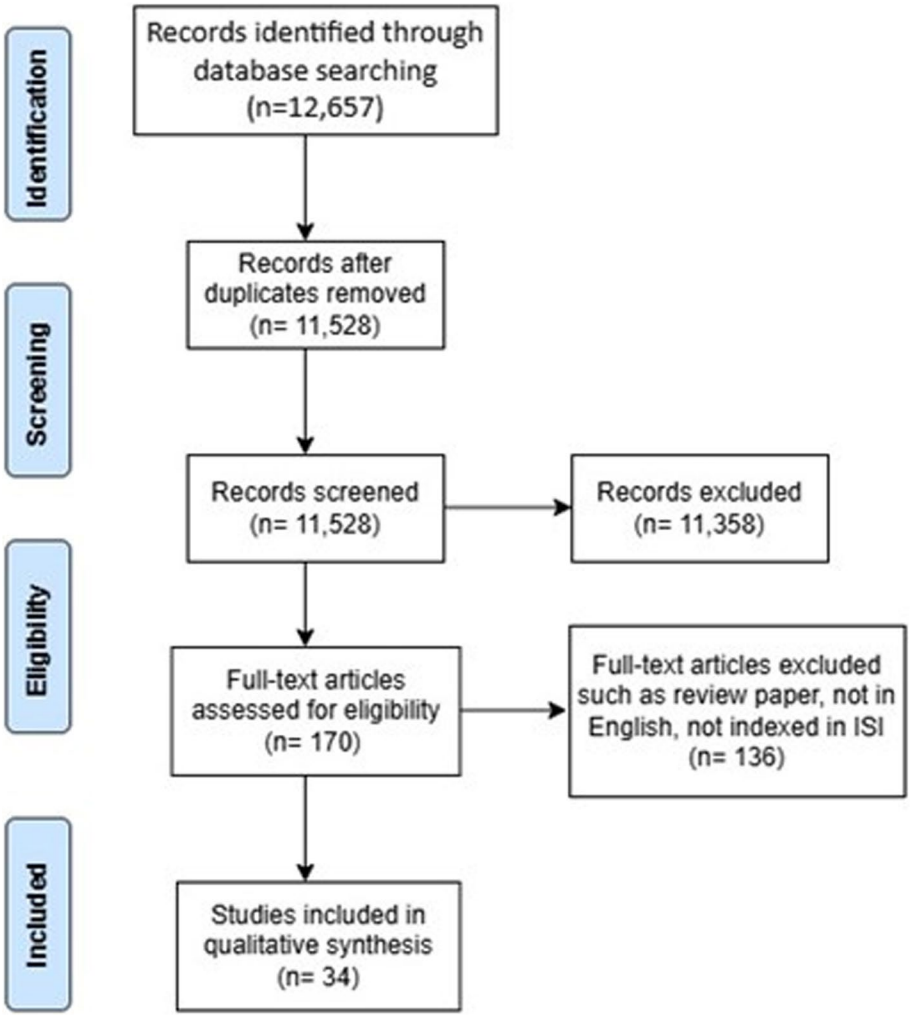

Fig. 2 Study selection using PRISMA diagram

Datasets

ISIC 2016, 2017, 2018, 2019, 2020

ISIC dataset, also known as the International Skin Imaging Collaboration dataset, is a collection of skin images used for research in the field of dermatology, particularly in the development of automated skin cancer detection systems. The dataset contains high-resolution images of skin lesions, along with corresponding metadata such as lesion type (e.g., melanoma, nevus), patient age, and lesion location. ISIC dataset consists of images acquired using various imaging modalities, including dermoscopy, clinical photography, and confocal microscopy. It serves as a benchmark dataset for evaluating the performance of machine learning and deep learning algorithms in the diagnosis and classification of skin lesions [24].

ISBI 2016, 2017

ISBI (International Symposium on Biomedical Imaging) dataset cover various modalities and medical conditions, and they are often used for tasks such as image segmentation, classification, and registration. ISBI dataset includes images from various medical imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, and microscopy. The resolution and dimensionality of the images vary depending on the specific dataset and modality. Some datasets consist of 2D images, while others contain 3D volumes or time-series data.

PH2

PH2 dataset is a collection of dermoscopy images of melanocytic lesions. The dataset contains a total of 200 dermoscopy images of melanocytic lesions, captured using

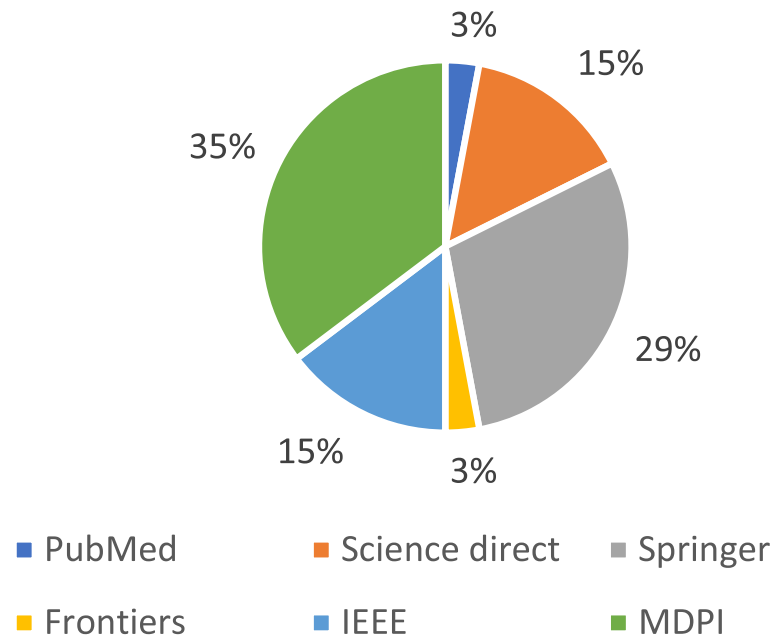

Fig. 3 The percentage of selected articles retrieved from various databases

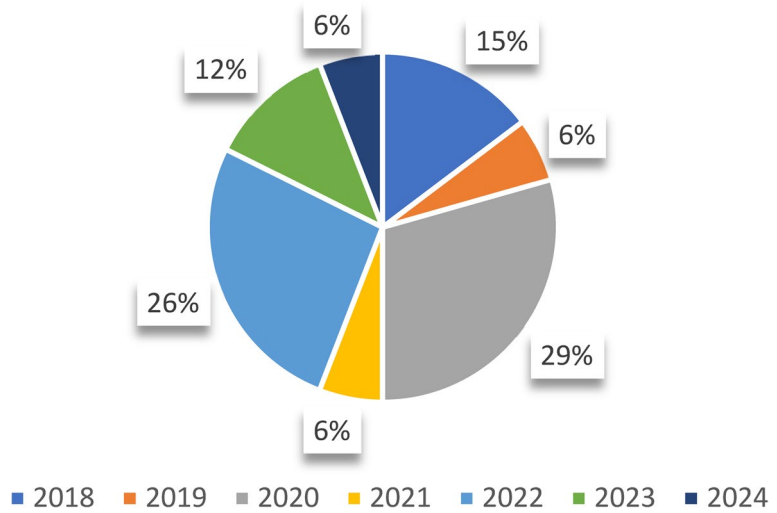

Fig. 4 the percentage of selected articles published from 2018 to 2024 in the subject of the study area

a DermLite II Pro dermoscope with a magnification of 10×10. The lesions in PH2 dataset are divided into two categories (benign and malignant). The images in PH2 dataset have a resolution of 768×560 pixels. Along with the images, the dataset also provides corresponding annotations, including lesion outlines and boundary points. These annotations can be used for tasks such as segmentation and classification. PH2 dataset is a valuable resource for automated melanoma diagnosis and prognosis using machine learning and computer vision techniques.

