IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SKIN CANCER DISEASE PREDICTION SYSTEM

[1]Tanushree Choukase, [2] Gaurav Suvarna

[1]Student, [2]Student

[1]Dept. of Computer Engineering,

[1] AJEENKYA DY PATIL SCHOOL OF ENGINEERING, Lohegaon, Pune-412015.

Abstract : In addition to being dangerous, skin conditions including melanoma, eczema, and impetigo are frequently communicable, so early and precise diagnosis is essential for successful treatment. However, diagnosis usually calls for specialized understanding of dermatology, and even experts sometimes misclassify disorders, which can result in improper therapy. Our project suggests a deep learning and image processing-based skin disease detection system that uses a Convolutional Neural Network (CNN) model for precise classification in order to overcome this difficulty. This system, which is a personal computer-based program, can be implemented in places with limited resources or in remote locations, improving access to first diagnostic assistance. The system evaluates and categorizes the condition by examining a user-provided photograph of the affected skin area, providing pertinent medical information along with the projected disease.

Keywords: AI/ML, Python, Frameworks, etc.

I.INTRODUCTION

Skin illnesses have a substantial impact on the patient's psychological well-being. It may cause the patient to lose confidence and perhaps become depressed. Skin infections can be fatal. It is a big issue that cannot be ignored but must be managed. As a result, early detection of skin illnesses is critical in preventing their spread. Human skin is unpredictable and nearly impossible terrain due to its intricacy of jaggedness, lesion formations, moles, tone, the presence of thick hairs, and other perplexing aspects. We are largely unaware of the signs of the majority of these disorders, while information is fast expanding; unfortunately, this makes it difficult for dermatologists to diagnose them.

This project focuses on developing a deep learning-based application for the detection and classification of skin diseases using Convolutional Neural Networks (CNN). The application will take skin images as input, process them through image enhancement techniques (such as noise removal and grayscale conversion), and then analyze the images using a trained CNN model. The 3 output will include the predicted type of skin disease along with relevant medical information. The system is intended to support dermatologists and general practitioners by providing quick, reliable, and accessible diagnostic assistance. The scope includes dataset collection, model training, performance evaluation, and deployment in a user-friendly interface.

II. NEED OF THE STUDY.

Skin illnesses are among the most common health concerns worldwide, necessitating prompt and correct diagnosis. However, diagnosis frequently relies on dermatologists' competence and might be influenced by factors such as low contrast between lesions and skin, as well as visual similarities between healthy and affected areas. To address these problems, there is an increasing demand for automated, computer-aided diagnostic solutions that provide dependable support. This project's goal is to diagnose skin illnesses from photos by using noise-reduction filters, converting them to grayscale for better analysis, and extracting critical features for reliable assessment. The technology improves diagnostic accuracy and consistency, allowing clinicians to make more informed judgments. This technology promotes early identification, enhances patient outcomes, and helps healthcare practitioners determine the severity and urgency of skin disorders.

III. RESEARCH METHODOLOGY

This project employs a systematic method for addressing the issue .The primary methodologies consist of:

1. Data Collection Source: The dataset is sourced primarily from Kaggle, consisting of thousands of labeled images representing various skin diseases. Diversity: The dataset includes a wide range of disease categories to ensure the model learns to identify multiple conditions effectively.
2. Data Preprocessing Image Resizing and Normalization: All images are resized to a uniform dimension and normalized to maintain consistency and improve model performance. Augmentation: Data augmentation techniques such as rotation, flipping, and scaling are applied to increase dataset size and improve the model's ability to generalize under different conditions.
3. Model Training (CNN) Architecture: A Convolutional Neural Network (CNN) is designed to extract and learn key features from the images, such as texture, edges, and color variations. Training Process: The model is trained iteratively using the preprocessed dataset, adjusting weights through backpropagation to learn and recognize disease patterns accurately.
4. Evaluation Metrics Accuracy, Precision, and Recall: These metrics are used to evaluate model performance. They help ensure the model is not only accurate but also reliable in identifying both diseased and healthy skin images.

