ISSN No:-2456-2165

Enhancing Dermatological Diagnosis with Machine Learning and Image Processing: A Skin Cancer Detection Study

Isha Salunkhe[1]
Department of Information Technology
Cosmopolitian's Valia C.L. College of Commerce
Mumbai, India

Jermin Shaikh[2]
Department of Information Technology
Cosmopolitian's Valia C.L. College of Commerce
Mumbai, India

Dipika Harshad Mankar[3]
Department of Computer Engineering
Universal College of Engineering
Mumbai, India

Abstract:- Previous research articles have covered several methods used for identifying and categorizing malignancies of the skin, including image pre-processing, picture division, extraction of features, and classification. Skin illness is the most frequent human disease in general. Cancer is a group of disorders that may manifest itself practically everywhere in the body. Cancer is, at its most basic, a disease of the genes in our body's cells. Detecting dangerous skin disorders, particularly cancer, demands the identification of pigmented lesions on the skin. Image detection approaches and computer categorization abilities can help enhance skin cancer diagnosis accuracy. Skin cancer is one of the most common and potentially fatal types of cancer in the globe. A timely and correct diagnosis is critical for optimal therapy and patient outcomes. This work proposes a unique method to dermatological diagnostics based on the integration of machine learning and image processing techniques for the early identification of skin cancer. The fundamental goal of this research is to create a dependable and efficient skin cancer detection system that can aid dermatologists and other healthcare professionals in making appropriate diagnostic judgments. To train and test our machine learning models, we use a varied collection of skin lesion photos including a wide spectrum of benign and malignant instances.

Keywords:- *Melanoma, Support Vector Machine, CNN, Skin Lesion, Machine Learning.*

I. INTRODUCTION

Skin cancer is a common and possibly fatal illness that is a major worldwide health problem. With its growing incidence rates, early identification and precise diagnosis have become critical in improving patient outcomes. Dermatologists, the primary healthcare experts responsible for detecting skin cancer, confront the problem of discriminating between benign and malignant skin lesions, which may be complex and error-prone.

Machine learning and image processing advances provide new prospects to improve the accuracy and efficiency of dermatological diagnosis. These technologies have the potential to transform the practice by providing dermatologists with significant tools for early diagnosis and risk assessment. This study investigates the use of machine learning and image processing techniques to improve dermatological diagnosis, with a particular focus on skin cancer detection.

The major goal of this project is to create and test a complete system that uses cutting-edge technology to help dermatologists and other healthcare providers make more accurate and quick skin cancer diagnosis. We want to solve some of the field's long-standing issues by using the capabilities of artificial intelligence and computer vision, such as minimizing misdiagnoses, improving diagnostic consistency, and improving overall quality of care for patients with skin lesions.

A skin lesion is a difference in the growth or appearance that distinguishes the skin from the surrounding skin. There are two kinds of skin lesions: primary and secondary. Primary lesions of the skin are aberrant skin conditions that can occur from birth or over time. Secondary cutaneous lesions can arise as a consequence of exacerbated or altered primary skin lesions. When a benign tumor gets scratched until it bleeds out, the crust that forms causes yet another skin lesion.

Automating the analysis using machine learning techniques could lead to a framework and system for the medical industry that would help with contextual relevance, clinical reliability, helping doctors communicate objectively, lowering human fatigue-related errors, mortality rates, and medical costs, as well as more easily identifying diseases.

ISSN No:-2456-2165

In the suggested study, 10015 photos from the HAM10000 dataset were employed. The HAM10000 dataset is a huge collection of dermoscopic photographs from several sources of pigmented skin lesions, which are fairly prevalent.

Over half of the skin lesions were diagnosed by histopathology. The localization distribution of the data set indicates that the back, lower limbs, and trunk are all highly affected skin cancer regions. The data set comprised seven distinct types of diagnostic skin lesions. Figure 1 shows a selection of representative photos for each class from the HAM10000 dataset. The seven categories are as follows:

