A comprehensive analysis of different models: skin cancer detection

Amruta Thorat[1], Chaya Jadhav[2]
[1]Department of Electronics and Telecommunication Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India
[2]Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India

Article Info

Article history:

Received Jun 26, 2024
Revised Dec 9, 2024
Accepted Dec 19, 2024

Keywords:

Alex-Net
Deep learning
EfficientNet
ImageNet
Machine learning
MobileNet-V2
ResNet

ABSTRACT

Due to fast-growing worldwide air pollution and ozone layer destruction, an alarming number of people are found to have skin cancer, more than any other kind of cancer combined. It is known to be one of the deadliest malignancies; if not identified and cured in its early stages, it is likely to spread to other body parts. Early detection is critical and helps prevent cancer from spreading. This allows for early decisions on diagnostic and treatment options. Early diagnosis and discovery, combined with the right treatment, can save lives. In this paper, we have done a detailed survey on various techniques and models developed for skin cancer detection and also discussed different security-related issues. This work thoroughly explores the several types of models utilized to identify cancer in the skin.

Corresponding Author:

Amruta Thorat
Department of Electronics and Telecommunication Engineering, Dr. D. Y. Patil Institute of Technology
Pimpri, Pune, India
Email: amrutathorat1206@gmail.com

1. INTRODUCTION

The National Cancer Institute (NCI) predicts that by 2040, there might be an astounding 29.5 million new cases [1]. World health organization (WHO) predicts one of the three most deadly cancers that can result from deoxyribonucleic acid (DNA) damage is skin cancer. The uncontrollable growing of tissues, is caused by this damaged DNA, and it is currently accelerating quickly. This type of uneven growth of cell patterns can be classed as either benign or malignant [2]. Skin cancer may also be caused by ultraviolet (UV) light exposure, a weakened immune system, a family history, and other factors.

The nature of human skin is incredibly complex. The skin shields all the organs within from the harsh external environment. It guards against infections and assists in temperature regulation. As shown in Figure 1, there are three layers of skin, which consist hypodermis epidermis and dermis [3]. The epidermis and dermis are the two principal membrane divisions. The epidermis is the skin's top layer and acts as a protective barrier by keeping bodily fluids in place and preventing bacterial growth. The skin's tensile strength and suppleness are derived from the connective tissues that comprise the dermis [3]. Figure 2 shows the different types of skin cancer, the three main categories of skin cancers are: i) Squamous-cell carcinoma (SCC), ii) basal-cell carcinoma (BCC), and iii) malignant melanoma [3], the other types are actinic keratosis (AKIEC) has a benign nature but can become malignant turning into a cancerous lesion. Basal cell carcinoma (BCC) and melanoma (MEL) are cancerous. Benign keratosis (BKL), dermatofibroma (DF), vascular skin lesion (VASC), and melanocytic nevi (NV) are non-cancerous [4]. There are studies for the automated identification of cancer in images of skin lesions. The analysis of these images, however, is very difficult due to some significant causes, such as the reflection of light from the skin's surface, differences in the color brightness of the lesions, and the lesions' varied sizes and forms.

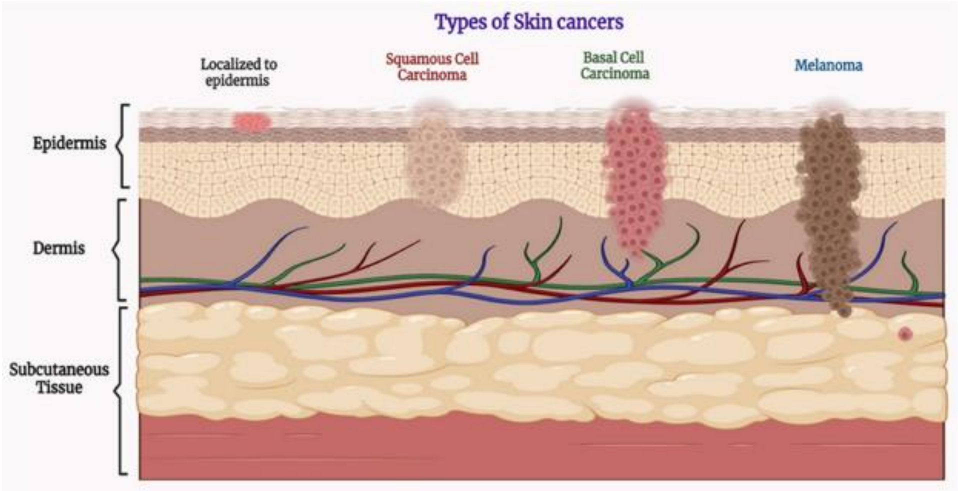

Figure 1. Layers of skin and basic three types of skin cancer [3]

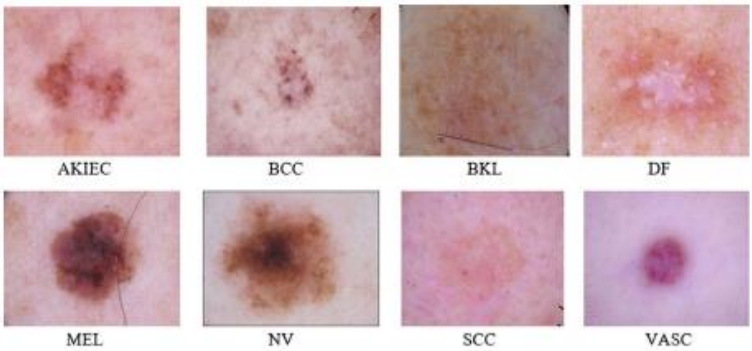

Figure 2. Types of skin cancers [4]

The nature of human skin is incredibly complicated. The challenging part is identifying skin cancer from clinical photographs. However, for physicians it is extremely difficult to manually diagnose skin issues. Consequently, computerized skin disease prediction is needed for both patients and doctors. Therefore, it is important to improve pathologists' accuracy and expertise in the early stages through evidential automatic skin cancer recognition.