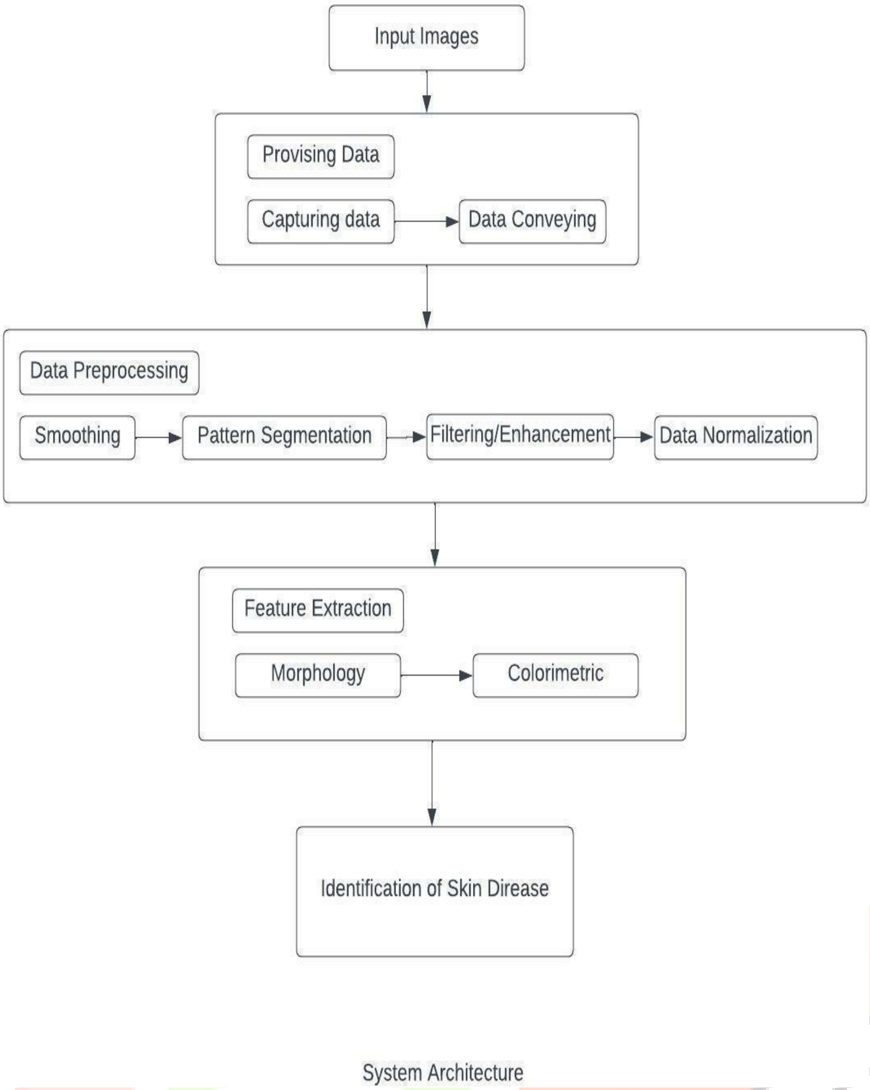

FIG 1. System Architecture

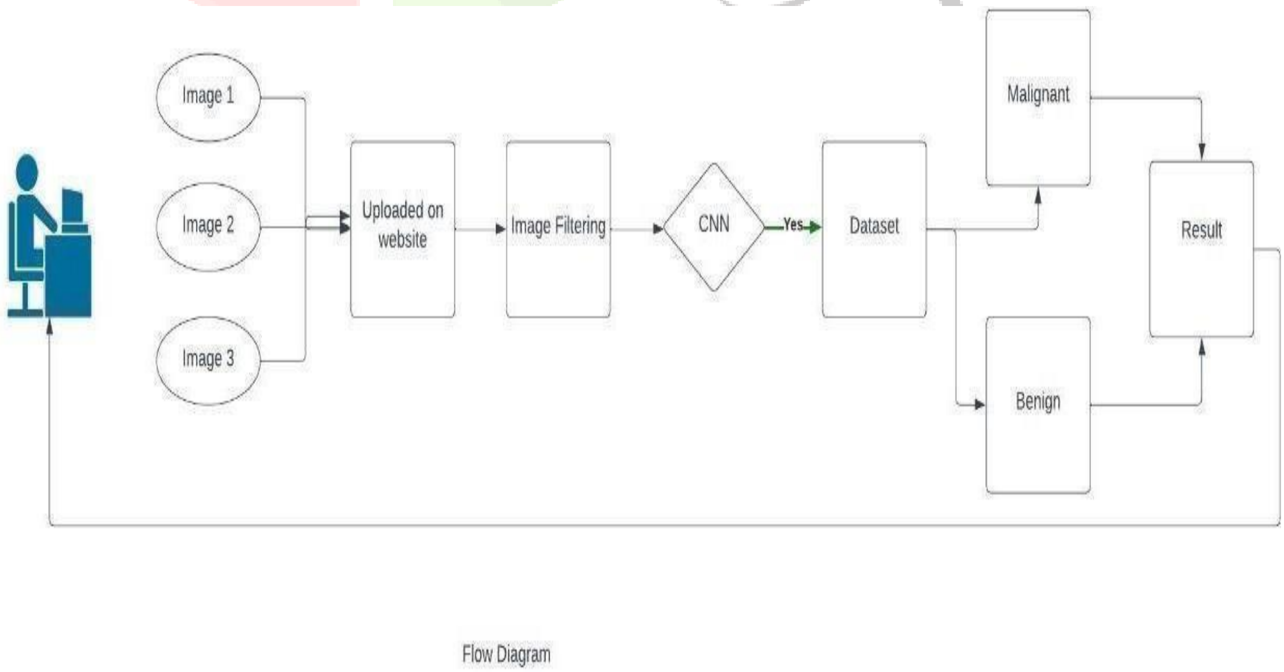

Fig 2 : Flow Diagram

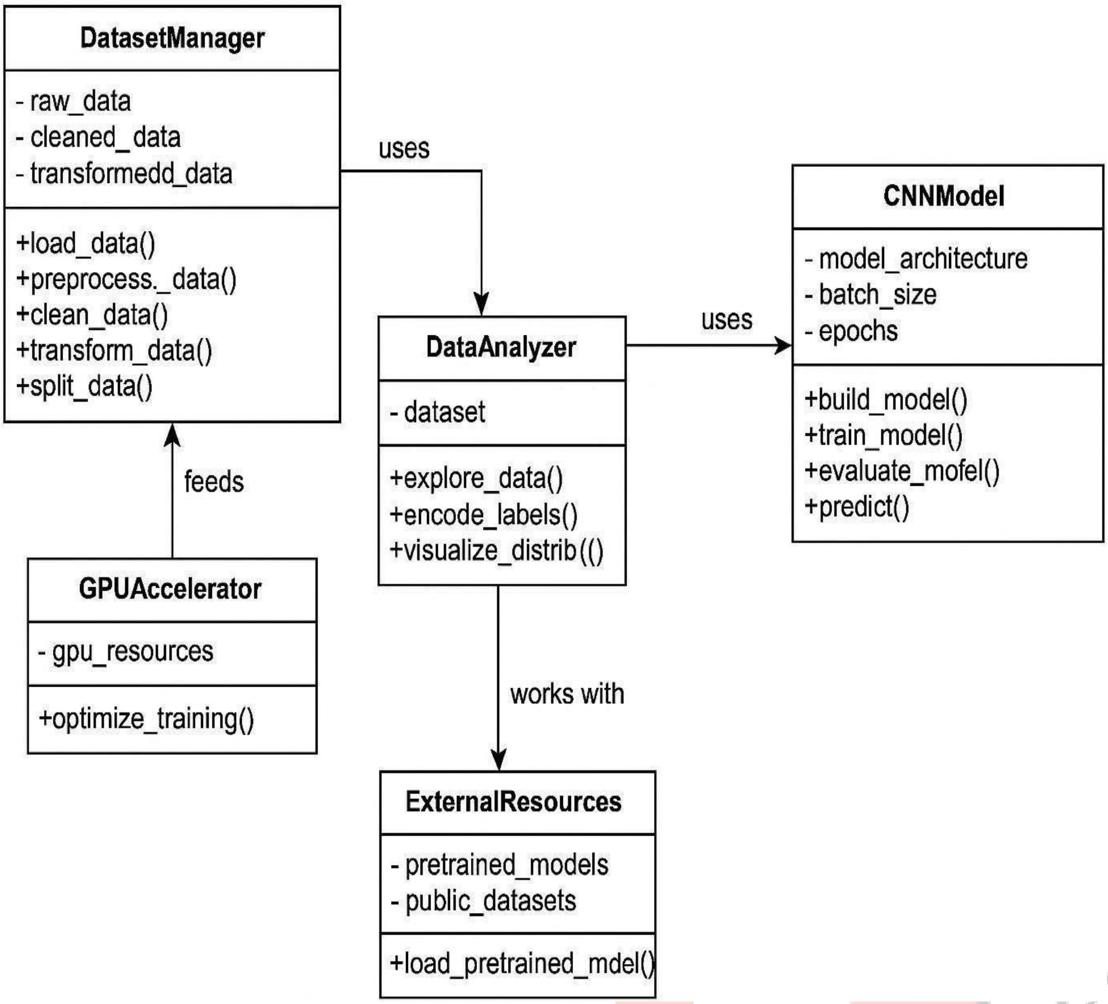

Fig 3: Class Diagram

IV. RESULTS AND DISCUSSION

Test Case

TC ID	Test Case	Input	Expected Output
TC01	Load dataset	HAM10000 dataset path	Dataset loaded successfully
TC02	Validate dataset format	CSV / image directory	Format accepted
TC03	Handle empty dataset	Empty path or file	Error message displayed
TC04	Normalize image data	Raw images	Pixel values scaled (0–1)
TC05	Resize images	Images of various sizes	Images resized to 224x224
TC06	Detect missing values	Dataset with nulls	Missing values identified
TC07	Clean missing values	Dataset with nulls	Nulls filled or rows removed
TC08	Encode labels	Text class labels	Labels converted to integers (0–6)
TC09	Augment image	Single input image	Multiple transformed images generated
TC10	Class distribution (EDA)	Labeled dataset	Chart or counts of each class