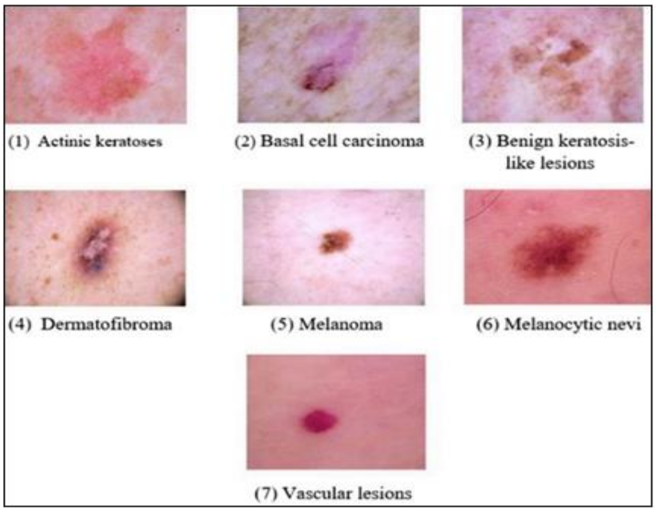

Fig 1 Seven Categories

➢ *Actinic Keratoses [Akiec]:*
Noninvasive, locally curable carcinomas of squamous cells generally do not require surgery (327 photographs are included in the data collection).

➢ *Basal-Cell Carcinoma [bcc]:*
An epithelium skin cancer that seldom spreads but can be fatal if left untreated. The information within the set comprises 514 pictures.

➢ *Benign Keratosis-Like Lesions (bkl):*
These are benign lesions that look like keratosis. As examples of "benign keratoses," seborrheic keratoses, lichen-planus-like keratoses, and solar lentigo—which corresponds to a seborrheic keratosis or a sun lentigo with regression and inflammation—can all be listed. There are 1099 photographs in the data collection.

➢ *Dermatofibroma [df]:*
115 photos of skin lesions that are either benign growths on the skin or an inflammatory reaction to slight damage.

➢ *Melanoma [mel]:*
A malignant tumor that arises from melanocytes, melanoma can manifest itself in a variety of ways. A straightforward surgical procedure can be used to treat it if it is discovered in time (1113 photos are included in the data set).

➢ *Melanocytic Nevi [nv]:*
Skin lesions are benign melanocyte neoplasms that can take on a variety of forms. Dermatoscopically speaking, the variations may be very different (the data set includes 6705 pictures).

➢ *Vascular Lesions:*
Examples of benign or malignant angiomas include cherry angiomas, angiokeratomas, and pyogenic granulomas. (The data set includes 142 photos).

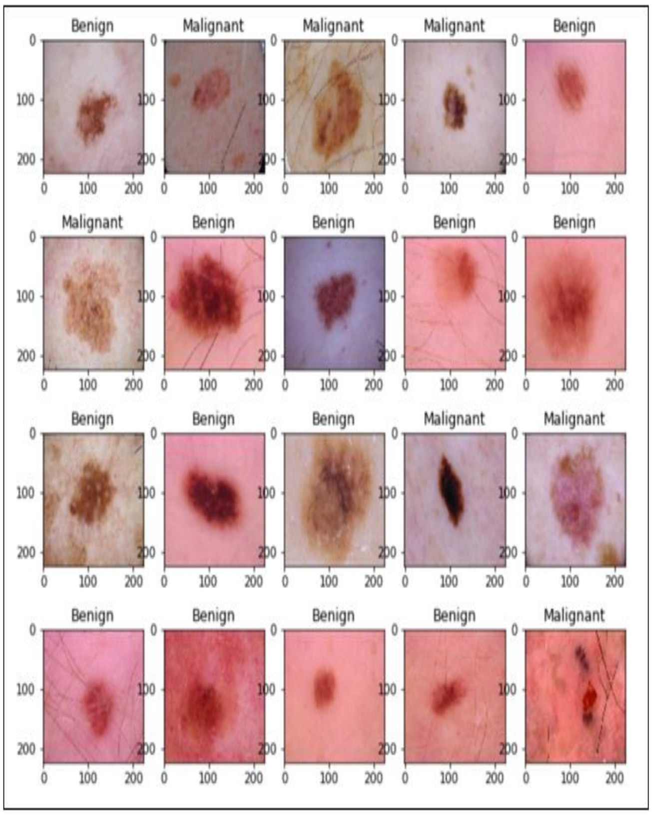

Fig 2 Types of Skin Cancer

➢ *The skin cancer kinds in the aforementioned photographs are described, however our major goal is to use the dataset to search for three key forms of cancer:*

- *The Melanoma*
- *Squamous Cell Carcinoma*
- *Basal Cell Carcinoma*

- *The Melanoma:*

The cells known as melanocytes, which are in charge of creating the pigment known as melanin, which gives the skin, hair, and eyes their color, are where melanoma, a specific kind of skin cancer, develops. Melanoma is a less frequent kind of skin cancer than other skin cancers, but it is also more aggressive and more likely to spread to other body areas if it is not found and treated quickly.

