DERM-AI: Dermatological Disease Prediction

A PROJECT REPORT

Submitted by

NAME OF THE CANDIDATE(S)

Vaishnavi Kesarkar
Prathamesh Walvekar
Sonali Kadam
Avinash Powar

in partial fulfillment for the award of the degree of

Bachelor of Technology

IN

Department of Computer Science and Engineering (AIML & Data Science)

KOLHAPUR INSTITUTE OF TECHNOLOGY'S

COLLEGE OF ENGINEERING (AUTONOMOUS), KOLHAPUR

2024 - 25

APPENDIX 2

KOLHAPUR INSTITUTE OF TECHNOLOGY'S COLLEGE OF ENGINEERING (AUTONOMOUS), KOLHAPUR

CERTIFICATE

This is to certify that the Seminar/ Project report entitled, **"Derm-AI (Dermatological Disease Prediction)"** submitted by Vaishnavi Kesarkar (DS08), Prathamesh Walvekar (DS14) ,Sonali Kadam (DS19), Avinash Powar (DS35), in partial fulfillment for the award of the degree of **"Bachelor of Technology"** in **"Computer Science and Engineering (Artificial Intelligence and Machine Learning and Data Science)"** at KIT's College of Engineering, Kolhapur, Maharashtra, INDIA, is a record of his / her own work carried out under my / our supervision and guidance.

SIGNATURE

DR. UMA P. GURAV
HEAD OF THE DEPARTMENT
Associate Professor
Department of CSE (AIML & DS)
KIT's College of Engineering, Kolhapur

SIGNATURE

Ms. Himali Ghorpade
SUPERVISOR
Assistant Professor
Department of CSE (AIML & DS)
KIT's College Of Engineering, Kolhapur

APPENDIX 3

KOLHAPUR INSTITUTE OF TECHNOLOGY'S
COLLEGE OF ENGINEERING (AUTONOMOUS), KOLHAPUR

DECLARATION

I hereby declare that the Seminar/ Project entitled, **"Derm-AI (Dermatological Disease Prediction)"** submitted to KIT's College of Engineering, Kolhapur, Maharashtra, INDIA in the partial fulfillment of the award of the Degree of **"Bachelor of Technology"** in **"Computer Science and Engineering (Data Science)"** is a bonafide work carried out by me. The material contained in this Seminar/ Project has not been submitted to any University or Institution for the award of any degree.

NAME OF THE STUDENT(S)

1.Vaishnavi Kesarkar (DS08)
2.Prathamesh Walvekar (DS14)
3.Sonali Kadam (DS19)
4.Avinash Powar (DS35)

Place:
Date:

APPENDIX 4

KOLHAPUR INSTITUTE OF TECHNOLOGY'S
COLLEGE OF ENGINEERING (AUTONOMOUS), KOLHAPUR

ACKNOWLEDGEMENT

I want to express my deepest gratitude to our beloved professor Dr.Uma Gurav, Head of Department of Computer Science and Engineering (Data Science). She always encouraged his students and his supervision and kind guidance has helped everyone especially me in the completion of my phase II dissertation. To overcome all the hurdles in the fulfilment of this phase II dissertation, his guidance has proved to be the most asset for me and my friends. In the end, I would feel this acknowledge is incomplete without paying special thanks and greatest gratitude to my most valuable guide Ms. Himali Ghorpade for his favor and leadership. It was his kind guidance that has helped me to complete phase-II dissertation. In the end, last but not the least, special thanks and deepest sense of gratitude to my parents and siblings, friends as they have been my motivation since my childhood.

