ISSN No:-2456-2165

# Automated System to Classify and Detect the Skin Lesion

Samiksha Dhote, Prajakta Dhumal, Prajwal Gaidhani, Indrajeet Ghadge and Prof.S.R.Nalamwar Department of Computer Engineering,AISSMS College of Engineering,Pune,India

**Abstract:- Deep learning and image processing techniques for skin disease identification are part of the suggested solution. Skin conditions are brought on by a variety of reasons, including DNA sequencing, radiation, and mutations, which result in skin defects. If skin conditions are not treated in a timely manner, they often spread to other parts of the body. In order to be treated, skin disorders must therefore be found in their early stages. These symptoms need to be identified early because skin disorders have been linked to mortality problems, lengthy, expensive therapies, and numerous characteristics.**

**We are creating web application for early stage skin disease detection. In the suggested solution, skin illnesses will be identified from the provided image collection using image processing techniques which uses a color image's inputs. Depending on the training dataset, we have used the Convolutional Neural Network which has an excellent visual representation power for the recognition or detection task. Therefore in order to diagnose skin diseases early with more accuracy and efficiency transfer learning will be used . It will provide the best training to the model and give a high accuracy, precision, recall, specificity for the correct anticipation of skin diseases among all picked which help doctor for early detection and prevent chronic disorder as well as economic and mental loss**

***Keywords:-*** *Data Preprocessing, Outlier Detection, Image Classification, Deep Learning, Convolution Neural Network, Transfer Learning.*

## I. INTRODUCTION

### *A. Background:*

Skin is one of the most important tissue of the human body. The occurrence of the Skin Diseases has the negative impact on the human society. Various skin diseases are being detected nowadays and these diseases are traumatizing the mentality of people with the social and economic significance. Along with the several skin diseases on the other hand skin cancer is lethal and can trouble the patient therefore the stage of the disease need to be known .It is good if the diagnosis is to be done at very early stage. In general, most of the common people do not know the type and stage of a skin disease. Some of the skin diseases show symptoms several months later, causing the disease to develop and grow further. This is due to the lack of medical knowledge in the public. Sometimes, a dermatologist may also find it difficult to diagnose the skin disease and may require expensive laboratory tests to correctly identify the type and stage of the skin disease. The advancement of lasers and photonics based medical technology has made it possible to diagnose the skin diseases much more quickly and accurately. But the cost of such diagnosis is still limited and very expensive. Therefore, we propose a deep learning approach to diagnose the skin diseases in the early stages such that the treatment of the disease will be much easier.

### *B. Problem statement*

Most of the world population is suffering from skin diseases so it is important to detect these skin diseases at early stages. The main aim is to create a solution which can Classify and Detect the Multiple Skin Diseases in humans by using the Deep Learning techniques. Also to create a user interface this will provide required solution to user. This automated system will be ease for the end user to operate the model, since it will not need to worry about the backend or working of the model to receive the required results.

### *C. Purpose*

- *Why we use Deep Learning?*

Deep learning is the part of machine learning during which multi layered neural networks modeled to figure out the solutions just like human brains-'learn' from the massive amount of information, among every layer of neural network, deep learning algorithms perform calculations and build predictions repeatedly, increasingly 'learning ' bit by bit the accuracy of the end result. As the human body get the senses by its sense organs the deep learning model ingests the info from multiple information sources and then it will analyze it in the real time projects

### *D. Scope*

The model of CNN used up over here can be implemented in the multiple domains. We can implement this classifier even on more large dataset consisting of various types of images. More complex models can be implemented and used over to increase the accuracy and for improving the classification of the skin diseases. Also for the better results to be accommodated we need to have the images of the early stages such that it would reduce the threat of chronic diseases and also the threat to life. Further using the various types of images we can classify the melanoma,non-melanoma, psoriasis, acne thus reducing the mortality issues and building up life.