The study explores advanced techniques and features designed for the detection and classification of skin cancer. It emphasizes several prominent deep learning models, including AlexNet, MobileNet-V2, EfficientNet, ImageNet, ResNet, VGG-16, DenseNet, InceptionV3, and image super-resolution (ISR), which have been instrumental in improving diagnostic precision. These models showcase significant advancements in accuracy, efficiency, and reliability, contributing to more effective automated skin cancer diagnosis.

Skin cancer impact on human life can be profound. Skin cancer can significantly affect an individual's quality of life. While early-stage skin cancers can often be treated effectively, advanced cases, especially melanoma, can be life-threatening. The disease can lead to physical and psychological burdens, including disfigurement, anxiety, and reduced life expectancy. Survival rates can be improved by early detection and treatment.

Skin tests for manual diagnostic methods usually include visual examination and biopsy. Skin tests usually include:

- Visual examination: Dentists use instruments such as dermoscopes to examine the skin and identify abnormalities based on symptoms (*e.g.*, asymmetry, irregularity, irregular border, discoloration).
- Biopsy: Perform a biopsy if a lesion is present. A small piece of skin is collected and examined under a microscope to detect the presence of cancer cells. Although successful, cell diagnosis relies heavily on the knowledge and skills of a dermatologist. Misdiagnosis or delayed diagnosis is possible, especially if symptoms are vague or atypical.

Machine learning (ML) and deep learning (DL) play a transformative role in skin cancer detection by enabling faster and more accurate diagnoses. Deep learning models, especially convolutional neural networks (CNNs), are trained on large datasets of dermoscopic images to detect and classify skin lesions, often surpassing the accuracy of human experts. These models assist dermatologists in identifying early-stage cancer, reducing the need for invasive biopsies through non-invasive image analysis. ML algorithms can also integrate additional patient data to improve diagnosis and predict cancer risk. By augmenting clinical workflows, artificial intelligence (AI) driven systems enhance diagnostic consistency, speed, and overall patient outcomes. This technology is helping to revolutionize skin cancer screening and early detection.

They can improve skin diagnosis by i) image classification, ii) early detection, iii) decision support, and iv) automated screening. Recent technological advancements have focused on the development of deep learning and machine learning, resulting in improved accuracy, speed, and scalability for early detection and enhanced efficiency. The main contributions include a detailed survey of the most prominent models for skin cancer detection and comparative analysis of the accuracy and reliability of different models.

This paper thoroughly surveys the various models and techniques developed for skin cancer detection, including deep learning models such as AlexNet, MobileNet-V2, EfficientNet, ImageNet, ResNet, VGG-16, DenseNet, and InceptionV3. The remaining sections of this paper are organized as follows: In section 2, relevant methods used by researchers on deep learning and machine learning for the identification of skin cancer are summarized. Section 3 outlines present the results and discussion of the models' performance using various datasets, including comparisons between different architectures. In section 4, concludes the paper with future research directions.

2. METHOD

Skin cancer continues to be one of the most widespread and dangerous types of cancer globally, underscoring the importance of early detection in enhancing survival rates. While traditional diagnostic techniques, including biopsies and dermoscopic evaluations, are effective, they tend to be labor-intensive and demand specialized knowledge. Recent developments in AI, especially in the realms of deep learning and machine learning, present promising avenues for automating and improving the precision of skin cancer detection.

2.1. Deep learning

2.1.1. Transfer learning on original and augmented data

The study explored two distinct categories of transfer learning (TL) methods: TL applied to original data and TL applied to augmented data, using the PAD-UFES-20 dataset. A significant finding was that data augmentation enhanced the performance of models, highlighting its importance in skin cancer classification tasks. Among the models evaluated, AlexNet demonstrated the best performance, showcasing its potential for integration into mobile applications aimed at improving skin cancer detection and diagnosis [5].

2.1.2. Comparison of dermoscopic and smartphone images for CNN-based detection

This study examined the diagnostic accuracy of dermoscopic images (DI) versus smartphone-captured images (SI) using a dual CNN with sonification for non-melanoma skin cancer (NMSC). Results showed that DI outperformed SI in accuracy, sensitivity, and AUC, suggesting telemedicine approaches may require dermoscopic images for optimal results. Methods included a preprocessing pipeline with hair removal, data augmentation, and resizing, with EfficientNet B4 yielding the best results (F1-score: 87%, Top-1 accuracy: 87.91%) [6], [7].

2.1.3. Skin lesion classification using ResNet, Xception, and DenseNet

The study utilized ResNet, Xception, and DenseNet models to classify skin lesions using the HAM10000 dataset. By employing a weighted ensemble technique, the study achieved an accuracy of 85.8%, outperforming the individual models. These findings underscore the effectiveness of ensemble methods in enhancing the performance of skin lesion classification tasks, providing a robust approach for improving diagnostic accuracy [8].

2.1.4. Challenges in skin image classification

The study emphasized the complexity of skin image classification due to the high variability in lesion appearance and characteristics. A CNN implemented in TensorFlow achieved an accuracy of 81.24%, showcasing its potential despite the challenges. However, transfer learning models in PyTorch, including Wide ResNet101, ResNet50, DenseNet121, and VGG19, significantly improved accuracy, ranging from 96.40% to 99.04%, highlighting their superior performance in addressing these complexities [9].