HAM 10,000

HAM10000 dataset is a collection of images of skin lesions, specifically focusing on melanoma and other types of skin cancer. It is widely used in dermatology, medical imaging, and machine learning for tasks such as image classification, segmentation, and diagnosis. The dataset contains 10,015 dermoscopy images of skin

Table 2 Extracted the paper's title, year of publication, and authors from various databases

Database	Paper's Title	Year	Authors
Science direct	A comparative study of deep learning architectures on melanoma detection [11]	2019	Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani
	Machine learning approach in melanoma cancer stage detection [12]	2020	Rashmi Patil, Sreepathi Bellary
	Skin lesion classification with ensembles of deep convolutional neural networks [13]	2018	Balazs Harangi
	Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks [8]	2018	Mohammed A. Al-masni, Mugahed A. Al-antari, Mun-Taek Choi, Seung-Moo Han, Tae-Seong Kim
	Explainable deep inherent learning for multi-classes skin lesion classification [69]	2024	Khalid M. Hosny, Wael Said, Mahmoud Elmezain, Mohamed A. Kassem
springer	A Convolutional Neural Network Framework for Accurate Skin Cancer Detection [1]	2020	Karl Thurnhofer-Hemsi, Enrique Domínguez
	Classification of skin cancer from dermoscopy images using deep neural network architectures [14]	2022	Jaisakthi S M, Mirunalini P, Chandrabose Aravindan, Rajagopal Appavu
	DenseNet-II: an improved deep convolutional neural network for melanoma cancer detection [10]	2022	Nancy Girdhar, Aparna Sinha, Shivang Gupta
	Melanoma diagnosis using deep learning techniques on dermatoscopic images [15]	2021	Mario Fernando Jojoa Acosta, Liesle Yail Caballero Tovar, Maria Begonya GarciaZapirain and Winston Spencer Percybrooks
	Recent Advances in Melanoma Diagnosis and Prognosis Using Machine Learning Methods [16]	2023	Sarah Grossarth, Dominique Mosley, Christopher Madden, Jacqueline Ike, Isabelle Smith, Yuankai Huo7, Lee Wheless
	Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier [60]	2023	Yousef S. Alsahaf, Mohamed A. Kassem and Khalid M. Hosny
	Refined Residual Deep Convolutional Network for Skin Lesion Classification [65]	2022	Khalid M. Hosny, Mohamed A. Kassem
	The skin cancer classification using deep convolutional neural network [66]	2018	Ulzii-Orshikh Dorj, Keun-Kwang Lee, Jae-Young Choi, Malrey Lee
	Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet [67]	2020	Khalid M. Hosny, Mohamed A. Kassem, Mohamed M. Fouad
	Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks [68]	2020	Khalid M. Hosny, Mohamed A. Kassem, Mohamed M. Fouad
MDPI	Automatic Malignant and Benign Skin Cancer Classification Using a Hybrid Deep Learning Approach [5]	2022	Atheer Bassel, Amjed Basil Abdulkareem, Zaid Abdi Alkareem Alyasseri, Nor Samsiah Sani, and Husam Jasim Mohammed
	Deep Learning–Based Methods for Automatic Diagnosis of Skin Lesions [18]	2020	Hassan El-Khatib, Dan Popescu and Loretta Ichim
	Deep Learning-Based Transfer Learning for Classification of Skin Cancer [19]	2021	Satin Jain, Udit Singhania, Balakrushna Tripathy, Emad Abouel Nasr, Mohamed K. Aboudaif and Ali K. Kamrani
	DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopy Images [20]	2023	Maryam Tahir, Ahmad Naeem, Hassaan Malik, Jawad Tanveer, Rizwan Ali Naqvi, and Seung-Won Lee
	Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopy Images [21]	2022	Solene Bechelli, Jerome Delhommelle
	Melanoma and Nevus Skin Lesion Classification Using Handcraft and Deep Learning Feature Fusion via Mutual Information Measures [22]	2020	Jose-Agustin Almaraz-Damian, Volodymyr Ponomaryov, Sergiy Sadovnychiy, Heydy Castillejos-Fernandez
	Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopy Images [9]	2022	Ranpreet Kaur, Hamid GholamHosseini, Roopak Sinha, Maria Lindén
	On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning [23]	2022	Mohammad Fraiwan, Esraa Faouri

Table 2 (continued)

Database	Paper's Title	Year	Authors
	Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network [24]	2018	Yuexiang Li, Linlin Shen
	The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning [25]	2020	Shunichi Jinnai, Naoya Yamazaki, Yuichiro Hirano, Yohei Sugawara, Yuichiro Ohe, Ryuji Hamamoto
	SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Image [59]	2022	Ahmad Naeem, Tayyaba Anees, Makhmoor Fiza, Rizwan Ali Naqvi and Seung-Won Lee
	SNC_Net: Skin Cancer Detection by Integrating Handcrafted and Deep Learning-Based Features Using Dermoscopy Images [61]	2024	Ahmad Naeem, Tayyaba Anees, Mudassir Khalil, Kiran Zahra, Rizwan Ali Naqvi and Seung-Won Lee
IEEE	Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network [26]	2020	LISHENG WEI, KUN DING, HUOSHENG HU
	Convolutional Neural Network Based Skin Lesion Analysis for Classifying Melanoma [6]	2019	Shetu Rani Guha, Dr. S. M. Rafizul Haque
	Deep Learning-Based System for Automatic Melanoma Detection [27]	2020	ADEKANMI A. ADEGUN AND SERESTINA VIRIRI
	Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer learning [62]	2020	Mohamed A. Kassem, Khalid M. Hosny and Mohamed M. Fouad
	Skin Cancer Classification using Deep Learning and Transfer Learning [63]	2018	Khalid M. Hosny, Mohamed A. Kassem, and Mohamed M. Foaud
Frontiers	Untangling Classification Methods for Melanoma Skin Cancer [28]	2022	Seeja R D, Suresh A
PubMed	Deep Learning Based Skin Lesion Segmentation and Classification of Melanoma Using Support Vector Machine (SVM) [29]	2023	Ayushi Kumar, Avimanyou Vatsa

lesions, captured from different patients. These images are taken under controlled conditions using a dermatoscope, which is a handheld device used by dermatologists to examine skin lesions in detail. Each image in HAM10000 dataset is labeled with information about the diagnosis, including the presence or absence of melanoma, as well as other types of skin lesions such as nevi, seborrheic keratosis, and basal cell carcinoma. The images in the dataset are of varying resolutions and quality, reflecting the diversity of real-world clinical images. Some images may exhibit artifacts, blurriness, or uneven illumination, which adds to the complexity of the dataset and the challenges involved in analyzing it. The dataset is typically divided into training, validation, and test sets, with a predefined distribution of images across different classes. This allows researchers to train and evaluate machine learning models on a consistent basis and assess their performance on unseen data. HAM10000 dataset is used for various tasks, including image classification to differentiate between benign and malignant skin lesions, segmentation to outline the boundaries of lesions, and detection to identify suspicious regions within images. The dataset serves as a benchmark for evaluating the effectiveness of different algorithms and techniques in automated skin cancer diagnosis.

MedNode

MedNode dataset is a collection of high-quality dermoscopic images focused on skin lesion classification. This dataset consists of 70 melanoma and 100 nevus images from the digital image archive of the Department of Dermatology of the University Medical Center Groningen (UMCG) used for the development and testing of the MED-NODE system for skin cancer detection from macroscopic images. The dataset's primary objective is to enhance AI-driven diagnostic tools by providing well-annotated images that support supervised learning [51, 58].

DermIS

DermIS (Dermatology Information System) dataset is a comprehensive resource containing a large variety of clinical images related to skin conditions. This dataset contains 500 benign and 500 malignant images with

dimensions of 600×450 pixels. Covering numerous dermatological diseases, including both benign and malignant skin conditions, DermIS is widely used for training AI models to diagnose and classify skin diseases. This dataset al.so serves as a valuable tool for medical education, providing detailed clinical descriptions and visual references that help in the understanding of skin diseases [52].

DermQuest

DermaQuest dataset consists of clinical images and associated metadata related to a wide range of skin conditions. It is particularly useful for machine learning applications aimed at the automatic classification of skin diseases. DermaQuest includes labeled images that span conditions such as skin cancer, eczema, and psoriasis, making it a versatile dataset for both AI development and medical research. Its primary purpose is to assist in the creation of diagnostic algorithms that can reliably identify various dermatological conditions based on clinical images, thus supporting advancements in automated dermatology diagnostics [53, 58].

DermPK

DermPK dataset contains 157 dermoscopic images. These images were gathered from the Multan Institute of Nuclear Medicine and Radiotherapy (MINAR) Pakistan. This dataset includes images from the following classes of skin cancer: Melanoma, Basal cell carcinoma, squamous cell carcinoma, benign keratosis, melanotic navi, and unknown [54].

ISIC (International Skin Imaging Collaboration) and HAM10000 are the most widely used datasets that publishes a large number of dermoscopy lesion images for skin cancer diagnosis every year (Supplementary File 4, 5) (Table 3).

Results

After refining the extracted studies and adding studies manually, the studies are examined separately in this section. In this review of studies, an attempt has been made to answer the questions raised in the previous section.

Study selection

The selected articles use different machine learning or deep learning algorithms which are listed below.

