International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Hybrid model for skin classification using Neural Networks and K-Nearest Neighbors

Dr. Meenakshi Garg , Ninad Agre , Rohit Mehta, Aniket waghamare
Dept. of Master of Computer Applications ,Vivekanand Education Society Institute for Technology, Mumbai, India.

Abstract. Globally common skin diseases require early detection to reduce serious complications and improve treatment results. Current diagnostic systems operate using traditional methods that frequently produce errors and show inconsistent results which emphasizes the necessity for improved diagnostic technologies. The research unveils DermAID which uses Convolutional Neural Networks (CNNs) within a multi-modal system to automate initial skin dis- ease detection. The system combines user-provided skin images with symptom questionnaires to enhance diagnostic precision· The study used 19,500 images from 23 dermatological categories including conditions like acne and melanoma which came from DermNet and processed those images with resizing, normalization, augmentation, and standardization before training robust CNN and KNN models For feature extraction and disease classification the architecture integrates convolutional layers with pooling layers and ReLU activation functions and fully connected layers. Through a focus on interpretability and generalizability this system effectively manages biases and accounts for real-world variations in dermatology datasets. The methodology shows AI potential for quicker diagnostics and cost savings in skin disease management especially where Mumbai's climate makes skin conditions more common despite quantitative details missing from the results. The research encourages AI-driven early intervention strategies which supply users with self-assessment capabilities and assist clinical operations during shortages of dermatologists. The paper examines DermAID's scalability as a global dermatological solution while addressing model transparency limitations and future improvement strategies.

Keywords: Machine Learning, convolutional neural network, K Nearest Neighbour Classifier.

1 Introduction

The skin diseases are extremely common all across the globe and are experienced by individuals of all ages. Some of the most encountered dermatological disorders are acne, eczema, psoriasis, such as fungal infections, and skin tumors. These diseases result from interactions or predispositions such as genetic makeup, ecological environment, and behaviour [1] . Skin conditions should therefore be diagnosed at an early stage in order to have a better chance of treating them especially malignant skin cancer such as melanoma. It also can help avoid some of the serious ramifications, minimize the disease's severity, and enhance the standard of living. Regular skin inspections and changes in the skin should be noted as fundamental prerequisites for early detection of the disease and the start of effective treatment.

2 Objective:

DermAID is a new concept in health care system that significantly utilizes artificial intelligence. Incorporate the use of Artificial Intelligence (AI) to transform the first screening test of skin ailments. The field of dermatology raises problems concerning the timeliness and accuracy of required services. as skin conditions are usually multifactorial and assessments are affected by the scarcity of dermatologists [2] . DermAID handling these issues since it utilizes intelligent artificial intelligence technology to diagnose skin conditions.

Algorithms to diagnose and interpret dermatological disorders, and thus open up the path to be able to deliver fast and accurate initial screenings. The system starts with the acquisition of patient data. including pictures of skin conditions and medical history related to the cases posted on the site [3] . DermAID's deep learning models, which have been trained from large data sets containing various skin disorders, show a high level of expertise in classifying patterns, textures, and abnormality of skin images. The AI algorithms are intended to distinguish between frequently occurring and unique

dermatological manifestations makes it possible for different aspects of skin diseases to be addressed comprehensively[4]. We Develop a Website, leveraging Convolutional Neural Networks (CNNs) to enable preliminary skin disease diagnosis. We Implement a multi-modal approach by combining user questionnaire based symptom data with their skin images. This will enhance the accuracy of diagnosis.
We Include a user-friendly feature within the app to empower users with the ability to conduct initial self assessments for skin conditions. This will encourage early intervention, self-care and helps to know about the disease.

3 Existing System:

In the existing system, the ML domain company has fewer opportunities to use advanced approach and techniques as it mostly work on conventional approach. Data analysis and processing. The company still relies on spreadsheets and databases process and archive data that is sometimes tedious and prone to error and there is sense of realizing an actual need for a system improvement and optimization.
This implies that dermatology is very rich and that standardization still poses a challenge. Unlike radiology, where while imaging protocols are comparatively more standardized, dermatology encompass multiple non standard procedures.
features (for instance, magnification, image coloration, skin markings). These variations can impact AI algorithms Interpretability and Trust

AI models and Artificial Intelligence have issues of transparency. It important therefore that clinicians grasp how the above model arrives at its decisions[6] . There are concerns with interpretability, and to build the trust of users is crucial
For example, the AI systems can relapse to the same bias that was found in the training data. Promoting equal treatment and dealing any sort of prejudice in relation to ethnicity, gender, or skin type is a requirement Generalizability: The smart algorithms, currently in use work very well especially when the environment is well defined Its role for clinical practice is still inconclusive[7] .
These sources of variation include: Real-life variation; patient diversity and cases are difficult and complicated.