TC11	Train CNN model	Preprocessed training data	Model trained without errors
TC12	Evaluate model	Test dataset	Accuracy, precision, recall printed
TC13	Predict cancer class	New image	Predicted label + confidence score

FIG 1.1: Prediction Home Page

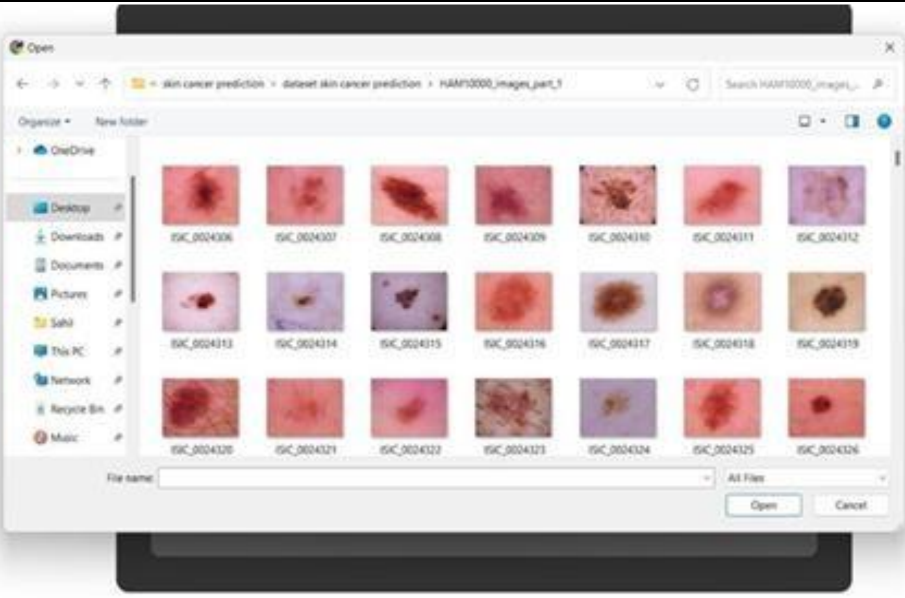

FIG 1.2 : Upload Image

FIG 1.3: Upload & Detect

Discussion

The completion of the Skin Cancer disease Prediction System signifies an important achievement. Identification of disease can help in reducing the problem of skin disease spread and will provide a better way to identify the skin problem. This will provide a low- cost way to do medical treatment without any delays. This will also help in early identification and early treatment of disease before they spread because most of the skin disease can get spread easily with touch. In our application we have used a model of Convolutional neural network . This will help in detection of skin disease in rural parts of India where there is already a huge lack of basic medical facilities.
Skin Cancer Disease Prediction System presents a promising application of deep learning in the healthcare domain, offering reliable and accessible diagnostic support. Utilizing Convolutional Neural Networks (CNNs) and trained on the HAM10000 dataset, the model has demonstrated high accuracy in classifying various types of skin cancer. This accuracy is crucial in reducing misdiagnoses and supporting early medical intervention. The system integrates image preprocessing, classification, and result

interpretation into a smooth and efficient workflow. Designed with user-friendliness in mind, the interface supports easy image upload and clear result visualization, making it accessible even to non-specialist users. Ultimately, this project bridges the gap between AI research and real-world medical applications, laying the groundwork for future.