Machine learning

Machine learning (ML) plays a crucial role in the diagnosis and prognosis of melanoma, contributing to more accurate and timely detection of skin cancer. ML algorithms are used to process dermoscopy images and extract relevant features such as asymmetry, border irregularity, color variation, and texture patterns [12]. Various image processing techniques, including edge detection and segmentation are employed to enhance image quality and highlight important diagnostic features. ML models, including supervised learning algorithms such as Support Vector Machines (SVM), Random Forests, Logistic Regression (LR), Linear Discriminant Analysis (LDA), k-nearest Neighbors classifier (KNN), Decision Tree Classifier (CART), and Gaussian Naive Bayes are trained on annotated dermoscopy image datasets to classify lesions as benign or malignant [21]. These models learn to distinguish between different types of skin lesions based on extracted features and patterns, enabling automated diagnosis and decision support for dermatologists. Logistic regression (LR) is particularly suitable for classifying binary problems and is often used in biomedical applications [30]. LR is a parametric method that calculates the probability of class membership for one of two classes in a data set. LDA is commonly used for supervised pattern classification and produces linear decision boundaries between classes [31]. KNN relies on proximity. This method uses a metric that assigns to each item a specific category depending on its closeness to similar data points, and its versatility makes it an appropriate choice to study classification problems [30]. ML algorithms are used to predict prognosis and risk of melanoma recurrence or metastasis based on dermoscopy images and clinical data. By analyzing various clinical and histopathological factors, ML models can assess the probability of disease progression and inform treatment decisions.

Deep learning

Deep convolutional neural networks (DCNN) It is a type of artificial neural network, specifically designed for processing and analyzing visual data, such as image and speech recognition [32], object detection [33] and natural language processing [34]. DCNNs are composed of multiple layers of neurons, including convolutional layers, pooling layers, and fully connected layers. The key feature of DCNNs is their ability to automatically learn hierarchical representations of features from raw pixel data. This hierarchical representation allows DCNNs to capture both low-level features (edges and textures) and high-level features (object shapes and patterns) in an image. Convolutional layers in DCNNs apply convolutional filters to input images, extracting local patterns and features. Pooling layers then down sample the feature maps obtained from convolutional layers, reducing the spatial dimensions of the data while preserving important

features. Fully connected layers integrate the extracted features and perform classification or regression tasks. DCNNs have demonstrated state-of-the-art performance in various computer vision tasks, including image classification, object detection, and image segmentation.

Xception architecture Xception is a deep convolutional neural network (CNN) architecture introduced by François Chollet [35]. The name Xception stands for Extreme Inception, indicating its derivation from the Inception architecture. Xception is built upon the idea of depthwise separable convolutions, which aim to reduce computational complexity while maintaining high performance. In traditional CNNs, convolutional layers apply a kernel to the entire input volume, resulting in a large number of parameters and computational costs. Depthwise separable convolutions break down the standard convolutional operation into two separate steps: depthwise convolutions and pointwise convolutions. Xception achieved state-of-the-art performance on various image classification benchmarks while significantly reducing computational complexity compared to previous architectures.

AlexNet architecture AlexNet [36] is the first breakthrough in the architecture of CNNs applied to large datasets. The basic architecture of AlexNet consists of an input layer, five convolutional layers, two normalization layers, three fully-connected layers, and a final layer with softmax activation function in the output to ensure each neuron predicts class probabilities of a particular image.

VGGNet architecture VGGNet or Visual Geometry Group Network is a deep convolutional neural network proposed by Karen Simonyan and Andrew Zisserman [37]. It achieves one of the top performances in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014. VGGNet utilizes smaller filters of 3×3, compared to AlexNet 11×11 filter, in order to provide better features extraction from images. The study also verifies that using much smaller filters in order to increase the depth of network instead of its width plays a critical role for gaining higher performance. There are two versions of this architecture: VGG16 and VGG19 with different depths and layers. VGG19 is deeper than VGG16. The increased depth allows the network to learn more complex features from input images.

ResNet architecture Deep residual neural network (ResNet) architecture is proposed by He et al. (2016) and won ILSVRC & COCO 2015 competitions [38]. Researchers are able to train deeper and more effective neural networks using ResNet with better recognition accuracies. ResNets with various depths such as ResNet50 and ResNet100 use the bottleneck features to improve efficiency in compare with its predecessor CNN models. it enables to train very deep networks and can boost the performance of the model by overcoming the vanishing gradient problem.

GoogleNet GoogLeNet was presented at the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 classification and detection challenges, also known as Inception. The main difference with respect to AlexNet, one of the first classification networks, is the deepness of the network, with a total of 22 layers. They increased the complexity of the network using more neurons at each level as well. It is composed of modules named as Inception, including a set of convolution layers and a max pooling, using a ReLU activation. The input is an RGB image of size 224×224 and the output is a probability vector of 1000 classes.

DenseNet DenseNet is a deep network based on a modification of the connections between layers. Whereas in the traditional nets one layer is connected only to the following one, DenseNet connects one layer to all the subsequent layers, i.e. all the preceding feature maps are used as input for the next layers. The network architecture is based on multiple densely connected "dense blocks", including convolutional, max pooling and activation layers as a transition between one block and the following one, for a total of 201 layers. Generally, CNN tries to change the size of the feature map by down sampling layers. But DenseNet facilitates both down-sampling and feature concatenation by dividing the network into multiple densely connected dense blocks. The feature map size inside the blocks remains the same. Outside the

Table 3 Summary of skin lesions dataset, containing melanoma images and full images

Dataset	Melanoma	Total
ISIC 2016	173	900
ISIC 2017	374	2000
ISIC 2018	2594	12,970
ISIC 2019	4522	25,331
ISIC 2020	584	33,126
ISBI 2016	273	900
ISBI 2017	374	2000
PH2	40	200
HAM 10,000	1113	10,015
MedNode	70	170
DermIS	43	69
DermQuest	76	137
DermPK	30	157

dense blocks convolution and pooling operations are performed to down sample and inside the dense block the size of the feature maps are same which helps to carry out concatenation [14].

MobileNet MobileNet is a deep learning architecture specifically designed for efficient deployment on mobile and embedded devices with limited computational resources. Developed by Google, MobileNet utilizes depthwise separable convolutions to reduce the computational complexity of traditional convolutional layers while maintaining high accuracy in image classification tasks. By decomposing standard convolutions into depthwise convolutions and pointwise convolutions, MobileNet significantly reduces the number of parameters and computational cost, making it well-suited for real-time applications on resource-constrained devices. Additionally, MobileNet offers flexibility through its adjustable depth multiplier parameter, allowing developers to trade-off between computational efficiency and model accuracy according to specific deployment requirements. With its lightweight and efficient design, MobileNet has become a popular choice for various mobile vision applications, including object detection and image recognition [47].

U-net The network merges a convolutional network architecture with a deconvolutional architecture to output the semantic segmentation. The architecture of U-Net consists of a contracting path and an expansive path. The contracting path contains encoder layers that capture contextual information and reduce the spatial resolution of the input, while the expansive path contains decoder layers that decode the encoded data and use the information from the contracting path via skip connections to generate a segmentation map. It consists of repeated applications of two 3×3 convolutions (Unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2×2 maxpooling operation with stride2 for down sampling. At each downsampling step the number of feature channels are doubled [29, 39] (Supplementary File 6).

EfficientNet EfficientNet is proposed to improve accuracy and efficiency of CNN by using a uniform scaling method for all dimensions, i.e. depth, width and resolution of the network, while scaling down the model [14]. In the proposed compound scaling method, it uses a grid search to find the relationship between different scaling dimensions of the base grid under a fixed resource constraint. With this method, a scaling factor is determined for the dimensions of depth, width and resolution. These coefficients are then applied to scale the baseline network to the desired target network [40].

DSCC_Net (deep learning-based skin cancer classification network) This model is designed to identify four different types of skin cancer (MEL, BCC, MN, and SCC). The proposed model has the capability of extracting dominant features from dermoscopy images that can assist in the accurate identification of the disease. This system was trained and tested on images of four main categories of skin cancer. The input image's size is fixed to a resolution of 150×150 pixels. In addition, the dataset was used according to the data normalization technique, in order to stop the model from being overfit [20].

FCRN (Fully Convolutional Residual Networks) In this method, to make full use of very deep networks, a set of schemes were proposed to ensure effective training and learning under limited training data. Residual learning was applied to cope with the degradation and overfitting problems when a network goes deeper. Then, a fully convolutional residual network (FCRN) is developed for accurate segmentation of skin lesion, and its capability is enhanced by incorporating a multi-scale background information integration scheme. Finally, it seamlessly integrates the proposed FCRN for segmentation and other remaining very deep networks for classification to form a two-stage framework. This framework enables the classification network to extract more representative and specific features based on segmented results instead of the whole dermoscopy images, further alleviating the insufficiency of training data [41].