4 Methodology

Dataset Used:
The data consists of images of 23 types of skin diseases. The total number of images are around 19,500, out of which approximately 15,500 have been split in the training set and the remaining in the test set [5].

Fig. 1:

Content

The images are in JPEG format, consisting of 3 channels, i.e. RGB. The resolutions vary from image to image, and from category to category, but overall these are not extremely high resolution imagery.
The categories include acne, melanoma, Eczema, Seborrheic Keratoses, Tinea Ringworm, Bullous disease, Poison Ivy, Psoriasis, Vascular Tumors, etc.

5 Data Preprocessing

Information preprocessing is a difficult step when preparing Information for Teaching a Convolutional Nerve-related Web (CNN) Representation. it helps better Check truth cut education sentence and keep overfitting. Here are some important Pre-methoding steps for CNNs:

1. resizing images CNN much take images of amp set sized then resizing ensures complete stimulus information has consistent dimensions (eg 224x224 or 128x128 pixels). This step is decisive for consistency across the Dataset [11].

2. normalization pixel values are much normalized to work them into amp coherent run typically [0 1] or [-1 1]. This helps the Representation converge faster by reducing the gradient explosion or vanishing Problems. # Normalize to [0 1] image = image / 255.0[12].

3. augmentation Information augmentation techniques care roll flipping cropping zooming and loose service to unnaturally Fancy the Dataset and present variance. This reduces overfitting by making the Representation more robust to unseen Information[13].

4. standardization subtracting the base and disjunctive away the stock difference (per channel) is different facility to plate the pel values ensuring that stimulus Characteristics bear amp stock rule distribution mean
= [0485 0456 0406] # green values for pre-trained Representations std = [0229 0224 0225] image = (image - mean) / std[14].

5. One-Hot Encoding (for labels) If the CNN is used for classification categorical labels need to be converted into one-hot encoded vectors to match the Representation's output.

Example for 3 classes [1 0 0] for class 1 [0 1 0] for class 2 [0 0 1] for class 3 6. make and splitting shuffling ensures that information is broadcast arbitrarily during education preventing prejudice inch the acquisition work[15].

The Dataset is also typically split into Teaching validation and Check sets to evaluate Representation Effectiveness on unseen Information. Pre-Methoding ensures the CNN Representation receives clean standardized and enriched Information which leads to better generalization and Effectiveness.

6 Model Architecture:

CNN

A Convolutional Neural Network (CNN) is a type of deep learning neural network that is well-suited for image and video analysis. CNNs use a series of convolution and pooling layers to extract features from images and videos, and then use these features to classify or detect objects or scenes.