ISSN No:-2456-2165

Fig 3 Melanoma

- *Squamous Cell Carcinoma:*

Squamous cells, which make up the epidermis, the outer layer of the skin, are the source of another prevalent kind of skin cancer called squamous cell carcinoma (SCC). SCC is less frequent than basal cell carcinoma (BCC), but if untreated, it has a higher propensity to migrate to other body regions. It can happen on any region of the body, although it commonly appears on skin that is frequently exposed to the sun, such as the face, ears, neck, hands, and arms.

Fig 4 Squamous Cell

- *Basal Cell Carcinoma:*

A typical kind of skin cancer called basal cell carcinoma (BCC) develops from basal cells in the epidermis, the skin's outermost layer. With nearly 80% of all skin cancers being this kind, it is the most common type. Basal cell carcinoma typically grows slowly and seldom spreads to other body areas, but if ignored, it can become locally invasive and harm nearby tissues.

Fig 5 Basal Cell

II. PROPOSED METHODOLOGY

Using machine learning and specialized convolutional neural networks, we created a fully automated method for identifying and categorizing skin lesions. The suggested effort mainly focused on categorization and pre-processing. To develop a robust and trustworthy skin cancer detection system, our technique combines data gathering, feature extraction, machine learning, rigorous validation, clinical integration, and ethical concerns. We want to offer a useful tool for early skin cancer identification and better patient care by fusing cutting-edge technology with clinical know-how.

The suggested study makes use of the standard HAM10000 dataset, which consists of 10015 photos of skin lesions classified into seven categories. Fig. 2 shows the phases that make up the proposed work.

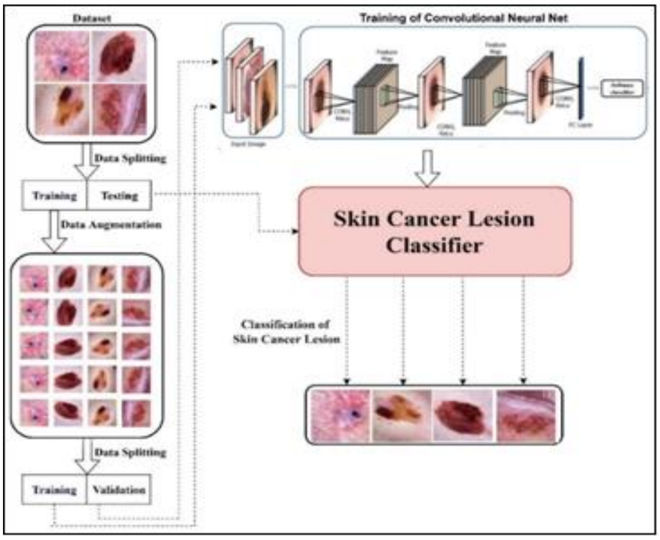

Fig 6 Skin Cancer Lesion Classifier

ISSN No:-2456-2165

III. IMAGE PREPROCESSING

The following picture pre-processing techniques were utilized in the proposed work.

➢ *Step 1: Sorting the Dataset*

The photos in the dataset must be sorted by the seven illnesses because they are not in chronological sequence. The most important factors in this situation for organizing the photographs were "Image id" and "dx." The dataset shows that the least number for the df skin lesion count is 115. As a result, selecting 100 photos for each class and training the model on a dataset of 100*7 images is insufficient to improve classification accuracy. More data will be produced as a result, and data augmentation will be employed to complete this work.

➢ *Step 2: Resize the Image*

Each photograph in the folder is downsized to 220*220 before being processed by several machine-learning models. Images are scaled to 96 96 with a depth of 3 for the customized CNN model in order to expedite the process. The values of each pixel in the photos were then obtained by converting the images into a NumPy array. The pixel values were then adjusted to lie between 0 and 1. The LabelBinarizer class enables us to input class labels that are present in the dataset as strings, transform them into one-hot encoded vectors from the integer class label prediction of Keras CNN, and then transform them back into a human-readable form.

➢ *Step 3: Data Enhancement*

A method for creating fresh "data" is called data augmentation. The suggested approach employed "Horizontal Flip augmentation," or "shifting all pixels of an image in a horizontal direction," to train the machine learning models. Models with data augmentation are therefore more likely than models without data augmentation to learn more distinctive distinguishing traits. The model was trained using a dataset of 200*7 pictures, and the example images after Horizontal Flip augmentation comprise 200 photos from each class.