## II. LITERATURE REVIEW

1] .Skin Lesion Classification of Dermoscopic Images using Machine Learning and Convolutional Neural Network :

In this research paper,they have worked on HAM10015 dataset images using machine learning and convolutional neural Networks.Their result show that their CNN has achieved accuracy of 95.18% which was better from all machine learning algorithms.We can improve this result by adding more layers,different types of layers ,transfer learning.

2] Skin Disease Classification from Image – In this research paper, they have used ISIC DATA SET And CNN classifier

3] Automatic diagnosis of skin diseases using convolutional neural network : In this research paper ,they have classified skin diseases using AlexNet Architure. The model was trained using Dermnet database ,total images used to train were 105 and 69 images were used to test the model's performance. They have achieved the accuracy for the skin diseases such as acne,keratosis,Eczema herpeticum and urticaria , they are as follows 85.7%, 92.3%,93.3%,92.8% respectively

4] Skin Lesions: HAM10000 dataset and BIODS220 PROJECT: In this research paper, they have implemented various approaches mix-up data augmentation, two types of supervised learning algorithms for the pre training the model and finally Efficient Net for best average recall. For results, they used 20 epochs for training with Adam optimizer and achieved 0.001 rate of learning, and for testing 1511 images were used to get best model possible which has accuracy of 0.891 and recall of 0.784. According to them this result can be improved by more training the dataset

5] Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge: In this research paper, they have used the data from dermatology department of Peking Union Medical College Hospital, which has 28,000+ images. Their project was built on global standards where each image went under two doctors which have dermatology as specialization. They have used two datasets; one consists all images and another consists of evenly distributed dataset. The algorithm they have used GoogleNet inception v3 code package to develop their algorithm. To avoid errors, the ran their experiment multiple times and achieved accuracy average and standard deviation for multiple runs for both datasets. The accuracy for dataset A 87.25% was achieved and for dataset B 86.63% was achieved. According to them, this result can be improved by using larger dataset and also including multiclass diseases in the dataset

6] Multiclass Skin Lesion Classification Using Hybrid Deep Features Selection and Extreme Learning Machine : In this research paper, they have used two dataset ,one is HAM10000andanother is ISIC2018 dataset. They have used deep learning and the best feature selected fusion to build the model . They have also implemented various classifiers like MC-SVM, DT, NaiveBayes, EBT on both dataset. The Accuracy Achieved on HAM 10000 are 84.72,79.63,81.50,82.96 respectively.Theyalsocompared accuracy after feature selection which occurred minor improvement in accuracy and major improvement in computation time. The Accuracy on ISIC2018 dataset were achieved as follows 83.90,82.56,83.50,82.16 respectively for each classifier, furthermore accuracy was increased with decrease in computational time.

## III. METHODOLOGY

### *A. System Architecture*

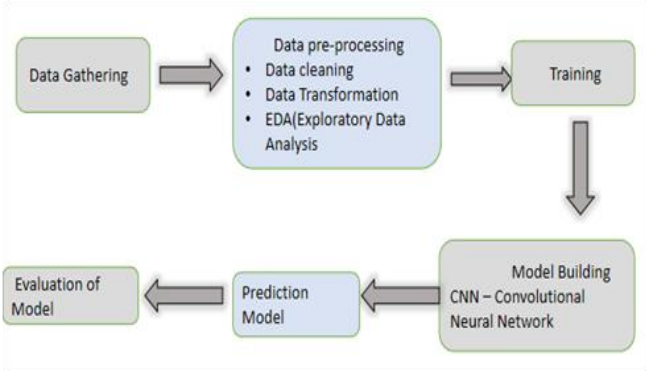


Fig. 1 Flowchart of system

### *B. Data Gathering:*

Machines learn from the data that we fetch into them. It is important to collect reliable data stat machine learning model can find out the correct patterns and improve the reliability. The Quality Of The data that you feed to the machine will determine the accuracy of model.

➢ *Dataset: Finalisation of the dataset is from: Kaggle*

The proposed system has been developed on dermatoscopic images which is collected from publicly available dataset based on Skin Diseases Named as - Dermnet. 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 the test set. The categories include Acne, Melanoma, Eczema,Seborrheic Keratoses, Tinea Ringworm, Bullous Disease, Poison Ivy, Psoriasis, Vascular Tumors,etc.