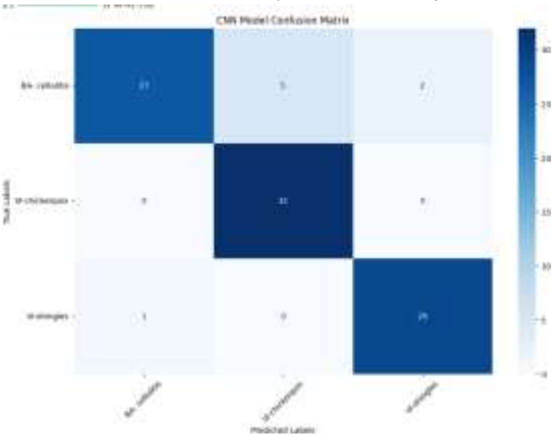

Fig. 2: CNN Model Confusion Matrix

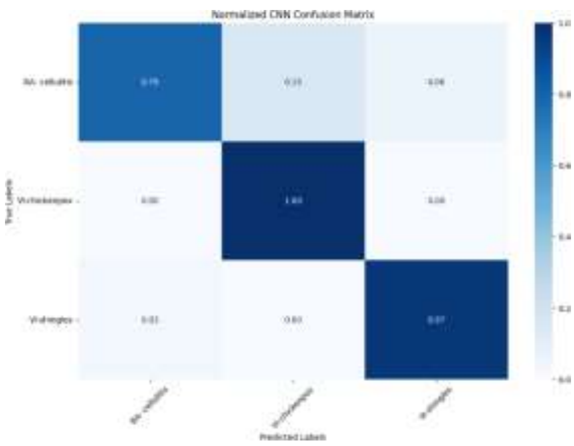

Fig. 3:Normalized CNN Confusion Matrix

How do CNN works:

CNNs work by applying a series of convolution and pooling layers to an input image or video. Convolution layers extract features from the input by sliding a small filter, or kernel, over the image or video and computing the dot product between the filter and the input. Pooling layers then down sample the output of the convolution layers to reduce the dimensionality of the data and make it more computationally efficient[7].

From ref[8]
Input Layer: Takes in the image.

Convolutional Layer: Detects small patterns (edges, textures).

Activation (ReLU) Layer: Filters out irrelevant signals.

Pooling Layer: Simplifies the image.

Fully Connected Layer: Makes sense of the detected patterns.

Output Layer: Gives the final classification.

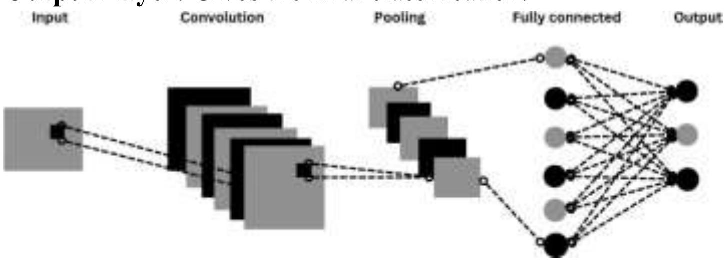

Fig. 3: CNN
KNN:

KNN is another instance-based, non-parametric learning algorithm. It does not learn like CNNs; it simply memorizes the training set and maps new instances to classes by similarity. With image classification and KNN, every image would need to be flattened to a feature vector.

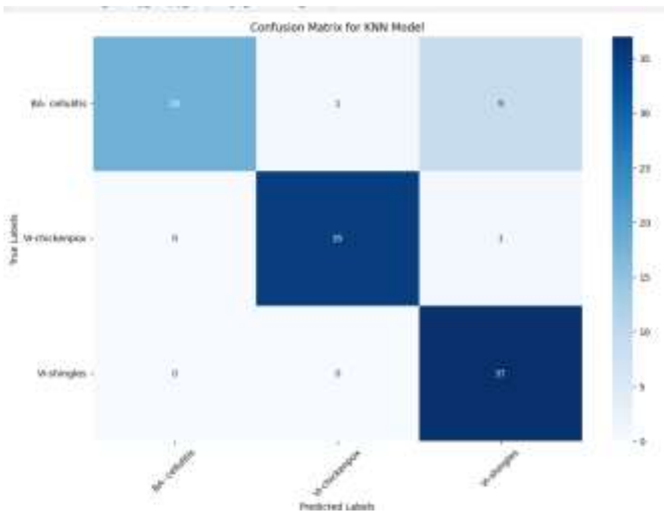

Fig. 4:Confusion Matrix for KNN model

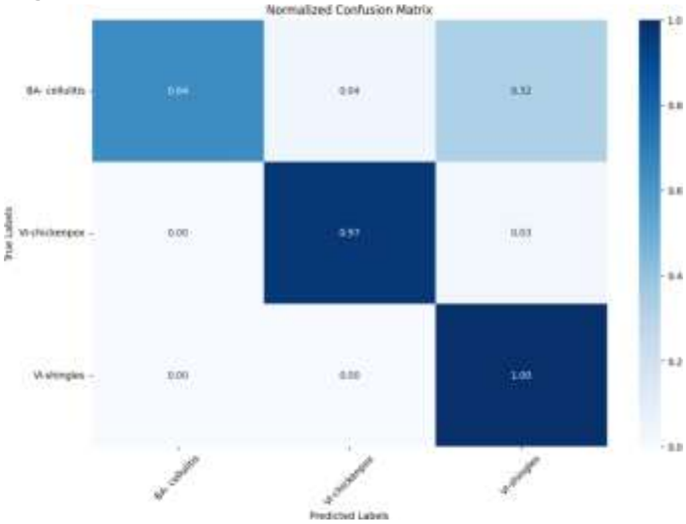

Fig. 5:Normalized Confusion Matrix

How do KNN works:

Load and Preprocess Data: Just like in CNN, load data set, train, validation, and test. But since KNN doesn't preprocess data in batches to the same extent, I may have to transform the data set to arrays[16].

2. **Flatten Images:** Flatten all images from a 2D (and 3D in the case of RGB) array into a vector of 1D. For instance, an image of 224x224x3 would be a vector of 224*224*3 = 150,528 features[17].

3. **Normalization:** Normalize the pixel values to a standard range (e.g., 0-1) so that all the features have equal weights in the distance computations[18].

4. **Use KNN**: Use a KNN classifier from toolkits like scikit-learn. Find the optimal value of k (number of neighbors) using cross-validation[19].

5. **Training**: As KNN does not have any training process like neural networks, the "training" is merely caching the dataset[20].

6. **Evaluation:** Label on test set and estimate accuracy. Additionally, check for overfitting by noting the gap between training and test accuracy[21].

7. **Visualization:** Optionally view some predictions to compare actual vs. predicted labels. Potential issues:

- **High Dimensionality**: Flat images are of very high dimensionality (150k features)

and are computationally costly and can cause the curse of dimensionality. This can result in the application of dimensionality reduction methods such as PCA.

- **Memory Usage:** It might be inconvenient to hold all the training samples in memory, particularly with 19,500 images.
- **Distance Metric**: Selecting an appropriate distance metric (Euclidean, Manhattan, etc.) and appropriately normalizing features.
- **Optimal k Selection:** Determining the ideal k that optimizes between bias and variance[22].

Results

Table 1. Table depicting accuracy across all algorithms explored.

Algorithm	Accuracy
convolutional neural network	92.8%
K Nearest Neighbour	89.00%

The CNN model reached a total accuracy of 92%, which implies that it predicted 92% of the entire cases correctly. The macro-average F1-score of 0.92 represents well-balanced performance for all classes, and the weighted-average F1-score of 0.91 represents class distribution-sensitive performance[9] and [10].

Table 2: Classification Performance of the CNN Model

Class	Precision	Recall	F1-Score	Support
BA-cellulitis	0.96	0.79	0.87	N/A
VI-chickenpox	0.86	1.00	0.93	N/A
VI-shingles	0.94	0.97	0.95	N/A
Accuracy	-	-	**0.92**	-
Macro Avg	0.92	0.92	0.92	-
Weighted Avg	0.91	0.92	0.91	-

Accuracy for the KNN model was 87%, that is, 87% of total cases were correctly classified .Macro-average F1-score is 0.86, which shows that all the classes are weighted equally. The weighted-average F1-score is 0.86, taking into account class distribution[9] and [10].

Table 3: Classification Performance of the KNN Model

Class	Precision	Recall	F1-Score	Support
BA-cellulitis	0.88	0.71	0.79	N/A
VI-chickenpox	0.91	0.95	0.93	N/A
VI-shingles	0.82	0.94	0.87	N/A
Accuracy	-	-	**87%**	-
Macro Avg	0.87	0.87	0.86	-
Weighted Avg	0.86	0.87	0.86	-

Discussion

- Interpret the results: Discuss how well the model performed and any insights gained.
- Compare your results with existing literature to show the effectiveness of your approach.
- Mention any limitations of your study and potential areas for improvement.

7. Conclusion

The findings affirm that the classification model based on CNN efficiently identifies BA-cellulitis,
VI- chickenpox, and VI-shingles with excellent accuracy (92%) and robust classification
performance. Enhanced recall and precision help in more accurate disease detection, reducing both false negatives and false positives. The results reveal the capability of deep learning models for automated skin disease diagnosis and indicate promising implications in clinical decision-making

Although the KNN model is good, it is slightly worse than the CNN model, particularly in recall for BA-cellulitis and precision for VI-shingles. This indicates that CNN is a better model for skin disease classification in this case.