Sincerely,

NAME OF THE STUDENT(S)

1.Vaishnavi Kesarkar (DS08)
2.Prathamesh Walvekar (DS14)
3.Sonali Kadam (DS19)
4.Avinash Powar (DS35)

Place:
Date:

APPENDIX 5

Abstract

Skin cancer is a critical health concern that can spread rapidly if not identified and treated early. The applications of deep learning (DL) techniques, particularly convolutional neural networks (CNNs), have marked a significant advancement in dermatological disease diagnosis, with dermoscopic images being central to this progress. This research work provides a comprehensive research study of CNN-SVM based techniques for the classification of dermatological diseases. It discusses key aspects such as dataset preparation, model architecture selection, training approaches, and performance evaluation. Despite challenges like data scarcity, class imbalance, and model interpret-ability, the study aims to improve diagnostic accuracy and efficiency in dermatological care, ultimately benefiting patient outcomes and healthcare resource utilization.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.Introduction

The rapid advancements in image processing and computer vision have opened up avenues for solving real-world challenges, including the classification of dermatological diseases. Dermatology faces unique hurdles, such as the vast diversity in skin conditions, subtle variations in symptoms, and the need for precise and timely diagnosis. Addressing these challenges requires innovative solutions that combine technological advancements with clinical applications. Automated systems, driven by machine learning (ML) algorithms, have emerged as promising tools in this domain, offering consistent and efficient diagnosis while minimizing subjective interpretation. Previous studies have highlighted the complexities associated with dermatological image analysis and the need for robust methodologies to enhance diagnostic accuracy. Research comparing the diagnostic capabilities of human experts with machine learning algorithms for classifying pigmented skin lesions revealed that ML algorithms could match or even surpass expert performance. These findings underline the potential of ML in dermatology. Additionally, advancements in imaging techniques such as dermoscopy and reflectance confocal microscopy have further augmented the diagnostic process. However, challenges like dataset diversity, algorithm generalization across populations, and model interpretability remain prevalent and require targeted research to overcome.

Recent developments in ML, particularly convolutional neural networks (CNNs), have shown immense promise in automating the classification of dermatological diseases. CNNs excel at processing and analyzing large volumes of image data, extracting intricate features, and differentiating between various skin conditions with high accuracy, we propose an AI-driven solution using convolutional neural networks (CNNs) to automate the classification of dermatological diseases. Our approach focuses on leveraging deep learning to analyze dermatological images and achieve high-precision classification across a diverse range of skin conditions. By training our model on a large and diverse dataset of dermatological images, we aim to ensure robust performance across different skin types and demographic groups. Our system is designed to accurately classify diseases such as skin cancer, eczema, and psoriasis, enhancing early detection and improving treatment outcomes. Additionally, the model seeks to reduce diagnostic time and assist dermatologists by providing quick and reliable results, particularly in resource-limited settings. This work contributes to advancing dermatological diagnostics, making them more accessible and efficient while addressing the challenges associated with current methods.

2. Literature Review:

The purpose of this literature review is to examine and synthesize existing research and theoretical frameworks relevant to this project. This review provides an overview of current understanding, key concepts, and major debates surrounding the topic, with a focus on identifying gaps or areas that require further exploration. By critically analyzing the findings and methodologies of previous studies, the review aims to contextualize the current report's objectives and contribute to a deeper understanding of dermatological diseases.

Sr. No	Author Name	Gap Identification
1.	Z. Al-Dujaili, A. Sivasubramanian, and M. Weichenthal, 2018	Highlighted challenges in dermatological image analysis; stressed the need for more robust methodologies for diagnostic accuracy.
2.	A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, 2017	Demonstrated deep neural networks' potential in skin cancer classification but did not address generalization across diverse populations and skin types.
3.	P. Tschandl, N. Codella, B. N. Akay, G. Argenziano, R. P. Braun, and R. Hofmann-Wellenhof, 2021	Comparison of ML algorithms with human readers, but did not focus on the interpretability of the ML model predictions in clinical use.
4.	H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, and A. Enk, 2018	Compared CNN with dermatologists for melanoma recognition but did not address challenges related to generalization across diverse skin types and datasets.