FrCN (full resolution convolutional networks) FrCN segmentation method does not require any pre-processing techniques, such as artifact removal or low contrast enhancement, since it directly learns the full resolution features of each individual pixel of the input data. this method learns the full resolution features of each pixel of the input data to achieve more accurate pixel-wise segmentation of the skin lesions. This is achieved by eliminating the subsampling layers in the networks and enabling the convolutional layers to extract and learn the full spatial features of the input image. Due to this, the proposed FrCN produces finely segmented contours of the skin lesions [8].

FRCNN (faster, region-based convolutional neural network) This is a state-of-the-art deep learning model used for object detection, particularly in the context of identifying and localizing skin lesions from dermoscopy

images. Faster R-CNN integrates region proposal networks (RPNs) with convolutional neural networks (CNNs) to streamline the object detection process. By generating region proposals directly from the convolutional feature maps, Faster R-CNN eliminates the need for separate algorithms for region proposal generation, resulting in faster and more accurate detection of skin lesions. The utilization of Faster R-CNN in the development of the skin cancer classification system enhances the system's ability to precisely localize and classify pigmented skin lesions, thereby improving diagnostic accuracy and patient outcomes in dermatological practice [25].

SMTP (Similarity Measure for Text Processing) SMTP (Chim and Deng, 2010) is distinguishing among appearance and non-appearance of an element that is viewed as more significant than distinction among qualities related to the current element. The lesion segmentation technique is a basic method of pattern recognition algorithms to distinguish melanoma skin malignant growth in patients at soonest stage, in any case, in further stages it gets one of the deadliest illnesses and its death rate is extremely high. Hence, a precise melanoma stage detection scheme is presented based on use of SMTP loss function to identify and diagnosis of dermoscopy images which classify melanoma based on its stage using CNN with SMTP [12].

As shown in Fig. 5, preprocessing is performed on the input images. In the next step, the important features are extracted using the ABCDE rule. Also, techniques like CNN are applied to raw data. After running the appropriate algorithms, it is determined whether the input image of melanoma is malignant or benign.

Results of individual studies and syntheses

In the diagnosis of melanoma, several measures are commonly used to evaluate the performance of diagnostic models. Some of the key evaluation metrics include: Specificity (the measure of non-melanoma cases that correctly classified), Sensitivity or Recall (the proportion of melanoma cases that are correctly identified by the diagnostic model), Accuracy (the number of correctly classified skin lesions divided by the total number of skin lesions), Precision (the percentage of correctly classified labels are positive), F-measure and AUC-ROC.

These evaluation metrics help assess the diagnostic accuracy, reliability, and performance of models in identifying melanoma cases. Depending on the specific requirements of the diagnostic task and the importance of false positives versus false negatives, different metrics may be prioritized.

According to the above table, in various articles, tasks such as classification, diagnosis, detection and segmentation have been done on ISIC, ISBI, HAM10000, PH2 datasets. Algorithms such as SVM, logistic regression, gaussian naive bayes, KNN, decision tree, XG-Boost were used in machine learning tasks. DCNN, AlexNet, ResNet, U-Net, VGGNet, Xception, Inception and others were also used in deep learning tasks. ResNet and VGG-Net were used more than others 17 and 14 times respectively. However, DenseNet (DenseNet201, DenseNet-II), XG-Boost, SNC_Net, DCNN, GoogleNet + Transfer learning, AlexNet + Transfer learning and SMTP (Similarity Measure for Text Processing) have higher accuracy (more than 0.96).

According to Fig. 6, DCNN and DenseNet are the networks that were used more than others and had acceptable performance (Supplementary File 7).

Figure 7 provides information about the relationship between the number of images and the corresponding accuracy. Most of the researches that have been done and had accuracy above 90%, the images in their dataset were less than 1000. Notably, no dataset had more than 5000 images. Despite the existence of advanced methods such as data augmentation, the lack of images can be considered a serious problem, especially in the field of decision-making about the diagnosis and prognosis of melanoma. It is clear that a number of algorithms have very high accuracy on small datasets, leading to overfitting.

Table 4 shows the method performed along with accuracy, sensitivity, specificity, precision, F-measure, AUC-ROC. Some criteria have not been calculated in a number of articles.

According to Fig. 8, XG-Boost, DenseNet, lightweight deep learning network and SMTP (Similarity Measure for Text Processing) have accuracy higher than 0.96 and more than others. FCRN and SVM are less accurate than others with a value of approximately 0.85 (Fig. 8).

XG-Boost has very little sensitivity. That is, many dermoscopy images were not correctly diagnosed as melanoma. In addition to accuracy, DenseNet and SMTP also have higher sensitivity. which shows that these algorithms have an acceptable performance on the desired datasets. DenseNet (DenseNet201, DenseNet-II) and SMTP have higher accuracy and sensitivity, used HAM10000, ISIC 2019 datasets respectively, which have larger sample size. Therefore, it is recommended to select a larger data set to achieve a more accurate diagnosis (Fig. 9).

It is also worth mentioning that the articles that used the combination of several methods, such as

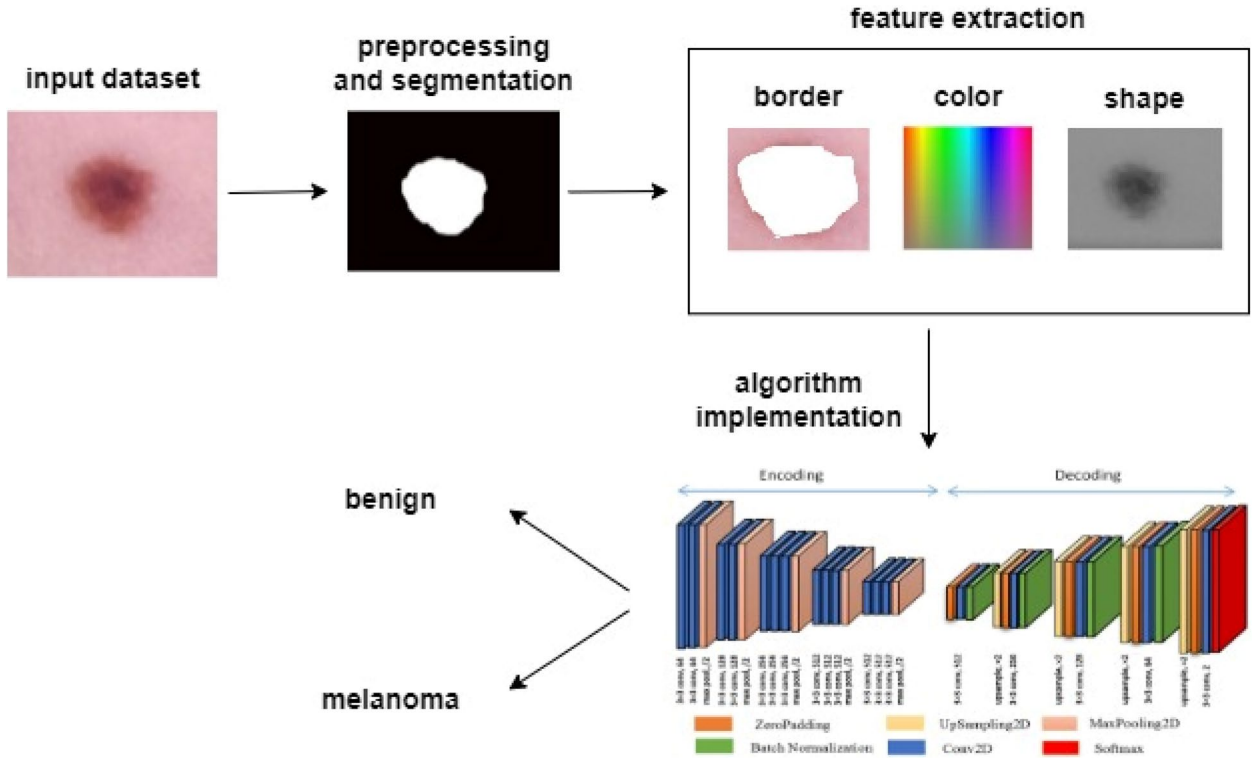

Fig. 5 System architecture

AlexNet + transfer learning and transfer learning + GoogleNet, had a higher accuracy and sensitivity in diagnosis.

