International Journal of Advanced Research in Engineering Technology & Science
ISSN: 2349-2819
Impact Factor: 7.10
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Email: editor@ijarets.org Volume-12, Issue-8 August – 2025 www.ijarets.org

AI based Skin disease identification: A Deep learning significantly to improves diagnostic accuracy

Dr Girish Padhan
Computer Science and Engineering,Biju pattnaik University of Technology

ABSTRACT:

Machine learning algorithms were used along with image processing techniques for the detection of skin diseases. Compared to machine learning, the dermatological technique for identifying the type of skin illness is quite expensive. Skin conditions can be surprising and difficult to diagnose. The colour of healthy skin differs from that of diseased skin.Aim of the project is to classify the type of disease using machine learning. In order to decide, these algorithms employ feature values from effected images as input. Three parts make up the process of determining the type of skin disease: feature extraction, training, and testing.The objective is to increase the accuracy of diagnosis for various types of skin diseases. Here we used herpes,keratosis, eczema, and acne as the four diseases in this study. The process makes use of machine learning technology to train itself with the various skin images. Three important features in image classification are texture, color, shape, and combination of these. In this project work, features such as entropy, variance, contrast, and energy are used to build machine learning algorithm such as logistic regression, Support Vector Machine, and Decision Tree. Accuracy is used to test the performance of the chosen algorithms.Skin diseases are common health problems around the world. The perils of the infections are invisible, which cause physical health distress as well as initiate mental depression. In addition, it sometimes leads to skin cancer in severe cases. Subsequently, diagnosing skin diseases from clinical images is one of the foremost challenging tasks in medical image analysis. Moreover, when performed manually by medical experts, diagnosing skin diseases is time-intensive and subjective. As a result, both patients and dermatologists require automatic skin disease prediction, which makes the treatments plan faster.

KEYWORDS: CNN,DT,LDA,K-means,RGB,HSV,XAI

INTRODUCTION:

Skin diseases are among the most common health conditions affecting millions of individuals worldwide, with a wide range of types, from benign rashes and infections to more serious conditions such as melanoma. Early detection and timely diagnosis of skin disorders are critical to ensuring effective treatment, minimizing the risk of complications, and improving patient outcomes. However, in many parts of the world, access to dermatological expertise is limited, leading to delayed diagnoses and treatment. By K.V.Swamy The integration of technology into healthcare, particularly in the field of dermatology, has opened up new avenues to address these challenges through automated systems capable of identifying and classifying various skin conditions.This project report presents the design, development, and implementation of a Skin Disease Detection System, an intelligent software tool that leverages advancements in machine learning and computer vision to aid in the accurate detection of skin diseases from images. The core objective of the system is to bridge the gap between the availability of dermatological services and the increasing demand for efficient, accurate, and accessible diagnostic support tools. By utilizing image processing techniques and powerful deep learning algorithms, the system is capable of analyzing images of skin lesions and predicting the presence and type of skin disease with a high degree of accuracy.The proposed system employs a Convolutional Neural

Network (CNN) model, trained on a diverse data set comprising thousands of skin images labeled with various dermatological conditions. These include common disorders such as eczema, psoriasis, acne, as well as more critical conditions like basal cell carcinoma and melanoma. Through feature extraction, pattern recognition, and classification, the model learns to distinguish between different disease types based on visual cues such as texture, color, shape, and border irregularities.The significance of this project lies not only in its potential to support clinical decision-making but also in its ability to provide diagnostic assistance in remote or under resourced regions where specialist care may not be readily available. Furthermore, the integration of this system into mobile or web-based platforms could enable broader accessibility and usability by both healthcare providers and the general public encouraging proactive skin health monitoring.This report details the various stages involved in the development of the Skin Disease Detection System, including problem analysis, data set selection and preprocessing,model architecture, training and validation procedures, performance evaluation, and the user interface design. It also discusses the challenges encountered during the project, such as class imbalance, image quality variations, and the ethical considerations related to medical data usage.