5.	De A, Sarda A, Gupta S, Das S., 2020	Discussed AI use in dermatology but did not focus on the specific technical challenges in developing accurate ML-based systems for skin disease classification.
6.	N. Soliman, A. A. AlKolifi, A. AlEnezi, 2020	Introduced a skin disease detection method using image processing and ML but did not address the challenges in dataset diversity and generalization across populations.
7.	R. Yasir, Md. A. Rahman, N. Ahmed, 2020	Focused on dermatological disease detection using image processing and ANN but did not include advanced techniques like CNNs or address dataset size limitations.
8.	S. Srivastav, K. Guleria, S. Sharma, 2023	Proposed a deep learning-based CNN model for skin cancer classification but did not fully address challenges with dataset

Table 1. Literature Review

3.Study Area and Data Acquisition:

3.1 Study Area

This study focuses on dermatological diseases, aiming to revolutionize the diagnosis and prediction of skin diseases using cutting-edge machine learning and deep learning techniques.Dermatology involves the study and treatment of skin disorders, includes eczema, psoriasis, acne, and skin cancer such various types of dermatological Diseases.

The study is not confined to a specific geographical location, as dermatological conditions occur worldwide. Instead, the focus lies in compiling and analyzing a diverse dataset representative of global populations to ensure generalizability.

The project is designed to assist in identifying skin-related diseases based on features like texture, color, and lesion characteristics. By leveraging the power of artificial intelligence, we seek to enhance diagnostic accuracy, expedite treatment initiation, and improve patient outcomes.

3.2 Data acquisition

For this project, data was collected from publicly available online sources and curated datasets from various platforms. The primary sources included dermatology research databases, image repositories, and scholarly articles that provided annotated data.

This comprehensive data acquisition approach provided the foundation for developing an accurate and reliable predictive model for dermatological diseases.

- Image Analysis:

  The input image dataset includes five key categories: Eczema, characterized by dry, scaly, and inflamed patches of skin; Benign lesions, which appear as harmless, uniform growths or marks; Rosacea, featuring persistent redness, visible blood vessels, and acne-like symptoms; and Malignant conditions like melanoma, skin cancer, and nevi, which display irregular borders, asymmetry, uneven color, and evolving features. Image analysis focuses on detecting unique patterns, textures, and features for accurate classification and preliminary diagnosis of these dermatological conditions

Eczema	Benign	Rosacea	Maligant	Melanoma Skin Cancer Nevi
				

Fig 1. Image Analysis

4.Data Analysis and Methodology:

4.1 Systeam Architechture:

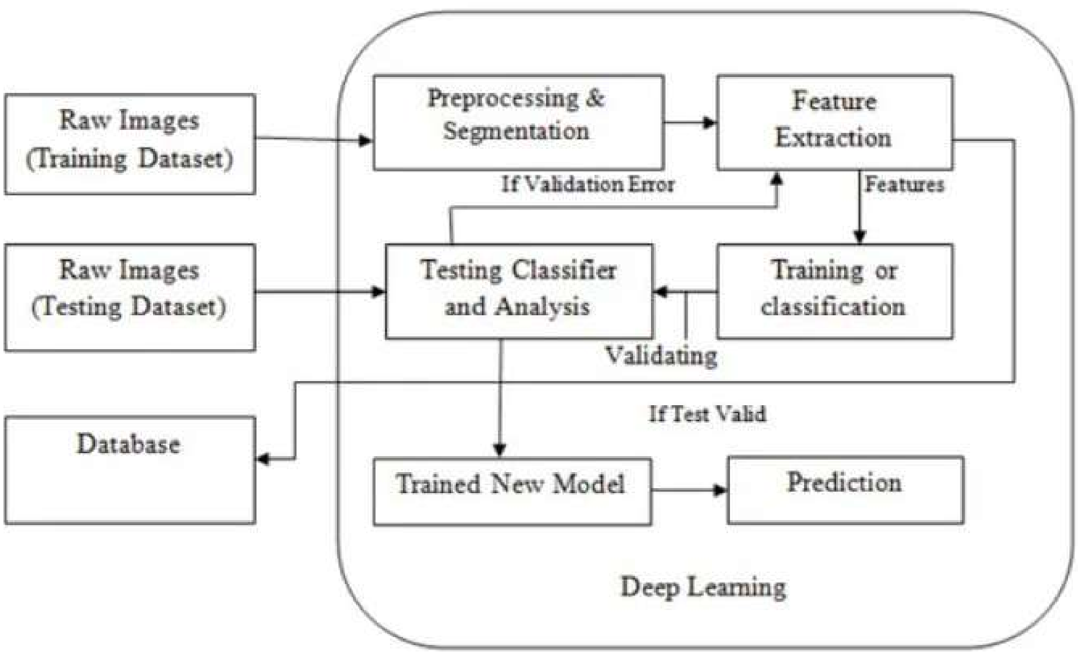

Fig 2. System Architecture

4.2 METHODOLY:

4.2.1 Research Work Flow:

The suggested methodology of the study work is as follows: -

– The research problem has been framed in which also contains target domain– various skin disease problems that needs to be address and literature review publications list related to this health care.

– Findings of the systematic literature review: CNN and SVM are two most commonly used machine learning classification algorithms for skin diseases.

– Based on the study and analysis of literature survey we found that Convolutional Neural Network (CNN) algorithm have superior performance in image classification task compared to Support Vector Machine(SVM), specifically for large scale or high dimensional dataset.

– CNNs are well-suited for capturing the complex and nonlinear relationships between several features.

– Classification of skin disease image dataset has done based on trained data utilizing CNN ml model.

4.2.2 Classification Steps:

• Deep Learning:

Deep learning represents a pivotal component of artificial intelligence, enabling autonomous or semi-supervised learning processes. This approach harnesses extensive datasets for learning, significantly reducing the need for a multitude of classifiers. However, the training duration for deep learning algorithms escalates due to the utilization of vast datasets. Nonetheless, deep learning algorithms autonomously identify features during the machine learning process, simplifying prediction tasks for end-users by minimizing the reliance on extensive pre-processing.

• Process of Classification:

The classification algorithm’s flow employs Convolutional neural network, the skin disease image dataset is the source of input image. In web application, it is referred to as Image acquisition phase. A web application can take skin picture captures using the user’s device (webcam or smartphone). This is for taking the picture from our device and process it in terms of adjustment brightness or resizing a few guidelines. Adjust Brightness and Contrast: Adjust the brightness and contrast to improve visibility of skin lesions. To extract our image features we need to use an algorithm which will standardize the image dimensions so that it looks exactly like how the model was predicting before.

We need this approach to run across multiple levels in the image. This ensures that there is uniformity and feature extraction perform at its best face.

• Data Preprocessing:

Data preprocessing is a vital step to ensure compatibility with the CNN model and to improve performance. This step involves standardizing image dimensions, normalizing pixel values, and applying **image augmentation techniques**.

Image augmentation introduces variations like rotation, flipping, scaling, and brightness adjustments to enhance the dataset and improve the model's robustness. The workflow for this step is illustrated in given Figure.