Discussion and conclusion

The findings of this systematic review provide an overview of the current perspective of machine learning and deep learning approaches in improving the accuracy and efficiency of melanoma diagnosis using dermoscopy images, compared to previous research. For example, DenseNet and DCNN consistently outperforms traditional diagnostic methods by achieving over 95% accuracy on challenging datasets such as HAM10000, ISIC2019 and 2020. This review also emphasizes that models such as ResNet and EfficientNet improve

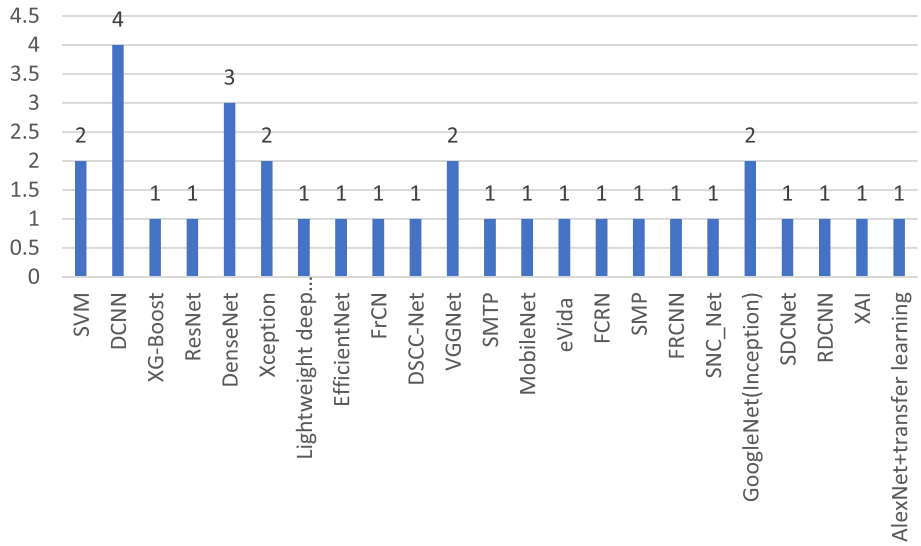

Fig. 6 The number of methods used separately in different articles

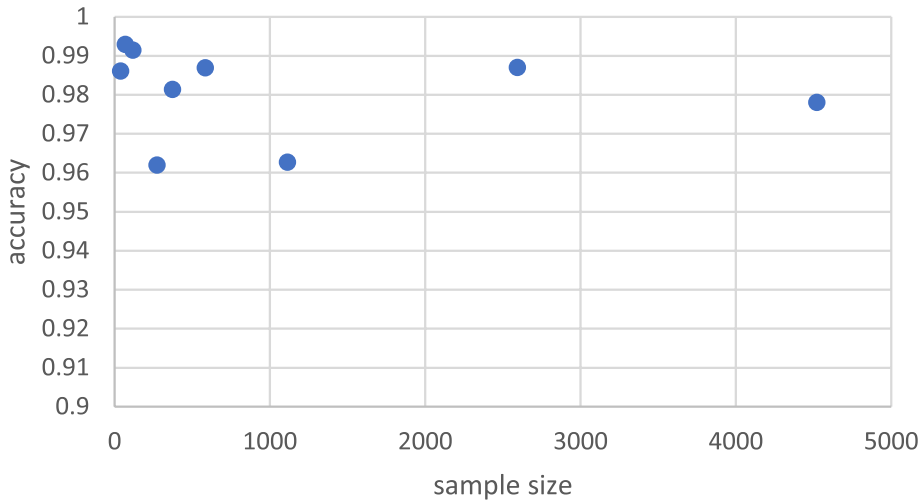

Fig. 7 The relationship between sample size and accuracy in different datasets

computational efficiency and facilitate real-time clinical use. This increased diagnostic accuracy and speed represents a significant improvement over previous machine learning and deep learning models with limited clinical application due to high computational requirements and lower accuracy. Overall, this review emphasizes the importance of integrating advanced machine learning and deep learning models into the diagnostic workflow to enhance early melanoma detection and patient outcomes. The results of this research will help future research and give users a better and more comprehensive view of conducting a suitable study and make previous studies more systematic. However, there are limitations such as the need for larger and more diverse data sets and improved model interpretability. Future work should focus on addressing these challenges to advance the clinical application of AI in melanoma diagnosis.

Limitations and future direction

A common limitation in many of the reviewed studies is the lack of access to large, well-labeled datasets, which are essential for training robust machine learning and deep learning models. Small sample sizes can lead to overfitting and reduce the generalizability of the models. The problem of training with small datasets can be reduced by data augmentation, image generation by an adversarial generative network, and transfer learning [48]. There should be a focus on developing common data sharing platforms and federated learning (FL) techniques, which allow models to be trained on distributed datasets without violating patient privacy. The creation of open-access annotated datasets for melanoma is also crucial [55].

Most studies focus on dermoscopy images alone, overlooking other critical data types such as genomic information, which could enhance diagnostic and prognostic accuracy. Genomic data, along with histopathological and clinical data, can provide a more holistic view of melanoma progression. Multi-modal approach can significantly improve predictive accuracy and enable personalized treatment recommendations. The existence of genomic datasets can identify gene activity using a model based on machine learning and help in more accurate diagnosis. It can also identify potential biomarkers of metastatic melanoma and therapeutic targets. Potential biomarkers that could serve as an alternative to painful biopsies and misleading imaging scans. However, more research is still needed to confirm these hypotheses [56].

While many studies demonstrate high accuracy, they often overlook model interpretability. Deep learning models, particularly CNNs, are often seen as black boxes, making it difficult for clinicians to properly compare results because different researchers use different datasets, analytical techniques, and computational resources [64]. To improve the clinical acceptance of machine learning and deep learning models, explainability in the models themselves should be prioritized. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME) or Grad-CAM (Gradient-weighted Class Activation Mapping) should be integrated to provide visual and interpretable feedback.

Several deep learning models require high computational resources for training and inference, which can limit their use in resource-constrained environments. To reduce the computational overhead while maintaining accuracy, lightweight model architectures, such as

Table 4 Summary of selected articles

Paper	Task	Dataset	Networks	Efficient Method	Accuracy	Sensitivity (Recall)	Specificity	Precision	F-measure	AUC-ROC
A comparative study of deep learning architectures on melanoma detection	Classification, detection, diagnosis	ISIC 2018	DCNN, AlexNet, VGGNet, ResNet, Xception	ResNet	0.9208	0.9253	-	0.9373	0.9274	-
A Convolutional Neural Network Framework for Accurate Skin Cancer Detection	Classification, detection	HAM10000	DenseNet201, transfer learning	DenseNet	0.9618	0.8104	-	0.9029	0.8542	-
Automatic Malignant and Benign Skin Cancer Classification Using a Hybrid Deep Learning Approach	classification	ISIC 2019	ResNet50, Xception, VGGNet16	Xception	0.909	0.886	-	-	0.896	0.917
Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network	Classification, detection, segmentation	ISBI 2016	U-Net, ResNet50, MobileNet, DesneNet	Lightweight deep learning network	0.962	0.934	0.974	-	-	-
Classification of skin cancer from dermoscopy images using deep neural network architectures	classification	ISIC 2019, 2020	DCNN, EfficientNet	EfficientNet	0.9174	-	-	-	-	0.9681
Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet	classification	ISIC 2018	AlexNet, Transfer learning	AlexNet + Transfer learning	0.987	0.956	0.9927	0.956	-	-

Table 4 (continued)

Paper	Task	Dataset	Networks	Efficient Method	Accuracy	Sensitivity (Recall)	Specificity	Precision	F-measure	AUC-ROC
Convolutional Neural Network Based Skin Lesion Analysis for Classifying Melanoma	classification	ISIC 2018	VGG-16, transfer learning	VGG-16	0.9107	0.79203	-	0.95039	-	-
Deep Learning Based Skin Lesion Segmentation and Classification of Melanoma Using Support Vector Machine (SVM)	Segmentation, classification	ISBI 2016	SVM, U-Net	SVM	0.8519	0.5	-	0.4259	0.46	-
Deep Learning–Based Methods for Automatic Diagnosis of Skin Lesions	Classification, diagnosis	PH2, ISIC 2019	GoogleNet, ResNet-101, NasNet-Large, transfer learning, SVM	Combining networks	PH2 = 0.95 ISIC 2019 = 0.93	PH2 = 92.5 ISIC 2019 = 92.5	PH2 = 96.66 ISIC 2019 = 0.9333	-	-	-
Deep Learning-Based System for Automatic Melanoma Detection	Classification, detection, segmentation	ISBI 2017, PH2	FrCN, U-Net, FCN, ResNet, DCNN	Combining networks	ISIC 2017, PH2 = 0.95	ISIC 2017 = 0.97 PH2 = 0.93	ISIC 2017 = 0.96 PH2 = 0.95	-	-	-
Deep Learning-Based Transfer Learning for Classification of Skin Cancer	classification	HAM10000	VGG19, InceptionV3, InceptionResNetV2, ResNet50, Xception, MobileNet	Xception	0.9048	0.48	-	0.58	0.52	-
DenseNet-II: an improved deep convolutional neural network for melanoma cancer detection	detection	HAM10000	ResNet, DenseNet, Inception, and VGG	DenseNet	0.9627	0.971	-	0.973	0.9735	-