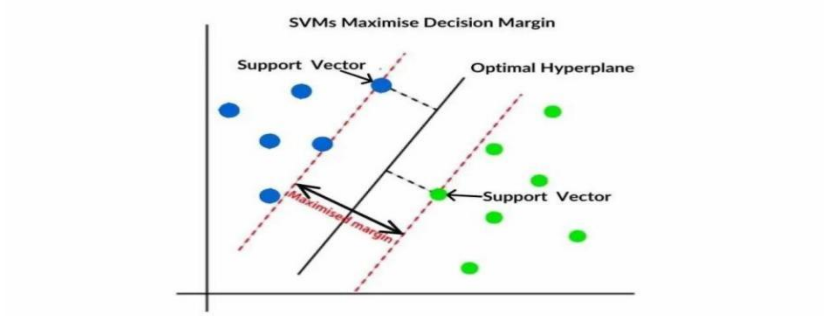
MACHINE LEARNING ALGORITHMS

a. Linear regression :

One of the most widely used and fundamental machine learning techniques is linear regression. A predictive analysis is conducted using this statistical technique.By V. B. Kumar For continuous, actual, numeric variables like salary, age, product cost, etc., predictions are made by linear regression. The dependent and independent variables are shown to have a linear relationship via the linear regression technique. The link is predicted by linear regression as a slope or straight line. It is a prediction algorithm to predict or estimate one variable using another. The Prediction of Ratings using Opinions is an example of linear regression. In this case ratings are the dependent variable, whereas Opinions are the independent variable.

b.Logistic regression: It is a type of supervised learning algorithm. By combining a number of known independent variables, it is utilized to forecast the definite dependent variable. When the dependent variable is categorical, the output is foreseen using logistic regression.

c.Support Vector Machine(SVM):



It has data points called support vectors closest to the plane in the SVM algorithm. Margin can define the SVM either the model is good or bad. The margin is the gap between the data vectors and the plane. A good margin is one that is vast, whereas a bad margin is one that is little.

<u>Application of SVM:</u>

●E-mail classification.

●Face detection – It classifies Faced structure and unface structure then draw a square boundary line to the structure.

●Handwriting recognization – Using SVM classification of character in data.

d. Decision Tree (DT):

It is a supervised learning algorithm with a structure like a tree. It has roots and nodes. Nodes of the decision tree have various types like Decision nodes and leaf nodes.Decision trees are often designed to resemble what a person thinks when making a decision, which makes them easier to understand.

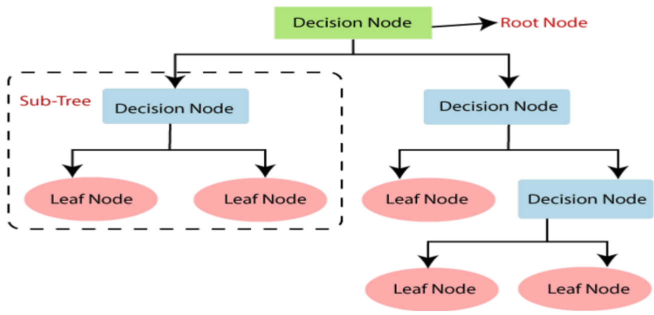
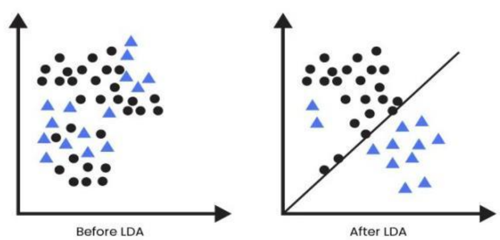


Fig. - 2.4: Decision tree

e.Linear Discriminant Analysis (LDA):

Linear Discriminant Analysis (LDA) is one of the common methods used in machine learning to solve problems that consists of more than two classes.By P. R. Hegde This is an algorithm of supervised machine learning. Its form is a linear discriminant analysis used to separate the dataset into classes by finding linear combinations. Basically, it is a dimensionality reduction approach.



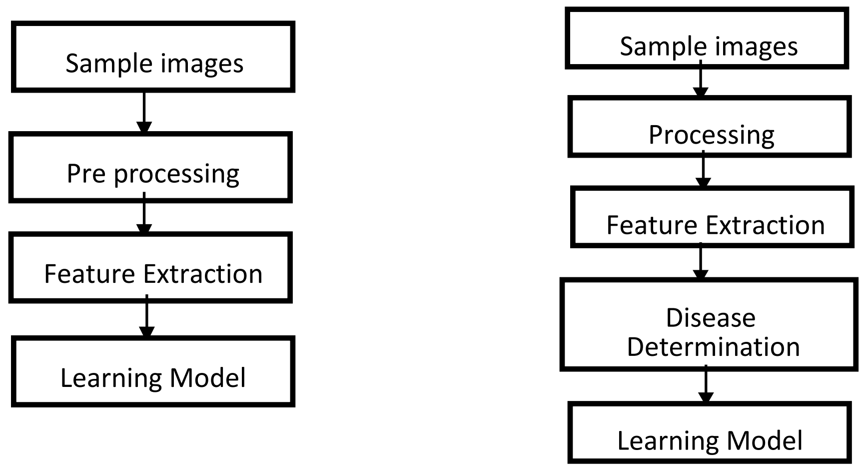
Given that certain high-dimensional data may contain noise and redundant data, we limit the number of variables in this dimensionality reduction to convert high-dimensional data into low-dimensional data.