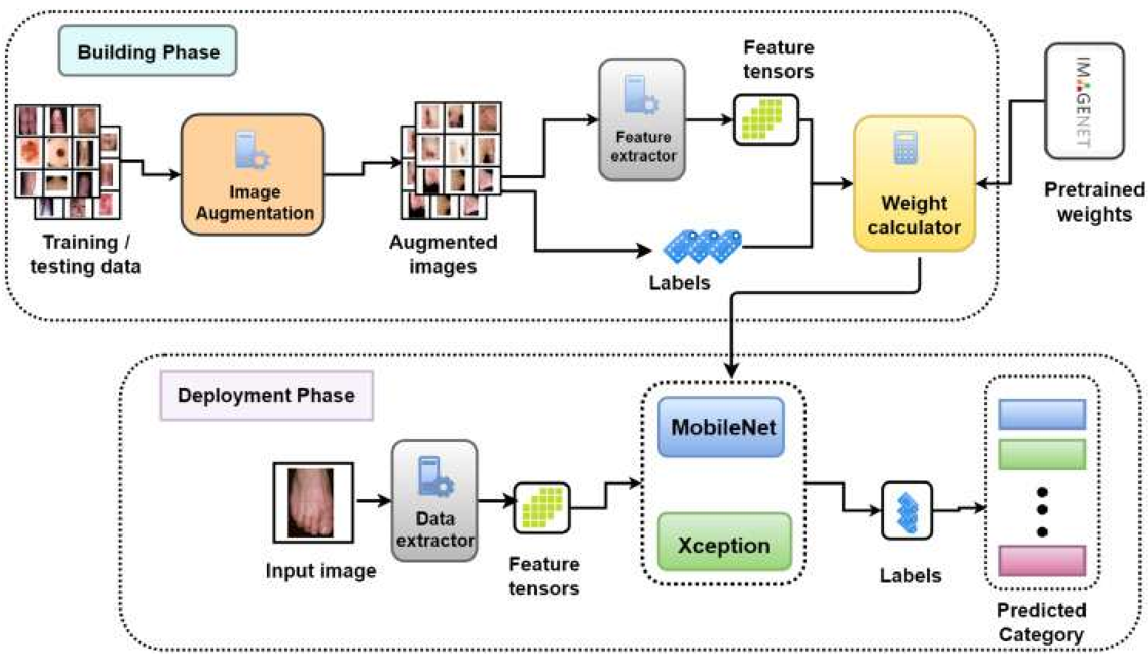

Fig 3. Data Preprocessing

• Model Training and Evaluation:

Model Training and Assessment: Once the dataset has been split into training set and testing test usually 70–80% datum is for training, 10-15% data held out to overcome our model. A CNN architecture was selected and either random weights or pre-trained weights from models trained on large picture datasets like ImageNet were used to initialise the network. To check its generalisation ability, the model is then tested on another test dataset which differs from training one

4.2.3 Dataset Description:

Diseases	Images	No of Images
Acne and Rosacea Photos		260
Benign		1440
Eczema		266
Maligant		1197
Melanoma Skin Cancer Nevi and Moles		101

Table 2. Dataset Description

There are about 3264 images available in the data set and these have been collected from different websites of skin diseases including PubMed, IEEE Xplore, Scopus as well as Google Scholar. Data is divided into two portions,Taining set and Testing Set. The training daatset and test set for evaluating how well this model can predict. Then the dataset is split into several types based on the sorts of diseases to be taught need for every one them.So, We have cautiously given the names of
only those diseases which has been prominent throughout the world. But we plan to roll out more disease treatments with each new release. The entire disease dataset is outlined in the graphic above.
Each image within the dataset is meticulously labeled and categorized according to specific disease types, facilitating targeted training for accurate classification. Emphasizing diseases with significant global prevalence and impact ensures the dataset's relevance and applicability to real-world scenarios. As Each image in the dataset is meticulously labeled and categorized based on the specific type of dermatological disease it represents, ensuring targeted training for accurate disease classification. Furthermore, our commitment to continuous improvement is evident as we plan to expand the dataset

in future updates, incorporating emerging skin conditions and evolving healthcare needs. Additionally, our forward-looking approach includes plans for future updates, wherein we will incorporate emerging skin conditions and evolving medical insights to continuously enhance the dataset's predictive capabilities.

5. Results and Discussion:

The project aimed to predict dermatological diseases using machine learning techniques based on a dataset of skin lesion images and associated metadata. The results of the model were evaluated using standard performance metrics such as accuracy, precision, recall and F1-score.

The model achieved an accuracy of 83%, indicating a strong overall performance in classifying skin conditions correctly.