Detection of skin diseases is a very important step to reduce death rates, disease transmission and the development of the skin disease. Clinical procedures to detect skin diseases are very expensive and
time-consuming. Image processing techniques help to build automated screening system for dermatology at an initial stage.

The extraction of features plays a key role in helping to classify skin diseases. In this research the method of detection was designed by using pretrained convolutional neural network. In conclusion, we must not forget that this research has an effective role in the detection of skin diseases in Mumbai because it has a very hot & humid weather for the presence of Sea; this indicates that skin diseases are widespread[8]. This research supports medical efficiency in Mumbai.

References

1. Sourav Kumar Patnaik, Mansher Singh Sidhu, Yaagyanika Gehlot, Bhairvi Sharma and P Muthu "Automated Skin Disease Identification using Deep Learning Algorithm" , Vol. 11(3), 1429-1436 (2018)
2. Ismail Oztel,* Gozde Yolcu Oztel, and Veysel Harun Sahin "Deep Learning-Based Skin Diseases Classification using Smartphones" Adv. Intell. Syst. 2023, 5, 2300211
3. Keshetti Sreekala ,N. Rajkumar ,R. Sugumar, K. V. Daya Sagar ,R. Shobarani ,K. Parthiban Krishnamoorthy , A. K. Saini,7 H. Palivela ,8 and A. Yeshitla "Skin Diseases Classification Using Hybrid AI Based Localization Approach", Volume 2022, Article ID 6138490.
4. Taye Girma Debelee "Skin Lesion Classification and Detection Using Machine Learning Techniques: A Systematic Review" Diagnostics 2023,13, 3147. https://doi.org/10.3390/diagnostics13193147
5. Jie Sun 1, Kai Yao, Guangyao Huang, Chengrui Zhang, Mark Leach, Kaizhu Huang and Xi Yang "Machine Learning Methods in Skin Disease Recognition ,A Systematic Review" Processes 2023, 11,1003. https://doi.org/10.3390/pr11041003
6. Nawal Soliman ALKolifi ALEnezi "A Method Of Skin Disease Detection Using Image Processing And Machine Learning" Procedia Computer Science 163 (2019) 85–92
7. Mostafiz Ahammed , Md. Al Mamun b, Mohammad Shorif Uddin "A machine learning approach for skin disease detection and classification using image segmentation" Healthcare Analytics 2 (2022) 100122.
8. https://www.geeksforgeeks.org/introduction-convolution-neural-network/
9. Anurag Kumar Verma · Saurabh Pal · B. B. Tiwari "Skin disease prediction using ensemble methods and a new hybrid feature selection technique" , Iran Journal of Computer Science (2020) 3:207–216
10. https://www.kaggle.com/datasets/subirbiswas19/skin-disease-dataset
11. Samir Kumar Bandyopadhyay ,Sandeep Poddar, Samir Kumar Bandyopadhyay, Payal Bose Amiya Bhaumik, Sandeep Poddar, "Machine Learning and Deep Learning Integration for Skin Diseases Prediction", Samir Kumar Bandyopadhyay et al. / IJETT, 70(2), 11-18, 2022
12. Ahmed A. Elngar ,Rishabh Kumar ,Amber Hayat ,Prathamesh Churi , "Intelligent System for Skin Disease Prediction using Machine Learning" , Journal of Physics: Conference Series 1998 (2021) 012037 , IOP Publishing