f. K-Means Clustering Algorithm:-

In machine learning or data science, clustering problems uses unsupervised learning method. Clustering is an unsupervised learning method which is used to solve problems in data science or machine learning. K-means clustering is used to solve this kind of problems. Uses an iterative process to split unsigned data into k unique groups, each a group in common with the others. No training needed, this is a simple way to independently identify groups in anonymous data and allows us to aggregate data into different groups. Clustering.

Methodology

The methodology of Classification of skin diseases consists of two phases namely the Training phase and the Testing phase. Training Phase is a phase where several images of disease classes are obtained, from which image features obtained are stored. The extracted features of different disease classes are used for creating the required learning model using different machine learning algorithms for classifying the skin diseases. In the Testing Phase, few images of disease classes are obtained, from which image features are obtained and are compared with respect to a determined model created using the Training phase. Based on the model given image input is classified accordingly.



Training Phase Testing Phase

FEATURE EXTRACTION:-

A good feature set contains discrete information that can distinguish one item from other items. It should be as strong as possible against generating different feature values for objects of same class. The selection process should be a small enough that it is useful to distinguish between different models, but similar models in same class. Features can be of four types: Texture features, Color features, Geometric features, and Statistical features.
Here, we have used two features for the feature extraction process and they are:

1. Texture feature
2. Color feature

1. Texture Features: The texture feature is a useful characterization for an image, where pixel properties are used to measure the color in the image while the group of pixels is used to measure a texture.

Fig – 3.1: Distribution of pixels

2. Color features: Since color is an important feature that people perceive when visualizing, it is calculated for feature extraction. Color is the best feature of all visual systems and is widely used in image acquisition systems. Color is usually defined in the 3D color space. They can be Hue, Saturation, and Brightness (HSB),

Hue, Saturation, and Value (HSV), or Red, Green, and Blue (RGB) (HSB). Hue, Saturation, and Value (HSV) and Hue, Saturation, and Brightness (HSB), the final two, are based on colour, saturation, and brightness as seen by humans. In our work, we have used four features they are energy, variance, entropy, and contrast.

3.Feature Selection: Feature extraction and feature selection are the two methods available for obtaining a subset of features. In contrast to feature extraction, which extracts different features, feature selection selects a subset of the initial feature collection. By removing features with poor or unexpected properties, feature selection aims to pick a subset of input variables while maintaining or increasing classification accuracy.

1.It is necessary to find all possible feature subsets that can be formed from initial values which leads to time consumption,

2.Every feature is meaningful for at least some discriminations, and

3.Variations within the interclass and intraclass are not that much high. Beyond a certain level, the addition of additional features results in a worse performance rather than a better performance.

4. Feature extraction: After selecting the required features for our work, the next stage for the classification is feature extraction in which image texture and color features of the skin images are extracted. Feature extraction is used mainly to increase the accuracy of disease-type detection. Three important features in image classification are texture, shape, color, and also a combination of these features. Here, texture and color features are considered for classifying skin diseases.Feature extraction refers to the process of transforming raw data into a mathematical process while preserving the information in the original database. It gives better results than applying machine learning directly to raw data.

1.Expand Data set Diversity -Include more diverse skin tones, age groups, and rare diseases.Use datasets from different geographical regions to improve model generalization.

2. Real-Time Detection via Mobile/Web App -Deploy the model in a mobile app or web interface. Real-time image capture + instant feedback for users.

3. Multi-Disease Classification -Expand beyond binary classification (e.g.,healthy vs diseased).Include classification for multiple skin conditions like eczema, psoriasis, vitiligo, etc.

4. Explainable AI (XAI) Integration -Add tools like Grad-CAM or LIME to explain predictions.Helps build trust with users and professionals.