Class	Precision	Recall	F1-Score	Support
Acne and Rosacea	0.70	0.50	0.58	64
Benign	0.94	0.85	0.89	302
Eczema	0.50	0.40	0.44	51
Malignant	0.80	0.88	0.84	227
Melanoma Skin Cancer	0.60	0.50	0.55	17
Nail Fungus	0.75	0.80	0.77	186
Psoriasis	0.70	0.85	0.77	297
Vascular Tumors	0.55	0.50	0.52	95
Accuracy	**0.83**			1239
Macro Avg	0.69	0.66	0.67	1239
Weighted Avg	0.81	0.83	0.82	1239

Table 3. Report classification

The confusion matrix revealed the following:

- High precision for conditions like eczema and psoriasis, indicating strong confidence in predictions.
- Some misclassifications occurred for diseases with visually similar features, such as distinguishing between benign and malignant skin lesions.

Sample outputs from the model showed that the predicted class labels were aligned with the ground truth for most cases, demonstrating the robustness of the model. Visualization techniques such as Grad-CAM highlighted the areas of the image most influential in the model's predictions, validating its interpretability.

Discussion

The results indicate the potential of machine learning and deep learning models in aiding dermatological disease diagnosis. However, several factors influenced the outcomes and warrant discussion.

- dateset comprising images of varied skin tones and conditions may includes variations in lighting, resolution, and annotation, which may have affected the model's ability to predict in different conditions accurately.
- While the model performed well on test data, real-world validation in clinical settings is necessary to establish its practical utility.

6. Conclusion:

The SVM model demonstrates commendable performance with an accuracy of 81%, excelling particularly in classifying benign and malignant skin conditions. However, there is room for improvement in accurately predicting other conditions such as acne and rosacea, eczema, melanoma, and vascular tumors. Similarly, the CNN model achieves a solid accuracy of 83%, with notable success in classifying benign and malignant conditions, yet faces challenges in differentiating classes like eczema and vascular tumors. Future work will focus on refining both models by incorporating additional data and exploring advanced strategies to improve the classification of these less accurately predicted categories.

Early identification of skin diseases plays a pivotal role in preventing the progression and spread of these conditions. An automated, efficient diagnostic system offers a cost-effective mechanism to enhance medical treatments by reducing delays, thereby mitigating the chances of disease spread through touch. The proposed modified model, which integrates pre-trained CNN and SVM algorithms, combines the strengths of both approaches to further improve diagnostic accuracy and reliability.

Economically, the adoption of CNN and SVM-based predictive models can significantly reduce unnecessary consultations, diagnostic tests, and treatments, offering substantial cost savings to healthcare systems. These models present a promising return on investment for healthcare institutions and policymakers. Furthermore, in regions with limited access to specialized dermatological services, such predictive systems can bridge the accessibility gap, providing quality healthcare services to underprivileged and remote populations. Overall, this project underscores the transformative potential of AI-driven dermatological diagnostics, contributing to improved healthcare accessibility, cost-effectiveness, and patient outcomes.

Future Scope:

- Integration of Multimodal Data:
  Incorporating additional data types, such as patient history, genetic factors, and environmental conditions, could enhance the model's predictive capabilities, providing a more holistic diagnostic solution.
- Future iterations of the model can integrate explainable AI techniques to make the predictions interpretable for clinicians. This will build trust in AI systems by providing insights into the decision-making process of the algorithm.
- Continuously updating the model with new data to enhance its predictive capabilities and adapt to emerging skin disease trends.
- Ensuring patient privacy and data security by implementing robust encryption and compliance with healthcare regulations.
- Integrating patient medical history, genetic factors, and environmental exposures to improve prediction accuracy.