2.1.5. Optimized CNN model with RMSprop and ADAM

A convolutional neural network model pre-trained on dermoscopic images was further refined using highway CNN features. The optimization process employed both RMSprop and ADAM, two widely used algorithms for deep learning tasks. Notably, the ADAM optimizer outperformed RMSprop, achieving a training accuracy of 90% and a validation accuracy of 82% [10].

2.1.6. Feature selection using genetic algorithm and particle swarm optimization

The study utilized EfficientNetB0 CNN features to enhance the accuracy of classification tasks. Feature selection was conducted using genetic algorithm (GA) and particle swarm optimization (PSO), two powerful optimization techniques widely applied in machine learning. Following this process, support vector machine (SVM) classification was performed, achieving an impressive accuracy of 89.17% [11].

2.1.7. Deep learning in diagnosing pigmented nevi

The study utilized CNNs to classify Dermoscopic images of pigmented nevi, aiming to distinguish between benign and malignant lesions. The use of CNNs provided a robust framework for analyzing complex patterns and textures within the images, which are critical for accurate classification. By leveraging these deep learning models, the study highlighted the potential of CNNs to outperform traditional diagnostic methods in terms of accuracy and efficiency [12].

2.1.8. Ensemble model with Xception, ResNet50, and VGG16

The study focused on fine-tuning an ensemble model composed of Xception, ResNet50, and VGG16 to enhance melanoma diagnosis. By employing a weighted fusion approach, the ensemble achieved an accuracy of 86.91%, surpassing the performance of many traditional diagnostic methods. These results underscore the effectiveness of integrating multiple deep learning architectures to harness their complementary strengths [13].

2.1.9. Image super-resolution and CNN for enhanced detection

The study employed a combined approach using ISR and CNNs to enhance the classification accuracy for various skin cancer types. The integration of ISR improved image quality, enabling CNN models to extract more detailed and informative features. Among the tested models, InceptionV3 demonstrated significant diagnostic improvements, highlighting the potential of this approach [14].

2.1.10. Evaluation of pre-trained networks on MED-NODE and DermIS datasets

Five pre-trained networks (AlexNet, ResNet-18, SqueezeNet, ShuffleNet, DarkNet-19) were evaluated. AlexNet and ResNet-18 achieved top precision and accuracy scores (up to 100%), with other models also performing well, albeit with slightly lower accuracy [15]. Methods includes:

- Data preparation and augmentation: Clearly outline dataset preparation steps (*e.g.*, hair removal, image resizing, data augmentation), specifying parameters used for each model.
- Model architectures and optimization techniques: Detail the configurations and optimizers (*e.g.*, ADAM, RMSprop) used for each CNN or transfer learning model.
- Experimental setup: Provide specifications for model training (*e.g.*, batch sizes, learning rates, epochs) and describe ensemble techniques, if applicable.
- Performance metrics: Report metrics like accuracy, sensitivity, AUC, and F1-scores for each model, including thresholds or decision criteria for classification.

2.2. Deep convolutional neural network (DCNN)

2.2.1. Comprehensive dataset description

Describe the HAM10000 dataset, used extensively in studies [16]–[18]. Detail skin lesion types, sample sizes, and any relevant metadata, including patient demographics used in study [18] for the multimodal model. Mention any data imbalance issues and the techniques employed to address these, as seen in study [17].

2.2.2. Preprocessing steps

Preprocessing techniques play a crucial role in improving the performance of skin lesion classification models by preparing the data for more effective analysis and learning.

- Data augmentation: Detail augmentation methods such as rotation, flipping, and scaling, as applied in study [16] to improve classification accuracy.
- Feature extraction and noise reduction: Study [16] emphasizes feature extraction and noise reduction as part of preprocessing to enhance model performance. Include steps to clarify each technique.

- Image resizing: Mention image resizing specifications for consistency across models, as highlighted in studies [16] and [18].

2.2.3. Model architectures

Recent advancements in deep learning have led to the development of innovative models that significantly enhance the accuracy and robustness of skin lesion classification.

- DCNN models: The proposed DCNN model in [16] integrates various preprocessing steps to improve performance, achieving high training and testing accuracy on HAM10000. Another DCNN developed in [17] demonstrates superiority over VGG16 and VGG19.
- Multimodal model (ALBEF): Study [18] presents a unique multimodal approach by combining dermoscopic images and patient metadata, achieving high accuracy and AUC-ROC.

2.2.4. Optimization algorithms

The optimization techniques in [16] aimed to enhance DCNN performance through adaptive optimizers, parameter tuning, and regularization strategies. ADAM, with a learning rate of 0.001 and a batch size of 32, enabled faster convergence and better outcomes. Regularization methods like dropout (0.5) and L2 weight decay (0.0001) helped prevent overfitting and improved generalization. Data augmentation techniques, including rotation, flipping, and scaling, increased training data diversity, further enhancing robustness. These combined methods played a crucial role in achieving superior performance in the study.

2.2.5. Evaluation metrics

Evaluation metrics such as accuracy, F1-score, and AUC-ROC are essential for assessing model performance, as emphasized in [16]–[18]. Accuracy measures the proportion of correct predictions but may be insufficient for imbalanced datasets, as noted in [16]. The F1-score, highlighted in [17], balances precision and recall, providing a better measure of performance in scenarios with skewed class distributions. AUC-ROC, as discussed in [18], evaluates a model’s ability to distinguish between classes across thresholds, making it crucial for clinical applications where reliability is paramount.

2.2.6. Experimental setup

A well-defined hardware and software setup is crucial for reproducibility in machine learning studies. This includes specifying the computing environment, such as graphics processing unit (GPU) models (*e.g.*, NVIDIA A100) or central processing units (CPUs), and the frameworks used, like TensorFlow or PyTorch, which are widely recognized for their robustness in deep learning. Studies [16]–[18] emphasize the importance of consistent training-validation splits, typically using an 80-20 or 70-30 ratio, along with setting random seed values to ensure reproducibility in dataset partitioning and model initialization. Additionally, hyperparameters such as batch size, learning rate, and number of epochs should be explicitly stated, as demonstrated in study [17], to allow researchers to replicate the results. Furthermore, study [18] highlights the significance of consistent evaluation metrics across repeated runs to enhance reliability in high-stakes applications.

