

# A Comparative Analysis of Machine Learning Models and Data Sources towards Effective Skin Disease Prediction

**Kavyashree G J<sup>1</sup>, Kavyashree Nagarajaiah<sup>2</sup>**

Assistant Professor, Dept. of AI & ML, Navkis College of Engineering, Hassan, India<sup>1</sup>

Associate Professor, Dept. of MCA, SSIT Tumkur, Tumkur, India<sup>2</sup>

**Abstract:** The human body consists of several organs and skin is the largest organ that covers human body. Any disorder that affects skin is known as skin disease. The main causes of skin lesion are bacteria, virus, fungal infections and some genetic factors. The accurate diagnosis of skin conditions is crucial for effective treatment and management. Dermatologist rely heavily on visual inspection and physical examination to identify and differentiate between various skin disease. This paper is highlighting frequently used datasets and figuring out the gaps, listing out existing machine learning algorithms and their accuracy. This paper gives the clear view of many algorithms' efficiency and limitations in existing approaches.

**Keywords:** Dermatology, Skin Lesion, Fungal Infection, Diagnosis, Accuracy, Machine learning algorithms.

## I. INTRODUCTION

Skin disease is affecting all the individuals irrespective of age groups, geographical areas. Skin acts as obstruction to the harmful factors in the environment, microorganisms, and any injuries caused by the external force. It is also vulnerable to a broad range of minor disorders, from commonly occurring issues such as acne, dermatitis, scaly skin conditions to more serious and long-lasting illness like skin cancer and some severe chronic diseases.

Over past years, there has been boom in research on skin diseases. This advancement in research has resulted in understanding mechanism of skin disease and advancement in techniques to diagnosis and therapies/ treatments.

In spite of all these developments, difficulties remain in early identification, categorizing accurately and efficient care taking for long time, particularly in the areas that don't have enough resources to support proper healthcare services.

The literature survey provides overview of present status of scientific studies on skin lesion. By carefully reviewing the available literature papers, this paper intends to recognize major trends, note existing gaps, and suggest pathways for future research.

## II. RELATED WORK

Employing machine learning model for detecting skin diseases has been explore in recent investigation. Below are some examples of studies addressing this area:

[2] Rifat Sadil et al. (2023) explored an in-depth investigation of CNN architectures combined with transfer learning for automated diagnosis of skin disease. The study emphasizes systematic image preprocessing and evaluates lightweight and deep models, including MobileNet and Xception, while benchmarking against ResNet-50, InceptionV3, Inception-ResNet and DenseNet architectures. Data insufficiency is compensated through real-time augmentation and the trained models are deployed using Flask-based web application to support remote diagnosis of skin disease. The study focused on classifying skin conditions such as atopic dermatitis, eczema, herpes, nevus, and melanoma with MobileNet and Xception achieving 96% and 97% respectively. The authors highlight the need for larger, clinically validated datasets and real-world deployment to improve robustness and practical applicability.

Meenakshi K et al. [4] introduced a machine learning- driven approach for the identification of melanoma by leveraging a combination of texture-based features. The proposed framework employed Otsu-based segmentation along with morphological processing to eliminate noise and hair artifacts from dermoscopic images. A diverse set of handcrafted

features, including ABCD characteristics, GLCM, LBP, and histogram of Gradients, was utilized to capture the distinctive patterns of melanoma lesions. These extracted features were classified using multiple supervised learning models such as XGBoost, Random Forest, lightGBM, and SVM, among which XGBoost delivered the best performance with an accuracy of 95.38%. The study was limited to a single disease class and used relatively constrained datasets which may affect its applicability to broader skin disease diagnosis.

The study reported in [5] by Shagun Sharma et al, proposed a skin lesion classification approach that differentiates Benign and Malignant cases by combining deep learning-based feature extraction with conventional machine learning techniques. A pre-trained squeezeNet CNN was employed to obtain discriminative features from dermoscopic images, capturing important melanoma characteristics. These features were analysed using various classifiers, including Neural Networks, Support Vector Machines, K-Nearest Neighbours, and Naïve Bayes, where the Neural Network mode achieved the highest accuracy of 88.2%. The system's performance was assessed using standard evaluation metrics such as accuracy, precision, recall, F1-score, AUC and ROC curves. The lack of external dataset validation, interpretability methods, and clinical deployment considerations limits its practical adoption in healthcare environments.