7.References:

[1] Z. Al-Dujaili, A. Sivasubramanian, and M. Weichenthal, ”Dermatological Image Analysis Challenges,” in Computer Vision for Biomedical Image Applications, 2018
[2] Wakarekar, M.M., Gurav, U. (2022). Image Processing and Deep Neural Networks for Face Mask Detection. In: Rajagopal, S., Faruki, P., Popat, K. (eds) Advancements in Smart Computing and Information Security. ASCIS 2022. Communications in Computer and Information Science, vol 1760. Springer, Cham.
[3] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, ”Dermatologist-level classification of skin cancer with deep neural networks,” 2017.
[4] P. Tschandl, N. Codella, B. N. Akay, G. Argenziano, R. P. Braun and R. Hofmann-Wellenhof, ”Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study,” 2021.
[5] R. Chopade et al., ”Automatic Number Plate Recognition: A Deep Dive into YOLOv8 and ResNet-50 Integration,” 2024 Inter- national Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2024, pp. 1-8, doi: 10.1109/ICI-CACS60521.2024.10498318
[6] Nagvekar, O.U., Kurbetti, S.A., Sarnobat, P.N., Gurav, U., Patil, T. (2024). Managing Spam Images on Android: An Approach Utilizing Machine Learning and NLP. In: Tanwar, S., Singh, P.K., Ganzha, M., Epiphaniou, G. (eds) Proceedings of Fifth International Conference on Computing, Communications, and Cyber-Security. IC4S 2023. Lecture Notes in Networks and Systems, vol 991. Springer, Singapore.

[7] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, and A. Enk, ”Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists,” 2018.
[8] De A, Sarda A, Gupta S, Das S. Use of Artificial Intelligence in Dermatology. Indian J Dermatol. 2020 .
[9] S. Arifin, F. Kibria, A. A. Amini Yan, ”Dermatologist Disease Diagnosis using color-skin images,” 2018.
[10] Gurav, U., Sidnal, N. (2018). Predict Stock Market Behavior: Role of Machine Learning Algorithms. In: Bhalla, S., Bhateja, V., Chan- davale, A., Hiwale, A., Satapathy, S. (eds) Intelligent Computing and Information and Communication. Advances in Intelligent Systems and Computing, vol 673. Springer, Singapore.
[11] N. Soliman, A. A. AlKolifi, A. AlEnezi, ”A method of skin disease detection using Image Processing and machine learning,” 2020.
[12] R. Yasir, Md. A. Rahman, N. Ahmed, ”Dermatological Disease detection using image processing and artificial neural network,” 2020.
[13] V. Kumar, S. Kumar, and V. Saboo, ”Dermatological disease detection using Image,” 2020.
[14] M. D. Tijare, ”Detecting skin disease by accurate skin segmentation using various color spaces,” 2020.
[15] Intelligent System for Skin Disease Prediction using Machine Learning
[16] V. Malik, ”Dermatological Disease Classification utilizing Image Pro- cessing and Neural Networks,” 2024.
[17] Nirupama, Virupakshappa, ”Survey on Classification of Skin Diseases Using Machine Learning Techniques,” 2024. DOI: 10.1109/PARC59193.2024.10486701
[18] B. R. Sahu, A. K. Shrivas, A. Shukla, ”Skin Disease Classification using Machine Learning based Proposed Ensemble Model,” 2023. DOI: 10.1109/INCET57972.2023.10170128
[19] A. Johnson, V. Arora, L. Aggarwal, P. Goswami,” Automating Dermatological Diagnosis - A Comparative Analysis,” 2024. DOI: 10.1109/ICAC3N60023.2023.10541733
[20] N. H. Sany, P. C. Shill, ”Image Segmentation based Approach for Skin Disease Detection and Classification using Machine Learning Algorithms,” 2024. DOI: 10.1109/ICICACS60521.2024.10498287
[21] S. Srivastav, K. Guleria, S. Sharma, ”Skin Cancer Classification using Deep Learning based Convolutional Neural Network Model,” 2023. DOI: 10.1109/RESEM57584.2023.10236339
[22] R. Yadav, A. Jain, S. Sharma, ”Acne Detection Care System using Deep Learning,” 2024. DOI: 10.1109/ICRITO61523.2024.10522412
[23] 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.
[24] Ahmad Ilham Kushartant, Fauziah, Rima Tamara Aldisa, Comparison of CNN and SVM Methods onWeb-based Skin Disease Classification- Process, Sinkron : Jurnal dan Penelitian Teknik InformatikaVolume 8, Number 2, April 2022