2.2.7. Comparison and analysis

Comparisons within studies highlight the superior performance of certain models due to their architecture and multimodal integration. In study [17], the deep convolutional neural network (DCNN) outperformed VGG models, thanks to its deeper layers and enhanced feature extraction capabilities, allowing it to capture more complex patterns in the data. Similarly, In study [18], the multimodal ALBEF model, which combines image and text data, surpassed image-only models by leveraging richer contextual information. The integration of multiple data sources, along with preprocessing techniques like data augmentation and normalization, contributed to the higher accuracy achieved by the multimodal model, enhancing its generalization and performance across diverse datasets.

2.3. Machine learning

2.3.1 Preprocessing techniques

Preprocessing techniques are crucial for improving the quality of input data and enhancing model performance in skin lesion classification tasks.

- Image enhancement: Include steps for Gaussian and Median filtering to reduce noise [19], [20], and mention digital hair removal techniques such as the Dull Razor method [19], [20].
- Segmentation methods: Describe color-based k-means clustering for segmentation as applied in [19] and the GrabCut technique in [20] to isolate regions of interest effectively.

2.3.2. Feature extraction

Feature extraction plays a pivotal role in improving the performance of skin lesion classification models by extracting relevant patterns and characteristics from images.

- ABCD and gray level co-occurrence matrix (GLCM) methods: Describe asymmetry, border, color, and diameter (ABCD) method and the GLCM for texture analysis [19]. Mention statistical feature extraction as per [20] and the CLBP_SMC method recommended in [21] for melanoma detection.
- Feature extraction via deep learning models: Include VGG16 for feature extraction, combined with XGBoost for classification as in [22], which achieved high accuracy for the skin cancer types mentioned.

2.3.3. Classification techniques

Classification techniques are crucial for accurately categorizing skin lesions and making reliable predictions in medical imaging.

- Machine learning classifiers: Describe classifiers used in studies, such as multi-class support vector machine (MSVM) achieving high accuracy on ISIC 2019 data [19], random forest (RF) for melanoma classification with completed local binary patterns (CLBP) [21], and support vector machine paired with GLCM [20], [23].
- XGBoost and artificial neural networks (ANN): Reference study [22] for combining VGG16 and XGBoost for classification, and study [23] for ANN models using hybrid features.

2.3.4. Datasets and skin cancer categories

The choice of datasets and the inclusion of various skin cancer categories are critical for training accurate and reliable models for skin lesion classification.

- ISIC 2019 dataset: Detail the use of the ISIC 2019 dataset with eight skin cancer types [19], [20].
- HAM10000 and PH2 datasets: Mention the use of HAM10000 and PH2 in study [23], specifying the types of lesions included and any preprocessing steps applied to ensure dataset quality.

2.3.5. Model evaluation metrics

Model evaluation metrics such as accuracy, confusion matrix, precision, recall, and F1-score are commonly used across studies [19]–[23] to assess classifier performance. These metrics help in evaluating how well different models, combined with various feature extraction methods, perform in skin lesion classification. Specifically, study [21] demonstrated the effectiveness of using random forest classifiers with CLBP for feature extraction, achieving high accuracy. The confusion matrix results showed that this combination not only delivered strong classification performance but also handled variations in skin lesion images effectively.

2.3.6. Experimental setup

Include details of hardware and software environments to support reproducibility. Specify ML frameworks used (*e.g.*, TensorFlow, Scikit-learn), batch sizes, and training-validation splits as recommended in each study. Mention seed values, random initialization processes, and cross-validation techniques to ensure replicable and stable results across methods [19]–[23].

2.3.7. Comparative analysis

The comparative results from studies [19], [22] highlight the effectiveness of specific combinations of feature extraction techniques and classifier models in achieving high accuracies for skin lesion classification. In particular, study [19] demonstrated how the integration of texture and statistical features with classifiers like SVM led to significant improvements in melanoma detection. Similarly, study [22] also emphasized the importance of selecting the right feature-extraction and classification models to optimize performance. Meanwhile, the study [24] provides a broader survey, offering insights into performance optimization strategies such as hyperparameter tuning and data augmentation, further enhancing melanoma classification accuracy across different approaches.

2.4. Image processing

2.4.1. Image processing techniques

Image processing techniques play a crucial role in enhancing the quality of input data for machine learning models, especially in skin lesion classification tasks.

- Statistical feature extraction: Describe the use of first and second-order statistical features for texture analysis as referenced in [25]. This should include specific features analyzed (*e.g.*, mean, variance, skewness) and how they contribute to identifying patterns relevant to skin lesion classification.

– Gray level co-occurrence matrix (GLCM): Detail the implementation of GLCM for texture analysis, focusing on how features like contrast, homogeneity, energy, and correlation were extracted from different color channels [25].

2.4.2. Novel metadata-enhanced classification technique

The novel metadata-enhanced classification technique offers a powerful approach to improving classification performance by integrating additional contextual information into the model training process.

– Metadata integration: Provide a clear description of the metadata-enhanced classification method presented in [26], which amplifies key features. Explain the process of boosting critical features in the classification pipeline and how it improves model robustness.
– Performance across models: Mention that the method was tested on two skin lesion datasets and outperformed other approaches in six out of ten scenarios [26]. List the classification models used and specify any preprocessing techniques or feature selection methods that were applied to optimize performance.