Premkumar Duraisamy et al [6] presented a predictive framework based on ML techniques to support the early identification of skin cancer. The research employed a publicly available Kaggle dataset consisting of 10016 samples consisting of clinical and demographic information which was processed using data cleaning, imputation and dataset portioning strategies. Relevant attributes were carefully selected to improve model efficiency, and several classifiers, including Logistic Regression, Random Forest, XGBoost, KNN and CNN were evaluated for disease prediction. Ensemble learning using both hard and soft voting mechanisms was applied to aggregate the outputs of individual models and enhance overall performance. Experimental results indicated that Logistic Regression outperformed other model with an accuracy of 85.02%. the study depends on a single non-clinical dataset, limited use of advanced image-based deep learning models, and lack of real-world clinical validation restrict its broader applicability.

[7] Jing En Leong et al. proposed a deep learning-based framework utilizing convolutional neural networks for the classification of multiple skin diseases from dermoscopic images. The study combined two publicly available Kaggle datasets to form a unified dataset of 18,505 images covering six major dermatological conditions. All images were pre-processed through resizing to 256\*256 RGB format and normalization, and the dataset was divided into training and testing subset using an 80:20 ratio. The CNN model was trained for 100 epochs, achieving a peak training accuracy of 91.9% while validation accuracy stabilized between 80% and 85%. However, challenges such as class imbalance, overfitting, limited dataset diversity, lack of explainability and absence of clinical validation may limit the mode's robustness and applicability in real-world healthcare settings.

[8] Kritika Sujay Rao et al. presented a convolutional neural network-based framework for automated multiclass classification of skin diseases. The proposed approach focused on identifying seven distinct types of skin lesions, including melanoma, melanocytic nevi, basal cell carcinoma, and other commonly occurring dermoscopic conditions. The model was evaluated on HAM10000 dataset, which comprises 10010 dermatoscopic images, and achieved an overall classification accuracy of approximately 93.35%. Furthermore, limitations such as restricted dataset diversity, lack of interpretability, and absence of clinical validation hinder its direct deployment in practical healthcare applications.

[9] Nazmul Hasan Sany et al, proposed a ML framework combining image segmentation and classification for detecting skin diseases such as Eczema, Melanoma, Dermatofibroma, Psoriasis, and vascular Lesions. The study evaluated multiple classifiers, with SVM achieving 93.5% accuracy, followed by Decision Tree (76%) and KNN (72.7%) using the ISIC dataset. Despite limitations segmentation complexity, lack of Deep Learning integration and absence of clinical validation were noted.

Sai Srinivas Vellela et al [11] proposed a multi-class skin disease classification model that combines colour and texture features with a CNN for automated detection of skin lesions, including Eczema, Acne, Benign and Malignant Lesions, and healthy skin. The model achieved high performance with an accuracy of 96.7% sensitivity of 95.6% and specificity of 94.5% using clinical and dermoscopic images from multiple online and institutional datasets. Despite these results, the study faced limitations such as limited dataset diversity, disease coverage, reliance on preprocessing, lack of cross-dataset testing, absence of clinical validation and limited explainability.

[12] Arunkarthick A K et al, proposed a ML-based framework for early prediction of Myositis, focusing on Polymyositis, Dermatomyositis and Inclusion Body Myositis. Multiple algorithms were evaluated, with Gradient Boosting achieving the highest accuracy of 96%, Random Forest achieved 92% accuracy and SVM with 90% accuracy, using a dataset of 82 patient records, despite the promising performance, the study is limited by the small dataset size, data imbalance, and restricted feature set. Additionally, the absence of deep learning methods and clinical validation highlights the need for further research to improve model generalization and clinical relevance.

Sameer Dev Sharma et al [13] proposed a real-time, web based skin disease prediction system using deep learning models implemented through the Python Django framework. The system integrates CNN, ANN and SVM to detect skin conditions such as Melanoma, Psoriasis, Eczema, Hives and cold sores. Among the models evaluated CNN achieved the highest accuracy of 95%, ANN achieved 91% and SVM with 89% , using a dataset of approximately 3000 images. While the real-time framework demonstrates promising performance, issues related to dataset diversity, class imbalance, limited disease scope, absence of clinical validation, and non-availability of mobile deployment remain unresolved.

Dr. Vaudha Rani et al [16] introduced a skin disease classification framework using machine learning and data mining technique. The approach employed image preprocessing CNN-Based feature extraction, and ReLU and softmax activation functions for multi-class classification of disease such as melanoma, acne and eczema. Experimental evaluation using ISIC and DermaNet datasets showed improved classification accuracy. However, the study is limited by dataset diversity, class imbalance, sensitivity to image quality and the need for larger real-world datasets to improve robustness and generalization.