Table 4 (continued)

Paper	Task	Dataset	Networks	Efficient Method	Accuracy	Sensitivity (Recall)	Specificity	Precision	F-measure	AUC-ROC
DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopy Images	Classification, diagnosis	ISIC 2020, HAM10000, DermIS	DSCC_Net, ResNet-152, VGG-19, MobileNet, VGG-16, EfficientNet-B0, Inception-V3	DSCC_Net	0.9417	0.9428	-	0.9376	0.9393	-
Explainable deep inherent learning for multi-classes skin lesion classification	classification	HAM10000, ISIC 2018	inherent deep learning	XAI	0.9289	0.5857	0.9557	0.7685	0.6314	-
Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopy Images	classification	HAM10000	logistic regression, linear discriminant analysis, k-nearest neighbors classifier, decision tree classifier, Gaussian naive Bayes, VGG16, Xception, ResNet50	VGGNet	0.88	0.71	-	0.68	0.70	-
Machine learning approach in melanoma cancer stage detection	Classification, detection	ISIC 2019	CNN, SMTP (Similarity Measure for Text Processing)	SMTP	0.96	0.9804	-	0.9444	0.9596	-
Melanoma and Nevus Skin Lesion Classification Using Handcraft and Deep Learning Feature Fusion via Mutual Information Measures	Detection, classification	ISIC 2018	Linear Regression, SVM, ResNet50, Xception, VGG16, MobileNet, DenseNet	MobileNet	0.924	0.8641	0.90	0.9208	0.8916	0.8964

Table 4 (continued)

Paper	Task	Dataset	Networks	Efficient Method	Accuracy	Sensitivity (Recall)	Specificity	Precision	F-measure	AUC-ROC
Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopy Images	classification	ISIC 2016, 2017, 2020	DCNN	DCNN	ISIC 2016: 0.8141, ISIC 2017: 0.8823, ISIC 2020: 0.9042	ISIC 2016: 0.813, ISIC 2017: 0.878, ISIC 2020: 0.903	-	ISIC 2016: 0.818, ISIC 2017: 0.878, ISIC 2020: 0.904		-
Melanoma diagnosis using deep learning techniques on dermatoscopic images	segmentation, classifcation, detection, diagnosis	ISBI 2017	ResNet152	eVida	0.904	0.820	0.925	-	-	0.872
On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning	Classification, detection	HAM10000	Resnet50, Xception, GoogleNet, MobileNet, DenseNet, Inception	DenseNet	0.829	0.736	0.96	0.785	0.744	-
Recent Advances in Melanoma Diagnosis and Prognosis Using Machine Learning Methods	Classification, diagnosis, prognosis	HAM10000	ResNet, VGG-16, InceptionV3	ResNet, VGG-16, InceptionV3	0.766	0.8956	0.879	-	-	-
Refined Residual Deep Convolutional Network for Skin Lesion Classification	classification	PH2, DermIS and Quest, MED-NODE, ISIC 2016, ISIC 2017, and ISIC 2018	residual deep convolutional neural network (RDCNN)	residual deep convolutional neural network (RDCNN)	0.9505	0.8286	0.97	-	-	-

Table 4 (continued)

Paper	Task	Dataset	Networks	Efficient Method	Accuracy	Sensitivity (Recall)	Specificity	Precision	F-measure	AUC-ROC
SCDNet: A Deep Learning-Based Framework for the Multi-classification of Skin Cancer Using Dermoscopy Image	classification	ISIC 2019	VGG16 with convolutional neural networks (CNN)	SDCNet	0.9691	0.9218	-	0.9219	0.9218	-
Skin Cancer Classification using Deep Learning and Transfer Learning	classification	PH2	SVM, ANN	DCNN	0.9861	0.9833	0.9893	0.9773	-	-
Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network	Classification, detection	ISIC 2017	FCRN, U-Net, AlexNet, VGG-16, Inception, ResNet	FCRN (Fully Convolutional Residual Networks)	0.857	0.490	0.961	0.729	-	0.912
Skin lesion classification with ensembles of deep convolutional neural networks	classification	ISBI 2017	ResNet, GoogleNet, VGG, AlexNet, Sum of the maximal probabilities (SMP)	SMP (Sum of the maximal probabilities)	0.866	0.556	0.785	-	-	0.891
Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks	segmentation	ISBI 2017, PH2	FrCN, FCN, U-Net, SegNet	FrCN(full resolution convolutional networks)	ISBI 2017: 0.9403 PH2: 0.9508	ISBI2017: 0.854 PH2: 0.9372	ISBI2017: 0.9669 PH2: 0.9565	-	-	ISBI2017: 0.9319 PH2: 0.9468

Table 4 (continued)

Paper	Task	Dataset	Networks	Efficient Method	Accuracy	Sensitivity (Recall)	Specificity	Precision	F-measure	AUC-ROC
Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer learning	classification	ISIC 2019	GoogleNet and multi-class SVM	transfer learning and pre-trained model with GoogleNet	0.9492	0.798	0.97	0.8036	-	-
Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks	classification	MED-NODE, DermIS & DermQuest and ISIC 2017	Alex-net, ResNet101, and GoogleNet	GoogleNet + Transfer learning	MED-NODE = 0.9929 DermIs&DermQuest = 0.9915 ISIC2017 = 0.9814	DermIs & DermQuest = 0.9892 MED-NODE = 0.9922 ISIC2017 = 0.9727	DermIs & DermQuest = 0.9941 MED-NODE = 0.9938 ISIC2017 = 0.986	-	-	-
Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier	classification	ISIC 2019, 2020	Transfer learning to VGG19, Fuzzy C-means and Red Fox Optimization	Deep Convolution Neural Network	0.9869	0.9543	0.9928	0.9543	0.9379	-
SNC_Net: Skin Cancer Detection by Integrating Handcrafted and Deep Learning-Based Features Using Dermoscopy Images	Classification, detection	ISIC 2019		SNC_Net (With SMOTE Tomek)	0.9781	0.9789	-	0.9831	0.9810	0.9967

Table 4 (continued)

Paper	Task	Dataset	Networks	Efficient Method	Accuracy	Sensitivity (Recall)	Specificity	Precision	F-measure	AUC-ROC
The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning	classification	Collected dataset	FRCNN, BCD, TRN	FRCNN (faster, region-based convolutional neural network)	0.915	0.833	0.945	-	-	-
The skin cancer classification using deep convolutional neural network	classification	Collected dataset	ECOC SVM, and deep convolutional neural network	ECOC SVM, and deep convolutional neural network	0.942	0.9783	0.9074	-	-	-
Untangling Classification Methods for Melanoma Skin Cancer	classification	ISIC 2021	XG-Boost, VGG16, Densenet121, Xception, InceptionV3, EfficientNetB0, ResNet50V2	XG-Boost	0.9722	0.1260	-	0.5568	0.1401	0.1578