2.4.3. Skin lesion classification model

The skin lesion classification model is designed to accurately categorize and classify skin lesions into distinct categories, leveraging deep learning techniques for high-performance results.

– Classification categories and dataset: Mention that the proposed model is capable of recognizing seven skin lesion categories and was tested on the HAM10000 dataset [27].
– Model architecture and training: Describe the model's architecture and hyperparameters (*e.g.*, learning rate, batch size, number of epochs) and note any data augmentation methods used to balance the dataset.
– Evaluation metrics: Report the achieved accuracy (90%), precision (0.89), and recall (90%) on the HAM10000 dataset as seen in [27].

2.4.4. Experimental setup

The experimental setup outlines the key components of the hardware, software environments, and data handling strategies used to ensure the reproducibility and reliability of the machine learning experiments.

– Hardware and software environments: Provide details on the hardware and software environments, including machine specifications, deep learning frameworks, and libraries used for feature extraction (such as OpenCV or Python's skimage).
– Training-validation split: Specify training-validation splits, cross-validation approaches, and random seed values for reproducibility.

2.4.5. Comparative performance analysis

The model comparisons reveal notable advancements in classification performance through innovative approaches. Study [26] demonstrated that integrating metadata into classification tasks significantly enhances performance across various scenarios, showcasing the value of enriched data representations. Similarly, study [27] achieved high classification metrics on the HAM10000 dataset, indicating the effectiveness of their approach in accurately diagnosing skin lesions. Table 1 further summarizes the dataset and model performance, providing a comprehensive overview of the results and emphasizing the impact of these methods in improving classification accuracy.

Clinical studies have shown significant progress in the use of deep learning for classification, especially in CNN architectures, transfer learning, and complementary tools. Advanced models such as EfficientNet show good results, while simple pre-designed models with optimization also perform well. Metadata collection and aggregation technology further improves accuracy. However, there are still problems in ensuring the reliability of the structure of different data. Future research should focus on combining AI with traditional methods to provide more accurate and easier skin cancer diagnosis, facilitate early intervention, and improve patient outcomes.

It has been observed that still work can be carried out, Firstly, feature extraction methods need to integrate multi-modal data sources, using advanced deep learning techniques to enhance accuracy and robustness. Secondly, optimization is needed for real-time processing, particularly for mobile applications, by developing lightweight algorithms. Thirdly, there is a need to improve algorithm performance by collecting comprehensive datasets and employing techniques like transfer learning and domain adaptation. Fourthly, exploring diverse ensemble learning architectures and integrating both image-based and non-image-based features can enhance diagnostic accuracy.

Table 1. Dataset summary and model performance

Method	Key findings	Dataset	Performance metrics
CNN [5]	Transfer learning with data augmentation showed improved performance, particularly with AlexNet.	PAD-UFES-20	Excellent results with AlexNet
CNN comparison [6]	Dermoscopic images outperformed smartphone images in diagnostic accuracy. EfficientNet B4 was the best performer.	HAM10000	F1-score: 87%, Accuracy: 87.91%
Ensemble models [8]	Weighted ensemble of ResNet, Xception, and DenseNet achieved better balanced accuracy than individual models.	HAM10000	Balanced Accuracy: 85.8%
Transfer learning [9]	Transfer learning with PyTorch achieved accuracies between 96.40% and 99.04%.	Multiple	Accuracy: 96.40% to 99.04%
DCNN [16]	DCNN with preprocessing methods achieved high training and testing accuracies.	HAM10000	Training Accuracy: 93.16%, Testing Accuracy: 91.93%
ML methods [19]	High accuracy achieved with advanced preprocessing and feature extraction.	ISIC 2019	Accuracy: 96.25%
ML and feature extraction [22]	Combining VGG16 for feature extraction with XGBoost classification achieved the highest accuracy.	ISIC dataset	Accuracy: 99.1%
ML and image quality enhancement [20]	Improved image quality and feature extraction led to better SVM performance.	HAM10000, ISIC2019	Outperformed ISIC2019
Image processing [25]	Statistical features and GLCM data achieved high accuracy and precision.	Two datasets	Accuracy: 97.00%
Enhanced classification [26]	Novel metadata technique improved classification across evaluated models.	Two skin lesion datasets	Outperformed other methods in six out of ten scenarios
Skin lesion recognition [27]	Model recognized seven skin lesion categories with good accuracy, precision, and recall.	HAM10000	Accuracy: 90%, Precision: 0.89, Recall: 90%

2.5. Available datasets

Different learning models are trained and evaluated on a number of publicly accessible datasets containing photos of skin cancer. Some of the most well-known datasets are listed in Table 2. The development and assessment of deep learning (DL) and machine learning (ML) models for the identification and classification of skin cancer cases could be greatly aided by these datasets.

Table 2. Overview of skin cancer datasets used for model evaluation

Datasets	Description	Total images	Access
International skin imaging collaboration (ISIC) archive [19], [20], [23], [27]	The ISIC Archive includes images labeled as benign, malignant, and various skin conditions. The archive is regularly updated with new data and is used in annual challenges to promote the development of automated diagnostic tools.	50,000	ISIC Archive
Human against machine training images (HAM10000) [8], [9], [20]	This collection includes pigmented skin lesions, such as melanoma and benign nevi. The collection of data is carefully chosen and includes metadata such as age, gender, and lesion location.	10,015	Kaggle
Dermofit image library [12]–[14]	This dataset created by the University of Edinburgh, includes pictures of skin lesions that are utilized in teaching and research. It covers a variety of skin disorders, including seborrheic keratoses and melanomas.	1,300	Requires a license purchase Dermofit Image Library
PH2 Dataset [15], [21]	The PH2 dataset 40 melanomas, 80 common nevi, and 80 atypical nevi. The images are manually segmented and annotated, making this dataset valuable for training models in segmentation and classification tasks.	200	PH2 Dataset
MED-NODE dataset [14], [15]	This dataset was created for the detection of melanoma. It includes images with labels for binary classification tasks.	170 (70 melanoma, 100 nevi)	MED-NODE dataset
PAD-UFES-20 dataset [5], [6]	This dataset includes clinical images of pigmented skin lesions collected from patients at the Federal University of Espírito Santo, Brazil. It also contains metadata like age, gender, and lesion localization.	2,298 images	IEEE DataPort