The study in [17] by Noorunisa Gaffoor et al, proposed a hybrid skin disease classification framework that combines ResNet-50 for deep feature extraction with a linear support vector machine for classification. The methodology employed image preprocessing techniques such as resizing, normalization, contrast enhancement, and noise removal using an 80:20 training and testing split. The study addressed ten dermatological conditions, including Atopic Dermatitis, Candidiasis, Eczema, Herpes, Melanoma, Psoriasis, and T-Warts. While the model achieved high accuracy of about 97% for certain diseases, the overall accuracy ranged between 75% and 85% and declined as the dataset size increased to 3000 images. Additionally, sensitivity to dataset composition, class imbalance and limited generalization highlight challenges in achieving reliable real-time clinical applicability.

The study reported in [18] proposed a skin disease detection and recommendation system that combines an ensemble of CNN models with GAN-based data augmentation and cloud computing. The framework uses computer vision techniques for image analysis and a cloud-based module to recommend treatments and preventive measures for various skin diseases, including melanoma. Experiments conducted on a GAN-augmented dataset of about 8000 images achieved 97% accuracy and 94% sensitivity, outperforming single labelled datasets and cloud services, and a lack of extensive real-time clinical validation.

[19] Ramya G et al, proposed a machine learning-based approach for skin disease detection using image preprocessing, segmentation, feature extraction, and SVM classification. The study addressed multiple dermatological conditions, including acne, melanoma, eczema, and psoriasis, but restricted the task to binary classification of normal and abnormal skin images. Experimental results showed that SVM outperformed other ML techniques in detection accuracy. However, the method is limited by its binary framework, small and less diverse datasets, absence of deep learning methods, and lack of real-time or clinical validation.

[20] Bharathi P T et al, proposed a ML-based framework for skin disease classification using image preprocessing and handcrafted feature extraction techniques. The approach employed first-order statistical features along with GLCM and GLRLM descriptors, which were classified using decision tree and random forest algorithms to identify diseases such as eczema, psoriasis, melanoma, and dermatitis. Experimental results on Kaggle dataset demonstrated high classification accuracy by a small dataset, lack of clinical validation, and absence of deep learning methods and real-time implementation analysis.

[24] Viswanatha REDDY Allugunti, proposed a CNN-based approach for skin disease classification that performs automatic feature extraction from dermoscopic images. The model incorporated image preprocessing and data augmentation and was compared with traditional classifiers such as Decision Tree, random Forest and Gradient Boosted Trees. Focusing on melanoma subtypes including lentigo maligna, superficial spreading melanoma and nodular melanoma, CNN achieved an accuracy of about 88.83%, outperforming conventional methods. The study is limited by evaluation on a single DermNet NZ dataset, restricted disease coverage, and absence of clinical validation and real-time deployment.

The study in [25] proposes a deep learning framework for skin lesion segmentation, dermoscopic feature removal, and lesion classification using the Lesion Feature Network (LFN) and Lesion Indexing Network (LIN). Fully convolutional residual networks are employed for segmentation and classification, with a lesion index calculation unit based on distance heat maps to refine classification results. A lightweight CNN is used for dermoscopic feature extraction and shows better performance than LIN for this task. Experiments on the ISIC 2017 dataset demonstrate learning methods in lesion segmentation and classification.

[26] Rashmi Patil et al, introduced a CNN-based method using a gradient descent based similarity Measure for Text Processing (SMTP) to classify melanoma stages. The approach evaluate both binary named as stage 1 and 2 and multi-class named as stage 1, 2 and 3 classification schemes. the SMTP loss function resulted in reduced training loss compared to conventional functions. This led to improved sensitivity, specificity and overall classification accuracy.

The study reported in [27] highlights the significance of colour and texture features in skin disease classification using traditional ML methods, statistical features, including entropy, variance and maximum HSV histogram values were extracted and classified using decision tree and SVM models. The approach demonstrated satisfactory classification performance. However, its reliance on visual features makes it sensitive to illumination and environmental variations.

The study in [28] addresses the limitations of traditional ML models in handling large, high dimensional dataset for skin disease detection. The proposed approach utilizes the convolutional neural network trained on an augmented dataset generated through data duplication. This strategy enhances the model's scalability and its capability to manage complex, high dimensional data. Experimental results demonstrate that the method achieves an average classification accuracy of approximately 92% across various skin disease.