V Future Work

- **Support for Additional Skin Conditions:**

Expand the model to detect a wider range of skin diseases (e.g., eczema, psoriasis, acne) beyond the current seven classes to enhance diagnostic capabilities.

- **Improved Deep Learning Architectures:**

Explore advanced CNN models like ResNet50, EfficientNet, or Vision Transformers to improve accuracy, efficiency, and feature extraction.**Dataset Expansion and Diversity:** Use larger and more diverse datasets from different demographics and imaging conditions to ensure better model generalization and reduce bias.

- **Real-Time Prediction Capabilities:**

Implement faster processing algorithms to enable real-time skin analysis for on-the-spot diagnosis.

- **Cloud Deployment:**

Host the model on cloud platforms (e.g., AWS, Google Cloud) to allow users to access predictions from anywhere without local processing.

- **Android Application Integration:**

Develop a full-featured Android app where users can capture or upload images, receive diagnoses, and view recommendations directly from their mobile devices.

- **User Profile and History Tracking:**

Allow users to maintain personal accounts where they can track previous predictions, monitor disease progression, and receivefollow-upadvice.

V1. References

1. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056
2. Codella, N. C. F., Nguyen, Q. B., Pankanti, S., Gutman, D., Helba, B., Halpern, A., & Smith, J. R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development, 61(4/5), 5:1–5:15. https://doi.org/10.1147/JRD.2017.2707401
3. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, 180161. https://doi.org/10.1038/sdata.2018.161
4. Brinker, T. J., Hekler, A., Enk, A. H., Berking, C., Haferkamp, S., Hauschild, A., ... & von Kalle, C. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer, 113, 47–54. https://doi.org/10.1016/j.ejca.2019.04.001
5. Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P. A. (2017). Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging, 36(4), 994–1004. https://doi.org/10.1109/TMI.2016.2642839
6. Kawahara, J., & Hamarneh, G. (2016). Fully convolutional neural networks to detect clinical dermoscopic features. In 2016 IEEE International Symposium on Biomedical Imaging (ISBI), 1362–1365. https://doi.org/10.1109/ISBI.2016.7493485
7. Abbas, Q., Celebi, M. E., & Garcia, I. F. (2013). Hair removal methods: A comparative study for

dermoscopy images. Biomedical Signal Processing and Control, 6(4), 395–404. https://doi.org/10.1016/j.bspc.2011.11.004

8. Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. Journal of Biomedical Informatics, 86, 25–32. https://doi.org/10.1016/j.jbi.2018.08.006

9. Codella, N. C. F., Lin, C. C., Halpern, A., & Smith, J. R. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging. ISIC Challenge Report. https://arxiv.org/abs/1710.05006

10. Pham, T. C., Luong, C. M., Visani, M., & Hoang, V. D. (2018). Deep CNN and data augmentation for skin lesion classification. In 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 261–265. https://doi.org/10.1109/ICAIBD.2018.8396162

11. Pytorch Documentation. (n.d.). Retrieved from https://pytorch.org/docs/stable/index.html

12. Torchvision Models. (n.d.). ResNet50. Retrieved from https://pytorch.org/vision/stable/models/generated/torchvision.models.resnet50.html

13. Flask Documentation. (n.d.). Retrieved from https://flask.palletsprojects.com/

14. Pillow Documentation. (n.d.). Retrieved from https://pillow.readthedocs.io/en/stable/

15. Werkzeug Secure Filename. (n.d.). Retrieved from https://werkzeug.palletsprojects.com/en/2.3.x/utils/#werkzeug.utils.secure_filename

16. HAM10000 Dataset. (n.d.). Retrieved from https://www.kaggle.com/datasets/kmader/skin-cancer-mnist- ham10000

17. ISIC Archive. (n.d.). Retrieved from https://www.isic-archive.com

18. PH2 Dataset. (n.d.). Retrieved from https://www.fc.up.pt/addi/ph2%20database.html

19. Scikit-learn: Machine Learning in Python. (n.d.). Retrieved from https://scikit-learn.org/

20. NumPy Documentation. (n.d.). Retrieved from https://numpy.org/doc/