3. RESULTS AND DISCUSSION

3.1. Results of different models using different datasets

In this section, we have compared different architectures used for skin cancer detection, and have observed results obtained for various parameters. From Table 3, we can see for the HAM10000 dataset, better results are obtained using SVM is 97%, While for the PAD-UFES-20 dataset, AlexNet shows a good result of 99% and for the ISIC2019 dataset MSVM shows the highest accuracy of 96.25%. A comparison of different methods for types of datasets is observed in Table 3.

Table 3. Accuracy values obtained by different methods for different datasets

HAM10000 dataset	Accuracy	PAD-UFES-20 smartphone images (SI)	Accuracy	ISIC 2019	Accuracy
ResNet Xception [8]	78.15%,	Alex-Net with TL and augmented data [5]	99%	MSVM [13]	96.25%
DenseNet [8]	81.9%	MobileNet-V2 with TL on augmented data [5]	94.071%	SVM [6]	95%,
ResNet50 [9]	90%	ResNet-50 with TL on augmented data [5]	94.918%	KNN [6]	94%,
SVM [6]	97%	CNN-TL [6]	74.85%	DT [6]	93%
KNN [6]	95%	-	-	EfficienNetB0 [11]	86.25%
DT [6]	95%	-	-	GA-Holdout [11]	90.09%
Multimodal fusion (ALBEF) [24]	94.11%			GA-cross-validation [11]	88.69%

3.2. Results of different CNN models on different parameters

From Figure 3, we observe, that different CNN architectures such as Alex-Net with TL and augmented data, MobileNet with TL on augmented data, and ResNet-50 with TL on augmented data [5], are compared for different parameters, i.e., accuracy, sensitivity, specificity, precision, and F1-score respectively [5]. The research [5] compared the results from two scenarios using three different methodologies. The best outcomes are obtained using an Alex-net with transfer learning that has been trained on improved photographs with an accuracy of 99.155%.

In Figure 4, we have compared different CNN TL architectures for different datasets *i.e.*, smartphone images and dermoscopy images. It is observed CNN TL (dermoscopy images) has more accuracy (87.80%) and sensitivity (95.50%) as compared to CNN TL (smartphone images) architecture [6], but CNN transfer learning (smartphone images) has more specificity (71.40%) and precision (94.10%) as compared to other architecture. From Figures 3 and 4 CNN architectures such as Alex-Net with TL on augmented data, MobileNet with TL on augmented data, and ResNet-50 with TL on augmented data [5], CNN TL architecture for different datasets *i.e.*, smartphone images and dermoscopy images [6], is compared for different parameters, *i.e.*, accuracy, sensitivity, specificity, precision, and F1-score respectively [5], [6]. CNN architectures-AlexNet with TL and data augmentation has the best result as compared to all the architectures.

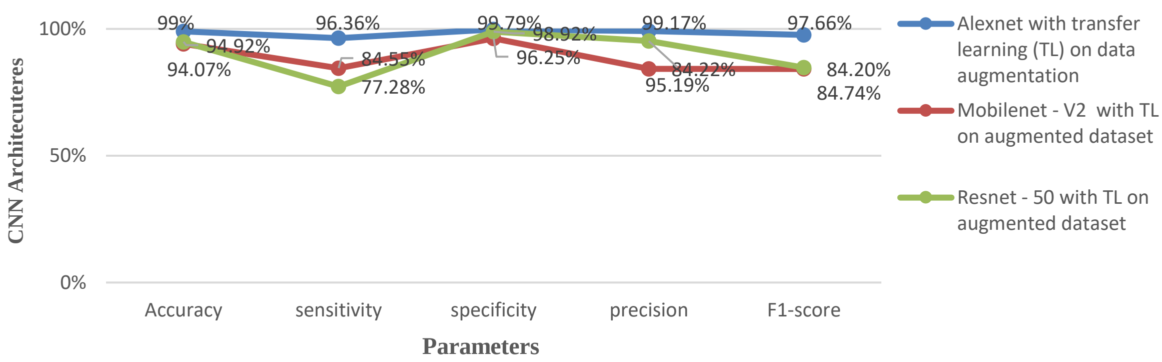

Figure 3. Comparative parameters for different architectures [5]

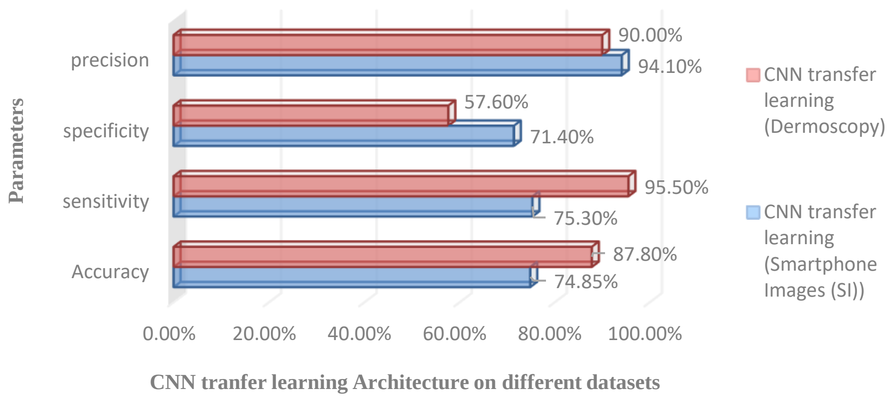

Figure 4. Comparative accuracy for different CNN transfer learning architecture [6]

3.3. Results of different ML models for different dataset

Figure 5 presents a comparison of various machine learning architectures, including SVM, KNN, and DT, applied to the ISIC 2019 and HAM10000 datasets. The results demonstrate that the SVM algorithm achieves the highest accuracy among the evaluated methods for both datasets. This superior performance highlights SVM's capability to effectively handle complex classification tasks in medical image analysis, as seen in the context of skin lesion classification. These findings, as supported by [6], emphasize the importance of selecting robust algorithms like SVM for achieving optimal results in such high-stakes applications.