### III. CONCLUSION

Recent literature demonstrates that machine learning and deep learning approaches for skin disease detection have achieved impressive accuracy rates (85%–97%) using algorithms like SVM, Random Forest, and deep architectures including CNN, ResNet, and MobileNet. Despite these advancements, significant challenges persist: most studies rely on limited, single-source datasets (ISIC, HAM10000) that lack diversity across populations and skin tones, few models undergo real-world clinical validation, and class imbalance affects prediction robustness. Additionally, the lack of explainability in complex deep learning models hinders clinical adoption, while mobile-based and accessible diagnostic systems remain underdeveloped for resource-limited settings. Future research must prioritize larger, diverse datasets, explainable AI (XAI), cross-dataset validation, and integration with tele dermatology to bridge the gap between experimental success and practical clinical deployment in dermatology.

### REFERENCES

- [1]. Mostafiz Ahammed, Md. Al Mamun, Mohammad Shorif Uddin, A machine learning approach for skin disease detection and classification using image segmentation, *Healthcare Analytics*, Volume 2, 2022, 100122, ISSN 2772-4425, <https://doi.org/10.1016/j.health.2022.100122>.
- [2]. Rifat Sadik, Anup Majumder, Al Amin Biswas, Bulbul Ahammad, Md. Mahfujur Rahman, An in-depth analysis of Convolutional Neural Network architectures with transfer learning for skin disease diagnosis, *Healthcare Analytics*, Volume 3, 2023, 100143, ISSN 2772-4425, <https://doi.org/10.1016/j.health.2023.100143>.
- [3]. K. Sagar, B. Sathwika, M. Maniteja, I. N. Reddy, C. R. Prasad and S. Yalabaka, "A Comprehensive Study on Skin Disease Classification using ML Algorithms," 2024 Asia Pacific Conference on Innovation in Technology (APCIT), MYSORE, India, 2024, pp. 1-5, doi: 10.1109/APCIT62007.2024.10673565.
- [4]. M. K, A. Adepur, V. V. T. Nagandla and S. Agarwal, "A Machine Learning Based Melanoma Skin Cancer using Hybrid Texture Features," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-5, doi: 10.1109/CONIT59222.2023.10205876.
- [5]. S. Sharma, K. Guleria, S. Kumar and S. Tiwari, "Benign and Malignant Skin Lesion Detection from Melanoma Skin Cancer Images," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-6, doi: 10.1109/ICONAT57137.2023.10080355.
- [6]. P. Duraisamy, D. J, A. S. P. F and A. J S, "Enhanced Skin Cancer Prediction with Analysis Using Machine Learning Algorithms," 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 2023, pp. 1341-1345, doi: 10.1109/ICSCNA58489.2023.10368604.
- [7]. J. E. Leong and C. P. Goh, "Identification of Skin Conditions with Convolutional Neural Networks: A Deep Learning Approach," 2024 3rd International Conference on Digital Transformation and Applications (ICDXA), Kuala Lumpur, Malaysia, 2024, pp. 205-209, doi: 10.1109/ICDXA61007.2024.10470649.
- [8]. Sun, J., Yao, K., Huang, G., Zhang, C., Leach, M., Huang, K., & Yang, X. (2023). Machine Learning Methods in Skin Disease Recognition: A Systematic Review. *Processes*, 11(4), 1003. <https://doi.org/10.3390/pr11041003>
- [9]. N. H. Sany and P. Chandra Shill, "Image Segmentation based Approach for Skin Disease Detection and Classification using Machine Learning Algorithms," 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2024, pp. 1-5, doi: 10.1109/ICICACS60521.2024.10498287.