doi:10.1088/1742-6596/1998/1/012037
13. Hui Liu, Yibo Dou, KaiWang, Yunmin Zou, Gan Sen, Xiangtao Liu1 & Huling. “A skin disease classification model based on multi scale combined efficient channel attention module” , (2025) 15:6116 | https://doi.org/10.1038/s41598-025-90418-0
14. T. Shanthi , R.S. Sabeenian , R. Anand , “Automatic diagnosis of skin diseases using convolution neural network” , Microprocessors and Microsystems, July 2020
15. V. Kumar, D. Banerjee, R. Chauhan, S. Kukreti and K. S. Gill, "Optimizing Skin Disease Classification: A Comprehensive Investigation of CNN-SVM Hybrid Models with Layer Configurations," *2024 3rd International Conference for Innovation in Technology (INOCON)*, Bangalore, India, 2024, pp. 1-6, doi: 10.1109/INOCON60754.2024.10511340.
16. Dibyahash BORDOLOI1)∗ , Vijay SINGH1) , Karthikeyan KALIYAPERUMAL2), Mahyudin RITONGA3) , Malik JAWARNEH4), Thanwamas KASSANUK5) , Jose QUIÑONEZ-CHOQUECOTA6) , “Classification and Detection of Skin Disease Based on Machine Learning and Image Processing Evolutionary Models” , Copyright © 2023 D. Bordoloi et al. Published by IPPT PAN. This work is licensed under the Creative Commons Attribution License CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).
17. Nazia Hameed [a] [b], Antesar M. Shabut [c], Miltu K. Ghosh [b] [d], M.A. Hossain [e] , “Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques”, 1 March 2020, 112961.
18. Karthik R [a], Tejas Sunil Vaichole [b], Sanika Kiran Kulkarni [b], Ojaswa Yadav [c], Faiz Khan [d] , “An efficient channel attention-based convolutional neural network for skin disease classification”, Biomedical Signal Processing and Control , March 2022, 103406
19. Belal Ahmad; Mohd Usama; Chuen-Min Huang; Kai Hwang; M. Shamim Hossain; Ghulam Muhammad , “Discriminative Feature Learning for Skin Disease Classification Using Deep Convolutional Neural Network” 24 February 2020 , IEEE
20. Sarvachan Verma & Manoj Kumar , “A hybrid machine learning model for skin disease classification using discrete wavelet transform and gray level co-occurrence matrix (GLCM)”, 05 June 2024
21. Kilari Veera Swamy; Vadhi Radhika; Ragodaya Deepthi Kadiyala, “Skin Disease Classification using Image Preprocessing and Machine Learning” , 14-16 March 2024.
22. K. Meena, N. N. Krishna Veni, B. S. Deepapriya, P. A. Harsha Vardhini, B. J. D. Kalyani & L. Sharmila, “A novel method for prediction of skin disease through supervised classification techniques” 17 August 2022.
23. Saleh Naif Almuayqil,Sameh Ghany and Mohammed Elmogy , “Computer-Aided Diagnosis for Early Signs of Skin Diseases Using Multi Types Feature Fusion Based on a Hybrid Deep Learning Model” 2 December 2022
24. Belal Ahmad, Mohd Usama, Tanvir Ahmad, Shabnam Khatoon, Chaudhary Maqbool Alam, “An ensemble model of convolution and recurrent neural network for skin disease classification” 28 October 2021
25. Dr V Vasudha Rani; G. Vasavi; Balajee 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), 28 February 2023
26. Evgin Goceri , “Diagnosis of skin diseases in the era of deep learning and mobile technology” , July 2021, 104458
27. Payal Bose, Prof. Samir K. Bandyopadhyay, Prof. Amiya Bhaumik, Dr. Sandeep Poddar, “Skin Disease Detection: Machine Learning vs Deep Learning” : 13 September 2021
28. Jagandeep Singh, Jasminder Kaur Sandhu & Yogesh Kumar , “An Analysis of Detection and Diagnosis of Different Classes of Skin Diseases Using Artificial Intelligence-Based Learning Approaches with Hyper Parameters” 19 October 2023
29. Keshetti Sreekala, N. Rajkumar, R. Sugumar, K. V. Daya Sagar, R. Shobarani, K. Parthiban Krishnamoorthy, A. K. Saini, H. Palivela, A. Yeshitla , “Skin Diseases Classification Using Hybrid AI Based Localization Approach” 29 August 2022
30. Bin Zhang, Xue Zhou, Yichen Luo, Hao Zhang, Huayong Yang, Jien Ma & Liang Ma, “Opportunities and Challenges: Classification of Skin Disease Based on Deep Learning” 24 November 2021
31. Ismail Oztel, Gozde Yolcu Oztel, Veysel Harun Sahin, “Deep Learning-Based Skin Diseases Classification using Smartphones” 19 September 2023.
32. Azka Ahmed1, Hafsa Ahmad1, MohsinKhurshid2,KamranAbid1, “Classification of Skin Disease using Machine Learning” , January-March 2023