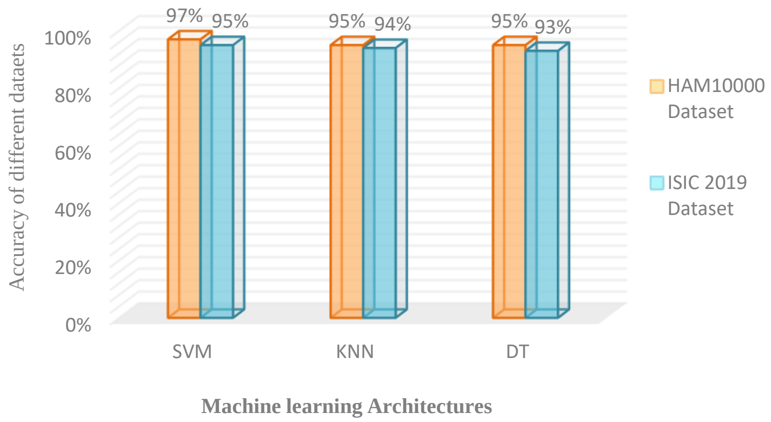

Figure 5. Comparative accuracy for different machine learning architectures for different datasets [6]

4. CONCLUSION

This study has demonstrated the effectiveness of DL and ML models, in addressing skin cancer diagnosis. By utilizing a deep learning-based approach, we were able to higher accuracy, and improved sensitivity, significantly outperforming traditional methods. Our findings highlight the potential of AI-driven diagnostic tools will help lessen the burden on healthcare workers and improve clinical decision-making. The results confirm the reliability and efficiency of our model in real-world scenarios. It is also observed different datasets works on different available models to obtain the best results, such as from our analysis we came to conclusion, for HAM1000 dataset, SVM shows highest accuracy of 97% for PAD-UFEUS dataset, Alex-Net with TL and augmented data shows 99% accuracy and for ISIC 2019 MSVM 96.25%

5. FUTURE WORK

Despite the promising results, there are several areas where further research and development are necessary. In future work, we aim to,

- Expand the dataset: Including larger, more diverse datasets to improve the model’s generalizability across different populations and conditions,
- Enhance model interpretability: Developing explainable AI (XAI) methods to make the model’s predictions more transparent and understandable to clinicians,
- Explore advanced models: Investigating newer deep learning architectures, to further enhance model performance,
- Real-world deployment: Testing the system in real-time clinical environments to ensure its practical applicability and evaluate its impact on clinical workflows,
- Multi-modal data integration: Incorporating other diagnostic data sources to improve prediction accuracy and provide more comprehensive diagnostic support.

These future research directions will help further refine our model, making it more robust and useful in diverse clinical settings.

REFERENCES

[1] A. Javaid, M. Sadiq, and F. Akram, “Skin cancer classification using image processing and machine learning,” in *Proceedings of 18th International Bhurban Conference on Applied Sciences and Technologies, IBCAST 2021*, Jan. 2021, pp. 439–444, doi: 10.1109/IBCAST51254.2021.9393198.