- [10]. M. Ferdous, J. Debnath and N. R. Chakraborty, "Machine Learning Algorithms in Healthcare: A Literature Survey," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225642.
- [11]. S. S. Vellela, R. D, C. Sowjanya, K. B. S K, L. Dalavai and K. K. Kumar, "Multi-Class Skin Diseases Classification with Color and Texture Features Using Convolution Neural Network," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1682-1687, doi: 10.1109/IC3I59117.2023.10398028.
- [12]. A. A. K and S. R. K, "Prediction of Myositis Disease using Machine Learning Algorithm," 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2023, pp. 534-539, doi: 10.1109/ICIRCA57980.2023.10220885.
- [13]. S. D. Sharma, S. Sharma, A. K. Pathak and N. Mohamed, "Real-time Skin Disease Prediction System using Deep Learning Approach," 2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON), Rajpura, India, 2023, pp. 1-6, doi: 10.1109/DELCON57910.2023.10127569.
- [14]. Zhang J, Zhong F, He K, Ji M, Li S, Li C. Recent Advancements and Perspectives in the Diagnosis of Skin Diseases Using Machine Learning and Deep Learning: A Review. *Diagnostics (Basel)*. 2023 Nov 22;13(23):3506. doi: 10.3390/diagnostics13233506. PMID: 38066747; PMCID: PMC10706240.
- [15]. S. M. Mian and Sushma, "Review of Machine Learning and Artificial Intelligence in Health Care," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 202-205, doi: 10.1109/ICACITE57410.2023.10183193.
- [16]. D. V. Vasudha Rani, G. Vasavi and B. Maram, "Skin Disease Classification Using Machine Learning and Data Mining Algorithms," 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), Gunupur, Odisha, India, 2022, pp. 1-6, doi: 10.1109/iSSSC56467.2022.10051620.
- [17]. N. Gaffoor and S. Soomro, "Skin Disease Detection and Classification Using ResNet-50 and Support Vector Machine: An Effective Approach for Dermatological Diagnosis," 2023 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), Bali, Indonesia, 2023, pp. 140-145, doi: 10.1109/IoTaIS60147.2023.10346059.
- [18]. N. D. Chowdary, S. Inturu, J. Katta, C. Yashwanth, N. S. H. V. Kanaparthi and S. Voore, "Skin Disease Detection and Recommendation System using Deep Learning and Cloud Computing," 2023 8th International Conference on Communication and Electronics Systems (ICES), Coimbatore, India, 2023, pp. 1064-1068, doi: 10.1109/ICES57224.2023.10192759.
- [19]. R. G., S. G. A., K. S. and R. R., "Skin Disease Detection based on Machine Learning Techniques," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 1249-1252, doi: 10.1109/SMART55829.2022.10047634.
- [20]. B. P T, P. G K and M. S, "Skin Disease Prediction Using Machine Learning Techniques," 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, India, 2023, pp. 1-6, doi: 10.1109/ICMNWC60182.2023.10436003.
- [21]. A. Afroz, R. Zia, A. O. Garcia, M. U. Khan, U. Jilani and K. M. Ahmed, "Skin lesion classification using machine learning approach: A survey," 2022 Global Conference on Wireless and Optical Technologies (GCWOT), Malaga, Spain, 2022, pp. 1-8, doi: 10.1109/GCWOT53057.2022.9772915.
- [22]. Debelee TG. Skin Lesion Classification and Detection Using Machine Learning Techniques: A Systematic Review. *Diagnostics (Basel)*. 2023 Oct 7;13(19):3147. doi: 10.3390/diagnostics13193147. PMID: 37835889; PMCID: PMC10572538.
- [23]. Vidhyashree B, Dr Prasanna Lakshmi G S (2024) A Comprehensive Survey on Skin disease detection using deep learning. *Frontiers In Health Informatics*, 13(3), 169-189.
- [24]. Viswanatha Reddy Allugunti. A machine learning model for skin disease classification using convolution neural network. *Int J Comput Programming Database Manage* 2022;3(1):141-147. DOI: 10.33545/27076636.2022.v3.i1b.53
- [25]. Rashmi Patil, Sreepathi Bellary. Machine learning approach in melanoma cancer stage detection, *Journal of King Saud University - Computer and Information Sciences*, 2020, ISSN:1319-1578, <https://doi.org/10.1016/j.jksuci.2020.09.002>.
- [26]. Patil R, Bellary S. Transfer learning based system for melanoma type detection. *Revue d'Intelligence Artificielle*. 2021;35(2):123-130. <https://doi.org/10.18280/ria.350203>
- [27]. K. V. Swamy and B. Divya, "Skin disease classification using machine learning algorithms," in Proc. 2nd Int. Conf. Commun., Comput. Ind. 4.0 (C2I4), Bengaluru, India, Dec. 2021, pp. 1-5.
- [28]. M. Maniraju, R. Adithya, and G. Srilekha, "Recognition of type of skin disease using CNN," in Proc. 1st Int. Conf. Artif. Intell. Trends Pattern Recognit. (ICAITPR), Hyderabad, India, Mar. 2022, pp. 1-4.