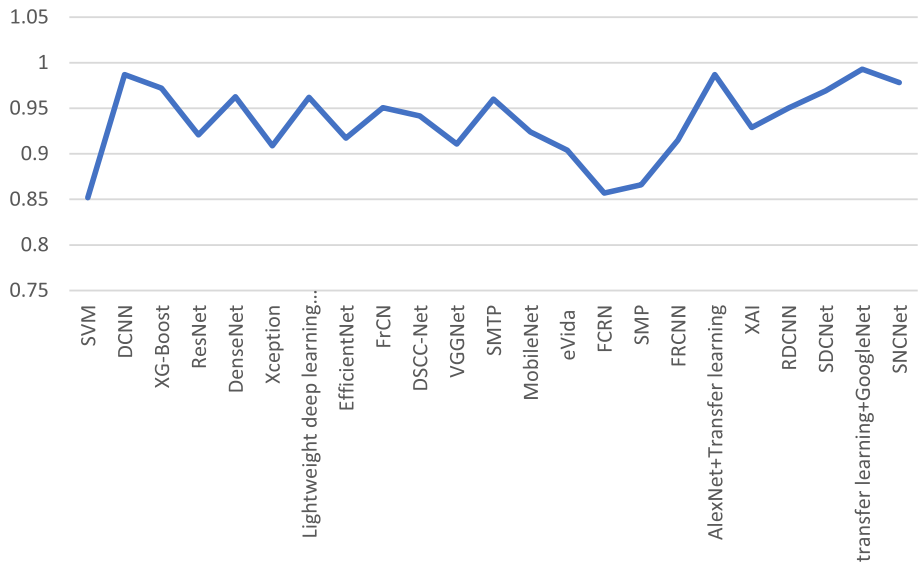

Fig. 8 The accuracy of each algorithm in different datasets

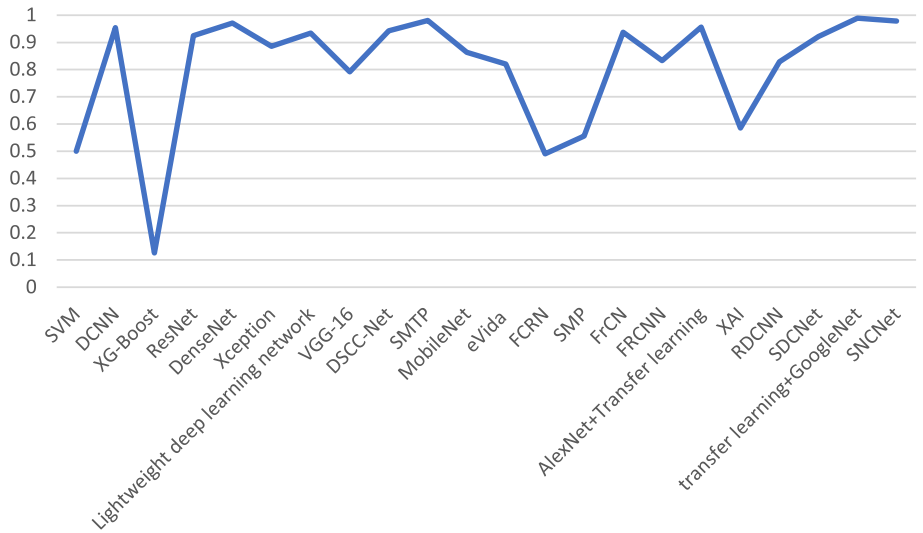

Fig. 9 The sensitivity of each algorithm in different datasets

MobileNet, or efficient compression techniques should be explored. These models can make AI-based diagnostics more accessible in a wider range of clinical settings [47].

While melanoma diagnosis has been widely studied, less emphasis has been placed on the use of machine learning and deep learning models to predict melanoma prognosis, such as survival rates and recurrence risks. More emphasis should be placed on integrating prognostic capabilities into machine learning and deep learning models. This could include incorporating additional clinical data, such as patient history, genetic markers, or treatment responses, to enhance predictive modeling for long-term outcomes.

Although generative AI models, such as Generative Adversarial Networks (GANs), have shown great potential in augmenting small datasets by generating synthetic images, their use in melanoma diagnosis remains underexplored. Generative AI could help overcome the issue of limited data by creating realistic dermoscopy images to improve model training. Generative AI techniques, such as GANs, should be used to expand the dataset by generating artificial yet realistic dermoscopy images. This can enhance model training and improve robustness, particularly when dealing with rare conditions like melanoma.

Abbreviations

ML	Machine Learning
DL	Deep Learning
CAD	computer-aided diagnostic
CNN	Convolutional Neural Network
UV	Ultraviolet
FL	Federated Learning
TL	Transfer Learning
SVM	Support vector machines
RF	Random Forests
LR	Logistic Regression
LDA	Linear Discriminant Analysis
KNN	k-nearest Neighbors classifier
CART	Decision Tree Classifier
ABCDE	asymmetry, borders, color, diameter, and evolution
ISIC	International Skin Imaging Collaboration
ISBI	International Symposium on Biomedical Imaging
MRI	magnetic resonance imaging
CT	computed tomography
DCNN	Deep convolutional neural networks
VGGNet	Visual Geometry Group Network
ResNet	Deep residual neural network
DSCC_Net	Deep learning-based skin cancer classification network
FCRN	Fully Convolutional Residual Networks
FrCN	Full resolution convolutional networks
FRCNN	Faster, region-based convolutional neural network
SMTP	Similarity Measure for Text Processing
AUC-ROC	Area under the Receiver Operating Characteristic Curve
GANs	Generative Adversarial Networks
LIME	Local Interpretable Model-Agnostic Explanations
Grad-CAM	Gradient-weighted Class Activation Mapping

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-13423-y.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
Hoda Naseri conducted the primary research and drafted the manuscript. Ali A. Safaei supervised the research process, provided critical revisions, and contributed to the final version of the manuscript.

Funding
Not applicable.

Data availability
The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 18 September 2024 Accepted: 31 December 2024
Published online: 13 January 2025

References

1. Thurnhofer-Hemsi K, Dominguez E. A convolutional neural network framework for accurate skin cancer detection. Springer. 2020. https://doi.org/10.1007/s11063-020-10364-y.
2. Kricker A, Armstrong BK, English DR. Sun exposure and non-melanocytic skin cancer, Armstrong BK, Kricker A. The epidemiology of UV induced skin cancer. Springer; 1994. https://doi.org/10.1007/bf01804988.
3. Ciążyńska M, Kamińska-Winciorek G, Lange D. The incidence and clinical analysis of nonmelanoma skin cancer. Nature. 2021. https://doi.org/10.1038/s41598-021-94435-7.
4. Hodis E. The somatic genetics of human melanoma. Ph.D. Thesis. Cambridge: Harvard University; 2018.
5. Bassel A, Abdulkareem A, Abdi Alkareen Akyasseri Z, Sani N, Mohammed H. Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. MDPI. 2022. https://doi.org/10.3390/diagnostics12102472.
6. Guha SR, Rafizul Haque SM. Convolutional neural network based skin lesion analysis for classifying melanoma. IEEE; 2019. https://doi.org/10.1109/STI47673.2019.9067979.
7. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, Hadj HAB, Thomas L, Enk A, et al. Reader study level-I and level-II groups, man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopy melanoma recognition in comparison to 58 dermatologists. Elsevier. 2018. https://doi.org/10.1093/annonc/mdy166.
8. Al-Masni MA, Al-Antari MA, Choi MT, Han SM, Kim TS. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Elsevier. 2018. https://doi.org/10.1016/j.cmpb.2018.05.027.
9. Kaur R, GholamHosseini H, Sinha R, Lindén M. Melanoma classification using a novel deep convolutional neural network with dermoscopy images. MDPI. 2022. https://doi.org/10.3390/s22031134.
10. Girdhar N, Sinha A, Gupta S. DenseNet-II: an improved deep convolutional neural network for melanoma cancer detection. Springer. 2022. https://doi.org/10.1007/s00500-022-07406-z.
11. Hosseinzadeh Kassani S, Hosseinzadeh Kassani P. A comparative study of deep learning architectures on melanoma detection. Elsevier. 2019. https://doi.org/10.1016/j.tice.2019.04.009.
12. Patil R, Bellary S. Machine learning approach in melanoma cancer stage detection. 2020.
13. Harangi B. Skin lesion classification with ensembles of deep convolutional neural networks. Elsevier. 2018. https://doi.org/10.1016/j.jbi.2018.08.006.
14. Jaisakthi SM, Mirunalini P, Aravindan C, Appavu R. Classification of skin cancer from dermoscopy images using deep neural network architectures. Springer. 2022. https://doi.org/10.1007/s11042-022-13847-3.
15. Jojoa Acosta MF, Caballero Tovar LY, Garcia-Zapirain MB, Percybrooks WS. Melanoma diagnosis using deep learning techniques on dermatoscopic images. Springer. 2021. https://doi.org/10.1186/s12880-020-00534-8.
16. Grossarth S, Mosley D, Madden C, Ike J, Smith I, Huo Y, Wheless L. Recent advances in melanoma diagnosis and prognosis using machine learning methods. Springer. 2023. https://doi.org/10.1007/s11912-023-01407-3.
17. American Cancer Society. 2018. Available: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html.
18. El-Khatib H, Popescu D, Ichim L. Deep learning–based methods for automatic diagnosis of skin lesions. MDPI. 2020. https://doi.org/10.3390/s20061753.
19. Jain S, Singhania U, Tripathy B, Abouel Nasr E, Aboudaif K, Kamrani M. Deep learning-based transfer learning for classification of skin cancer. MDPI. 2021. https://doi.org/10.3390/s21238142.
20. Tahir M, Naeem A, Malik H, Tanveer J, Ali Naqvi R, Lee SW. DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopy images. MDPI. 2023. https://doi.org/10.3390/cancers15072179.