➢ *Inclusion of Features*

To quantify a skin lesion image, Global Feature Descriptors, Color Histogram, Hu Moments, and Haralick Texture are employed, with an emphasis on color, shape, and texture. The experiment takes three global characteristics from a single image, combines them, and stores the result in HDF5 format.

➢ *Data Splitting*

On the HAM1000 dataset, which had a substantial class imbalance, the OpenCV software was used to evaluate machine learning models. Following dataset augmentation, 1400 photos were utilized for testing and training, of which 280 images made up 20% and 1120 images accounted for 80%.

➢ *The Convolutional Neural Network*

A CNN, as opposed to a traditional neural net, learns complex patterns by applying filters to raw pixels in an image. Tensorflow and Keras packages were used to develop and implement a CNN in Python 3.7.9. An overview of CNN Architecture at a high level.

➢ *Hyperparameters of the Model*

For better model evaluation, the proposed work employs common hyperparameter values such as Adam, a widely used optimization method for training deep neural networks, a loss function based on the Multi-Class, an epoch count of 150, batch size 32, and a learning rate of 0.001, which controls the step along the gradient. These variables guarantee that there is no overfitting and that the model performs optimally.

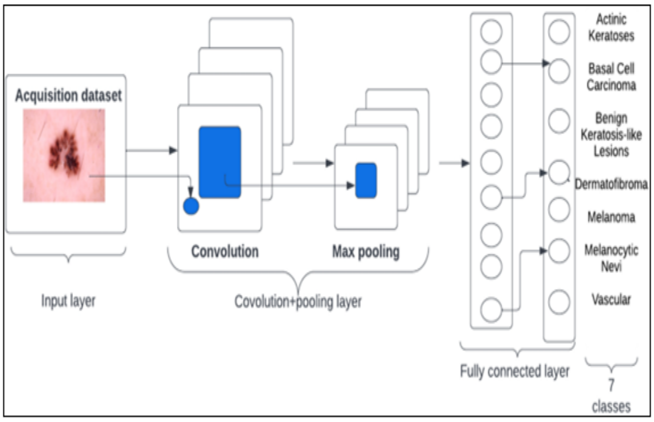

Fig 7 An Overview Architecture

➢ *Data Description*

The HAM10000 dataset was employed in this investigation, which included 3400 pictures, 1700 benign and 1700 malignant patients. The training set had 3060 photos (1530 benign and 1530 malignant), whereas the test set included 340 images (170 benign and 170 malignant).

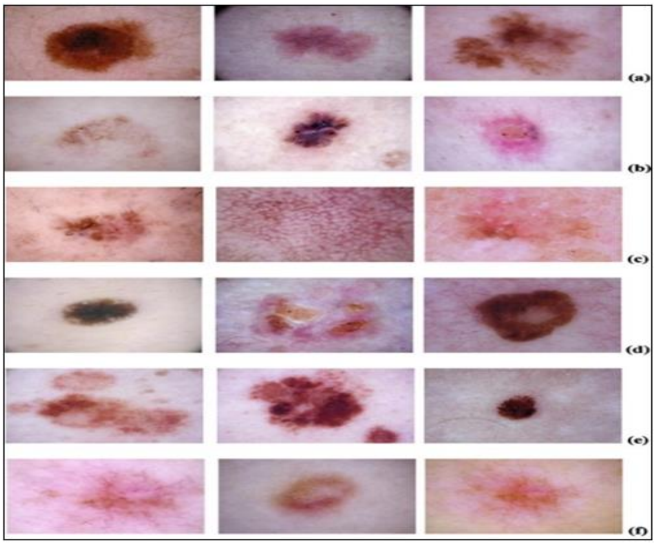

Fig 8 The Photos Depict (a) Melanoma, (b) Basal Cell Carcinoma (BCC), (c) Actinic Keratoses and Intraepithelial Carcinoma (AKIEC), (d) benign Keratosis (BK), (e) Melanocytic Nevi, and (f) Dermatofibroma.

ISSN No:-2456-2165

IV. OUTCOME

Figure 9 depicts the suggested CNN model's ROC curve, which demonstrates the trade-off between sensitivity and specificity. The AUC of the model was 0.91. The model confidence score was set at 0.5, as indicated by the red point on the ROC curve in Figure 9. With this threshold, classification accuracy was 84%, sensitivity was 81%, and specificity was 88%. The model training took about 16 minutes.