5. Dermatologist Collaboration -Collaborate with medical professionals to validate model predictions.Use expert feedback to refine the model and reduce false positives/negatives.

6. Improve Model Accuracy -Use advanced architectures (e.g., EfficientNet, Vision Transformers).Perform hyperparameter tuning, ensemble modeling, or transfer learning.

7. Image Preprocessing Enhancements -Explore better preprocessing techniques like hair removal, illumination normalization, or color constancy.

8. Cross-Platform Compatibility -Ensure your model runs well across different hardware (e.g., Android, iOS, Raspberry Pi).

9. Longitudinal Monitoring -Implement functionality to track a skin condition over time.Could help in monitoring treatment effectiveness.

10. Integration with Medicine-Allow users to send model predictions directly to a dermatologist.Create a bridge between automated diagnosis and real-world medical consultation.

RESULTS AND DISCUSSION:

Skin disease images Dataset:

The proposed algorithms were trained and tested from the features of images which were obtained from Kaggle website. This anonymous dataset was collected from patients suffering from different skin diseases and categorized the images accordingly. These images are obtained from the public access "Dermnet", one of the biggest online dermatology resource designed for providing online medical education. The image is in JPEG format and has three channels RGB. Resolution from image to image and category to category varies but most of the images do not have very high-quality images.

Fig 5.1. Acne

Fig 5.2. Herpes

Fig 5.3. Eczema

Fig 5.4. Keratosis

The obtained dataset has many disease image collections like Acne, Exanthems, Lupus, Melanoma and such 23 disease sets with total of 19,500 images around, from which we have selected the diseases that occur more often like Acne and Herpes. After downloading the dataset, we have filtered the images which have images with disease shown clearly so that it gives better feature values during extraction. Table 5.1 indicates the count of images considered for different disease classes.

Count of images considered for each disease set

Disease	Count
Acne	290
Eczema	312
Herpes	297
Keratosis	314

Normal Skin	87

By using different machine learning algorithms highest efficiency obtained was 72.3% by Support Vector Machine (SVM) algorithm. Confusion matrices of these respective algorithms are below.

Table 5.2: Confusion matrix for Logistic Regression algorithm

	Class:1	Class:2	Class:3	Class:4	Class:5	Classification overall
Class:1	33	0	0	10	1	44
Class:2	0	35	12	0	1	48
Class:3	0	18	19	0	1	38
Class:4	10	0	0	37	1	48
Class:5	0	1	0	0	16	17
Truth overall	43	54	31	47	20	195

In table 5.2 indicates the Logistic Regression using sigmoid function where Comparable to

A two-layer supervised learning neural network design, this function acts as an activation function for artificial neurons, diagonal elements of obtained matric indicates correct predictions which sum up to give 140 i.e., total 140 predictions were correct out of 195 and giving the total accuracy as 71.79%.

Table 5.3: Confusion matrix for DT algorithm

	Class:1	Class:2	Class:3	Class:4	Class:5	Classification overall
Class:1	28	0	0	16	2	46
Class:2	0	21	19	0	0	40
Class:3	1	12	32	0	1	46
Class:4	20	0	0	34	0	54
Class:5	2	0	0	0	7	9
Truth overall	51	33	51	50	10	195

Table 5.3 indicates the Decision Tree using Gini Criterion where a function that evaluates

CONCLUSIONS:

Machine learning algorithms combined with image processing techniques were utilized to detect and classify skin diseases from digital images. The work specifically focused on four common skin conditions: herpes, keratosis, eczema, and acne. Accurate and timely detection of these skin diseases is vital for effective treatment and improving patient outcomes. Traditional methods often rely on manual diagnosis, which can be time-consuming and subjective. Therefore, this project explored the potential of automated skin disease classification systems using computational techniques.

To achieve classification, images of affected skin were processed to extract significant features that characterize the texture and visual properties of the skin lesions. The choice of features plays a pivotal role in the effectiveness of classification, as skin conditions often present with overlapping visual characteristics. Color, though an obvious feature, is not always a reliable parameter for skin disease detection due to the similarities across different conditions. Therefore, the focus was placed on textural features, which better capture the underlying structure and pattern of skin abnormalities.