21. Bechelli S, Delhommelle J. Machine learning and deep learning algorithms for skin cancer classification from dermoscopy images. MDPI. 2022. https://doi.org/10.3390/bioengineering9030097.
22. Almaraz-Damian JA, Ponomaryov V, Sadovnychiy S, Castillejos-Fernandez H. Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual Information measures. MDPI. 2020. https://doi.org/10.3390/e22040484.
23. Fraiwan M, Faouri E. On the automatic detection and classification of skin cancer using deep transfer learning. MDPI. 2022. https://doi.org/10.3390/s22134963.
24. Li Y, Shen L. Skin lesion analysis towards melanoma detection using deep learning network. MDPI. 2018. https://doi.org/10.3390/s18020556.
25. Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R. The development of a skin cancer classification system for pigmented skin lesions using deep learning. MDPI. 2020. https://doi.org/10.3390/biom10081123.
26. Wei L, Ding K, Hu H. Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. IEEE; 2020. https://doi.org/10.1109/access.2020.2997710.
27. Adegun AA, Viriri S. Deep learning-based system for automatic melanoma detection. IEEE; 2020. https://doi.org/10.1109/ACCESS.2019.2962812.
28. Seeja RD, Suresh A. Untangling classification methods for melanoma skin cancer. Frontiers. 2022. https://doi.org/10.3389/fdata.2022.848614.
29. Kumar A, Vatsa A. Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM), Asian Pacific. J Cancer Prev. 2023;20:1555.
30. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform. 2002;35(5–6):352–9. https://doi.org/10.1016/S1532-0464(03)00034-0.
31. Li M, Yuan B. 2d-lda: a statistical linear discriminant analysis for image matrix. Pattern Recogn Lett Elsevier. 2005. https://doi.org/10.1016/j.patrec.2004.09.007.
32. Fayek HM, Lech M, Cavedon L. Evaluating deep learning architectures for speech emotion recognition. Elsevier. 2017. https://doi.org/10.1016/j.neunet.2017.02.013.
33. Pathaka AR, Pandeya M, Rautaray S. Application of deep learning for object detection. Elsevier. 2018. https://doi.org/10.1016/j.procs.2018.05.144.
34. Sun S. A review of natural language processing techniques for opinion mining systems. Elsevier. 2016. https://doi.org/10.1016/j.inffus.2016.10.004.
35. Chollet F. Deep learning with depthwise separable convolutions. 2016. https://doi.org/10.48550/arXiv.1610.02357.
36. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. 2012. https://doi.org/10.1145/3065386.
37. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. https://doi.org/10.48550/arXiv.1409.1556.
38. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015. https://doi.org/10.48550/arXiv.1512.03385.
39. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Spriner. 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
40. Tan M, Le QV. Efficientnet: rethinking model scaling for convolutional neural networks. 2019. https://doi.org/10.48550/arXiv.1905.11946.
41. Yu L, Chen H, Dou Q, Qin J. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE. 2016. https://doi.org/10.1109/TMI.2016.2642839.
42. D'Orazio J, Jarrett S, Amaro-Ortiz A, Scott T. UV radiation and the skin. 2013. https://doi.org/10.3390/ijms140612222.
43. McCourt C, Dolan O, Gormley G. Malignant melanoma: a pictorial review. Ulster Med J. 2014;83:103 PMID: https://pubmed.ncbi.nlm.nih.gov/25075139 ; PMCID: PMC4113154.
44. Holmes GA, Vassantachart JM, Limone BA, Zumwalt M, Hirokane J, Jacob SE. Using dermoscopy to identify melanoma and improve diagnostic discrimination. Fed Pract. 2018;35(Suppl 4):S39-45 PMID: 30766399; PMCID: PMC6375419.
45. Grant SR, Andrew TW, Alvarez EV, Huss WJ, Paragh G. Diagnostic and prognostic deep learning applications for histological assessment of cutaneous melanoma. Cancers (Basel). 2022;14(24):6231. https://doi.org/10.3390/cancers14246231. PMID: 36551716; PMCID: PMC9776963.
46. Das K, Cockerell CJ, Patil A, Pietkiewicz P, Giulini M, Grabbe S, Goldust M. Machine learning and its application in skin cancer. Int J Environ Res Public Health. 2021;18(24):13409. https://doi.org/10.3390/ijerph182413409. PMID: 34949015; PMCID: PMC8705277.
47. Howard G, Zhu A, Chen M, Kalenichenko B, Wang D, Weyand W, Andreetto T, Adam MH. MobileNets: efficient convolutional neural networks for mobile vision applications. 2017. https://doi.org/10.48550/arXiv.1704.04861.
48. Kassem MA, Hosny KM, Damaševi˘cius R, Eltoukhy MM. Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. Diagnostics. 2021;11:1390. https://doi.org/10.3390/diagnostics11081390.
49. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71.
50. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. PLoS Med. 2009;6:E1000097. https://doi.org/10.1371/journal.pmed.1000097.
51. Giotis N, Molders S, Land M, Biehl MF, Jonkman, Petkov N. MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. Expert Syst Appl. 2015;42:6578–85. https://doi.org/10.1016/j.eswa.2015.04.034.
52. Farhat Ullah Skin Lesion Dermis Dataset. 2022. Available online: https://www.kaggle.com/datasets/farhatullah8398/skin-lesion-dermis-dataset. Accessed 15 Jul 2022.
53. DermQuest. 2012. http://www.dermquest.com.
54. Naeem A, Anees T. "DermPK", Mendeley Data, V1. 2023. https://doi.org/10.17632/fdhyjjypbd.1.
55. Riaz S, Naeem A, Malik H, Naqvi RA, Loh W-K. Federated and transfer learning methods for the classification of melanoma and nonmelanoma skin cancers: a prospective study. Sensors. 2023;23:8457. https://doi.org/10.3390/s23208457.
56. Naeem A, Khan AH, Ayubi S, Malik H. Predicting the metastasis ability of prostate cancer using machine learning classifiers. 2023. https://doi.org/10.56979/402/2023.
57. Naeem A, Anees T. A Multiclassification Framework for Skin Cancer detection by the concatenation of Xception and ResNet101. 2024. https://doi.org/10.56979/602/2024.
58. Hosny KM, Kassem MA, Foaud MM. Classification of skin lesions using transfer learning and augmentation with Alex-net. PLoS One. 2019;14(5):e0217293. https://doi.org/10.1371/journal.pone.0217293.
59. Naeem A, Anees T, Fiza M, Naqvi RA, Lee S-W, SCDNet. A deep learning-based framework for the multiclassification of skin cancer using dermoscopy images. Sensors. 2022;22:5652. https://doi.org/10.3390/s22155652.
60. Alsahaf YS, Kassem MA, Hosny KM. Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier. Springer J Big data. 2023;10:105. https://doi.org/10.1186/s40537-023-00769-6.
61. Naeem A, Anees T, Khalil M, Zahra K, Naqvi RA, Lee S-W. SNC_Net: skin cancer detection by integrating handcrafted and deep learning-based features using dermoscopy images. Mathematics. 2024;12:1030. https://doi.org/10.3390/math12071030.
62. Kassem MA, Hosny KM, Foaud MM. Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access. 2020. https://doi.org/10.1109/ACCESS.2020.3003890.
63. Hosny KM, Kassem MA, Foaud MM. Skin cancer classification using deep learning and transfer learning. IEEE. 2018. https://doi.org/10.1109/CIBEC.2018.8641762.
64. Naeem A, Anees T. DVFNet: a deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images. PLoS One. 2024;19(3):e0297667. https://doi.org/10.1371/journal.pone.0297667.
65. Hosny KM, Kassem MA. Refined residual deep convolutional network for skin lesion classification. Springer. 2022. https://doi.org/10.1007/s10278-021-00552-0.

66. Dorj UO, Lee KK, Choi JY, Lee M. The skin cancer classification using deep convolutional neural network. Springer. 2018. https://doi.org/10.1007/s11042-018-5714-1.
67. Hosny KM, Kassem MA, Foaud MM. Classification of skin lesions into seven classes using transfer learning with AlexNet. Springer. 2020. https://doi.org/10.1007/s10278-020-00371-9.
68. Hosny KM, Kassem MA, Foaud MM. Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks. Springer. 2020. https://doi.org/10.1007/s11042-020-09067-2.
69. Hosny KM, Said W, Elmezain M, Kassem MA. Explainable deep inherent learning for multi-classes skin lesion classification. Elsevier. 2024. https://doi.org/10.1016/j.asoc.2024.111624.

Publisher's Note