[2] H. Nahata and S. P. Singh, "Deep learning solutions for skin cancer detection and diagnosis," in *Learning and Analytics in Intelligent Systems*, vol. 13, Springer International Publishing, 2020, pp. 159–182.
[3] N. Hasan *et al.*, "Skin cancer: understanding the journey of transformation from conventional to advanced treatment approaches," *Molecular Cancer*, vol. 22, no. 1, Oct. 2023, doi: 10.1186/s12943-023-01854-3.
[4] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020, doi: 10.1109/ACCESS.2020.3003890.
[5] S. Medhat, H. Abdel-Galil, A. E. Aboutabl, and H. Saleh, "Skin cancer diagnosis using convolutional neural networks for smartphone images: A comparative study," *Journal of Radiation Research and Applied Sciences*, vol. 15, no. 1, pp. 262–267, Mar. 2022, doi: 10.1016/j.jrras.2022.03.008.
[6] A. Dascalu, B. N. Walker, Y. Oron, and E. O. David, "Non-melanoma skin cancer diagnosis: a comparison between dermoscopic and smartphone images by unified visual and sonification deep learning algorithms," *Journal of Cancer Research and Clinical Oncology*, vol. 148, no. 9, pp. 2497–2505, Sep. 2022, doi: 10.1007/s00432-021-03809-x.
[7] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, Dec. 2022, doi: 10.1016/j.neuri.2021.100034.
[8] Z. Rahman and A. M. Ami, "A transfer learning based approach for skin lesion classification from imbalanced data," in *Proceedings of 2020 11th International Conference on Electrical and Computer Engineering, ICECE 2020*, Dec. 2020, pp. 65–68, doi: 10.1109/ICECE51571.2020.9393155.
[9] A. Jibhakate, P. Parnerkar, S. Mondal, V. Bharambe, and S. Mantri, "Skin lesion classification using deep learning and image processing," in *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, Dec. 2020, pp. 333–340, doi: 10.1109/ICISS49785.2020.9316092.
[10] V. A. Kumar, V. Mishra, and M. Arora, "Convolutional neural networks for malignant and benign cell classification using dermatoscopic images," in *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, Feb. 2021, pp. 1040–1044, doi: 10.1109/ICICV50876.2021.9388605.
[11] E. Başaran and Y. Çelik, "Skin cancer diagnosis using CNN features with genetic algorithm and particle swarm optimization methods," *Transactions of the Institute of Measurement and Control*, Feb. 2024, doi: 10.1177/01423312241253926.
[12] S. Yang, C. Shu, H. Hu, G. Ma, and M. Yang, "Dermoscopic image classification of pigmented nevus under deep learning and the correlation with pathological features," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–11, May 2022, doi: 10.1155/2022/9726181.
[13] X. Shen, L. Wei, and S. Tang, "Dermoscopic image classification method using an ensemble of fine-tuned convolutional neural networks," *Sensors*, vol. 22, no. 11, p. 4147, May 2022, doi: 10.3390/s22114147.
[14] A. Lembhe, P. Motarwar, R. Patil, and S. Elias, "Enhancement in skin cancer detection using image super resolution and convolutional neural network," *Procedia Computer Science*, vol. 218, pp. 164–173, 2022, doi: 10.1016/j.procs.2022.12.412.
[15] H. Hussein, A. Magdy, R. F. Abdel-Kader, and K. A. El Salam, "Binary classification of skin cancer images using pre-trained networks with I-GWO," *Inteligencia Artificial*, vol. 27, no. 74, pp. 102–116, May 2024, doi: 10.4114/intartif.vol27iss74pp102-116.
[16] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, p. 100036, Sep. 2021, doi: 10.1016/j.mlwa.2021.100036.
[17] N. Aburaed, A. Panthakkan, M. Al-Saad, S. A. Amin, and W. Mansoor, "Deep convolutional neural network (DCNN) for skin cancer classification," in *ICECS 2020 - 27th IEEE International Conference on Electronics, Circuits and Systems, Proceedings*, Nov. 2020, pp. 1–4, doi: 10.1109/ICECS49266.2020.9294814.
[18] A. Adebiyi *et al.*, "Accurate skin lesion classification using multimodal learning on the HAM10000 dataset," *medRxiv*, pp. 1–17, May 2024, doi: 10.1101/2024.05.30.24308213.
[19] M. K. Monika, N. A. Vignesh, C. U. Kumari, M. N. V. S. S. Kumar, and E. Laxmi Lydia, "Skin cancer detection and classification using machine learning," *Materials Today: Proceedings*, vol. 33, pp. 4266–4270, 2020, doi: 10.1016/j.matpr.2020.07.366.
[20] M. Ahammed, M. Al Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthcare Analytics*, vol. 2, p. 100122, Nov. 2022, doi: 10.1016/j.health.2022.100122.
[21] A. Gautam and B. Raman, "Skin cancer classification from dermoscopic images using feature extraction methods," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, Nov. 2020, vol. 2020-Novem, pp. 958–963, doi: 10.1109/TENCON50793.2020.9293863.
[22] M. R. Thanka *et al.*, "A hybrid approach for melanoma classification using ensemble machine learning techniques with deep transfer learning," *Computer Methods and Programs in Biomedicine Update*, vol. 3, p. 100103, 2023, doi: 10.1016/j.cmpbup.2023.100103.
[23] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Multi-models of analyzing dermoscopy images for early detection of multi-class skin lesions based on fused features," *Processes*, vol. 11, no. 3, p. 910, Mar. 2023, doi: 10.3390/pr11030910.
[24] Z. B. Kadeem and Q. O. Mosa, "Skin melanoma diagnosis using machine learning and deep learning with optimization techniques: survey," *BIO Web of Conferences*, vol. 97, p. 52, 2024, doi: 10.1051/bioconf/20249700052.
[25] M. A. M. Almeida and I. A. X. Santos, "Classification models for skin tumor detection using texture analysis in medical images," *Journal of Imaging*, vol. 6, no. 6, p. 51, Jun. 2020, doi: 10.3390/JIMAGING6060051.
[26] A. G. C. Pacheco and R. A. Krohling, "An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3554–3563, Sep. 2021, doi: 10.1109/JBHI.2021.3062002.
[27] H. K. Kondaveeti and P. Edupuganti, "Skin cancer classification using transfer learning," in *Proceedings of IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation, ICATMRI 2020*, Dec. 2020, pp. 1–4, doi: 10.1109/ICATMRI51801.2020.9398388.

BIOGRAPHIES OF AUTHORS

Amruta Thorat holds a master's degree in VLSI and embedded systems from SPPU and currently pursuing her Ph.D. in the field of electronics and communication engineering under SPPU. Her current research interest is in the field of biomedical, deep learning. She has been working at Dr. D. Y. Patil Institute of Technology (DIT), Pimpri, Pune since 2021, presently she is working as an assistant professor in the Department of Electronics and Telecommunication Engineering. She can be contacted at: amrutathorat1206@gmail.com.

Chaya Jadhav holds a masters from govt. College of Engineering, COEP Pune and she completed her Ph.D. in electronics and communication engineering in 2019 from Vignan University Guntur, A.P. She has recognition as a Ph.D. Research guide at SPPU. Guiding 3 research scholars, she is also recognized as an approved PG Teacher at SPPU. She is currently doing research in the field of medical image processing and machine learning. Worked at Dr. D. Y. Patil Institute of Technology (DIT), Pimpri, Pune since 2001, presently she is working as a professor in the Department of Computer Engineering. Total teaching experience is more than 25 years. She can be contacted at email: chaya.jadhav@dypvp.edu.in.