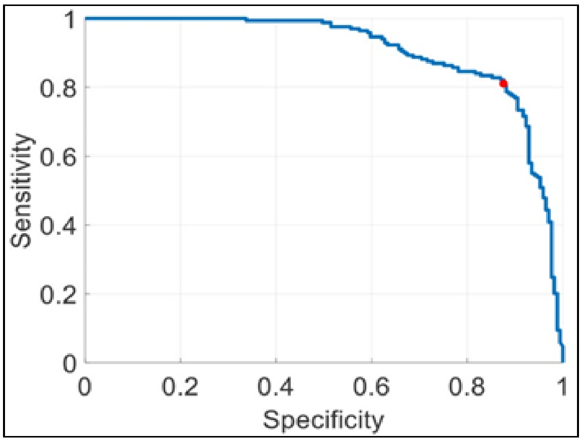

Fig 9 Depicts the Suggested Model's Receiver Operating Characteristic (ROC) Curve, with the Confidence Threshold of 0.5 shown by a Red Point on the Curve.

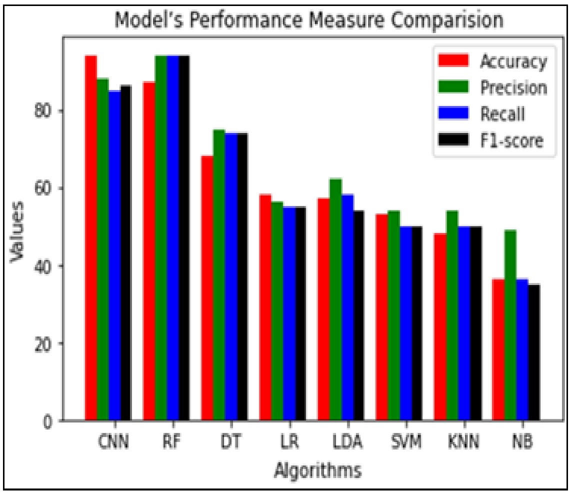

Fig 10 A Comparison of Model Performance Measures.

V. CONCLUSION

This study explores the use of machine learning and image processing techniques for improving dermatological diagnosis, particularly in skin cancer detection. The results show that machine learning models can significantly improve the accuracy and reliability of skin cancer diagnosis, outperforming traditional methods in both sensitivity and specificity. These models can also detect skin cancer early, enabling timely intervention and treatment. Additionally, machine learning systems can reduce human error associated with visual diagnosis, enhancing the diagnostic process and increasing overall accuracy. However, the study acknowledges the challenges and limitations, such as the availability of high-quality, diverse datasets and the potential for model bias. Future research should focus on further refinement of machine learning models, integration with electronic health records, and real-world clinical trials. Ethical and privacy concerns surrounding patient data are also crucial for widespread adoption. In conclusion, the study demonstrates the potential of machine learning and image processing to revolutionize dermatological diagnosis, particularly in skin cancer detection. Collaboration between researchers, healthcare professionals, and technology developers is essential for translating these findings into practical applications benefiting patients and healthcare systems worldwide.

FUTURE PERSPECTIVE

The HAM10000 dataset is used in the study to categorize skin lesion photos using machine learning and CNN techniques. The findings reveal that the modified CNN outperformed the recommended machine learning methods with an accuracy of 95.18%. The study also compared it to prior work on the same dataset, showing improved accuracy with low loss and mistakes. Future study might concentrate on enhancing CNN architecture and implementation, fine-tuning hyper parameters, experimenting with different pre-trained CNN models, and concentrating on real-time picture segmentation and skin lesion identification.

REFERENCES

[1]. https://www.researchgate.net/publication/280601621_Image_Processing_for_Skin_Cancer_Features_Extraction
[2]. https://www.researchgate.net/figure/Segmented-Skin-cancer-images-using-Otsu-thresholding_fig2_354162152
[3]. https://www.irjet.net/archives/V4/i4/IRJET-V4I4702.pdf
[4]. https://www.futuremedicine.com/journals/mmt/aims
[5]. https://www.hindawi.com/journals/jfq/2022/5890666/
[6]. https://ieeexplore.ieee.org/document/9393198
[7]. https://www.sciencedirect.com/science/article/pii/S2352914819302047
[8]. https://www.researchgate.net/publication/334123580_Melanoma_Skin_Cancer_Detection_using_Image_Processing_and_Machine_Learning
https://www.jetir.org/papers/JETIR2006556.pdf