The primary features used in this study were entropy, variance, contrast, and energy. These features represent the statistical properties of the images and provide a quantitative basis for differentiation among the skin diseases. Once the feature extraction phase was completed, three machine learning algorithms were applied for classification: Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT).
To evaluate the effectiveness of these algorithms, accuracy was employed as the primary performance metric. The results revealed several insights into the behavior of each algorithm under different conditions. The Support Vector Machine algorithm demonstrated superior performance when classifying between two or three diseases. This result aligns with the known strength of SVM in handling binary or limited-class problems, especially in high-dimensional feature spaces. However, as the number of disease categories increased to four, the accuracy of the SVM began to decline. This decrease in performance may be attributed to increased complexity in distinguishing between multiple classes with overlapping feature distributions.
Interestingly, Logistic Regression yielded better results than SVM when dealing with just two disease classes. This finding suggests that for simpler classification tasks, Logistic Regression can offer a balance between simplicity and performance. On the other hand, Decision Tree classifiers showed moderate performance across all scenarios, but they lacked the consistency and accuracy demonstrated by the other two methods. Despite their interpretability, Decision Trees are prone to overfitting, especially with small datasets and subtle feature variations, which might have contributed to their lower performance.
Overall, the findings suggest that no single algorithm universally outperforms others in all situations. The performance of a model largely depends on the number of classes, the quality of features, and the distribution of the dataset. Moreover, the relatively small size of the dataset used in this study likely constrained the full potential of these algorithms. With a larger and more diverse dataset, along with the inclusion of more discriminative features—such as shape descriptors, deep learning-derived features, or multi-spectral data—the performance could be significantly improved.
In addition to detection and classification, this work opens the door to further developments, such as integrating basic treatment recommendations for diagnosed skin conditions. While beyond the current scope, such an extension could enable a more holistic decision support system for dermatological care.
In conclusion, this project demonstrated that machine learning models, when equipped with carefully selected image features, can provide meaningful assistance in the diagnosis of skin diseases. Although improvements are needed in data diversity and algorithm optimization, the initial results are promising and suggest a strong foundation for further research. Future work may focus on incorporating deep learning methods, enhancing real-time classification capabilities, and expanding the range of detectable skin conditions, ultimately aiming to create a robust, automated diagnostic tool for dermatological applications.

ACKNOWLEDGEMENTS:

We are grateful to Dr. Girish Padhan,Biju pattnaik University of Technology,Rourkela (Odisha), India for the help of skin disease detection and sequential solution in future.

REFERENCES

[1] K. V. Swamy, B. Divya, “Skin Disease Classification using Machine Learning Algorithms”, 2021 2nd International Conference Communication Computing and Industry 4.0 (C2I4), 2022, doi: 10.1109/C2I454156.2021.9689338 .

[2] V. B. Kumar, S. S. Kumar and V. Saboo, "Dermatological disease detection using image processing and machine learning", 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), doi: 10.1109/ICAIPR.2016.7585217

[3] P. R. Hegde, M. M. Shenoy and B. H. Shekar, "Comparison of Machine Learning Algorithms for Skin Disease Classification Using Colour and Texture Features", 2018 International Conference on Advances in Computing Communications and Informatics(ICACCI), doi: 10.1109/ICACCI.2018.8554512

[4] D.M.T. Rasanga, G.K.A.A. Tharushika, Ishara Weeruthunge, P. Bandara, "Machine Learning-Based Skin and Heart Disease Diagnose Mobile App", 2022, doi: 10.1109/ECAI52376.2021.9515126

[5] N.V Kumar, P.V. Kumar, K. Promodh, y> Karuna. "Classification of Skin Diseases using Image Processing and SVM", 2019, DOI: 10.1109/ViTECoN.2019.8899449

[6] Honey Janoria, Jasmine Minj and Pooja Patre, "Classification of Skin Disease from Skin images using Transfer Learning Technique", 2020 4th International Conference on Electronics,Communication and Aerospace Technology (ICECA), doi: 10.1109/ICECA49313.2020.9297567.

[7] T. Chauhan, S. Rawat, S. Malik and P. Singh, "Supervised and unsupervised machine learning based review on diabetes care", 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 10.1109/ICACCS51430.2021.9442021.

[8] Khushi Kumari Jha, Roshan Jha, Ankita Kumari Jha, Md Al Mahedi Hassan, Saurav Kumar Yadav, Tr Mahesh, "A Brief Comparison On Machine Learning Algorithms Based On Various Applications: A Comprehensive Survey", 2021 IEEE International Conference on Computation System and Information Technology for SustainableSolutions(CSITSS)doi: 10.1109/CSITSS54238.2021.9683524.