### *C. Data Preprocessing and Enhancement :*

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model.

➢ *Import all the crucial libraries :*

The primary Python libraries utilized in this machine learning data pretreatment & Image processing are:

- NumPy (performing scientific calculations)
- Pandas (toolkit for handling and analyzing data)
- Matplotlib (2D charting tool)
- scikit-learn (open-source image processing Python libraries)

➢ *Import the dataset :*

The dataset(s) that have been gathered for the current Machine Learning project must be imported in this phase.

- Load Data with Python Standard Library.
- Another way to load your machine learning data is using Pandas i.e pandas.read_csv() function.
- NumPy and the numpy.loadtxt() function can also be used to load machine learning data in Python.

*D. Data Cleaning :*

Once the selection of the dataset is done the user need to prepare the dataset for the model building The dataset can be prepared by the following methods:

Fetching together all the data and then randomizing it. This process helps to make sure that the data upon which the model need to be built is evenly distributed and the order of the data must not affect the learning process

Cleaning of the data to remove the null values,duplicate values, remove outliers, data type conversions etc.For building of the model there may be a need of restructuring of the data by Principal Component Analysis which supports in dimensionality reduction where we can change number of rows, columns and index of data.

*E. Data Transformation :*

Data Transformation involves converting data from one format into another. It involves transforming actual values from one representation to the target representation.

*F. Exploratory Data Analysis (EDA) :*

In this we explore different features of the dataset,their distributions and actual counts. We Have Tried to visualize the data by means of correlation matrix, Heatmap, scatter Plot.

Visualization of the dataset to understand the relationship between the several classes and the variables present in the dataset.

Data visualization provides a good, organized graphical representation of the data. It makes easier to understand, observe, analyze the dataset

*G. Splitting The Data :*

Splitting of the dataset into two components:

➢ *Training dataset:*

The training dataset is the information that the model learns from. Training data is the subset of original data that is used to train the machine learning model.

It is generally larger in size compared to the testing dataset.

The general ratios of splitting train and test datasets are:80:20, 70:30, or 90:10.

➢ *Testing dataset:*

A testing set is used to check the error and the accuracy of the model after training. The testing data is like new data to the model.

*H. Choosing a Model :*

A machine learning model determines the output that the user get after training of the model by the several algorithms on the fetched dataset. Over the time the scientists developed various algorithms for training of the model, that can accomplish the data. Apart the user needs to choose the model according to the continuous and the categorical data.

*I. Model Building* **:**

➢ *CNN*

We have used Convolutional Neural Network (CNN) to determine the type of skin lesion and detect the disease.

CNN(Convolutional Neural Network) is used to recognize the images and classify it. We had used CNN in which the input is taken as image,processed it and then it had been classified the certain category or the class of the specified disease. Eventually, we had used deep learning that utilize CNN Model with several convolutional layer, pooling layer to determine the progression over the images to classify it into a specific category

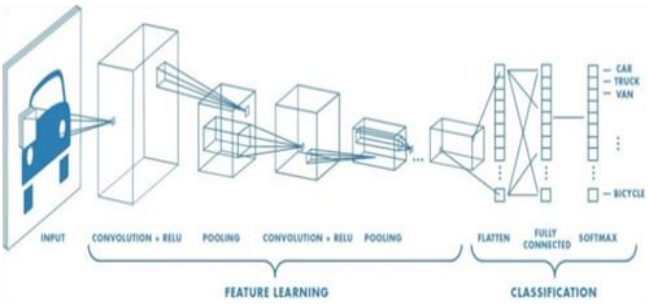


Fig 2 :- CNN

The convolutional neural network is classified as following layers :

*1) Convolution Layer:*

The first layer is convolution Layer. It works feature extraction by sliding the filter over the input image. The output will be convoluted feature is for the image and their sum for every sliding action.

The layer is meant for extracting the generic features and the complex part are removed the slayer.

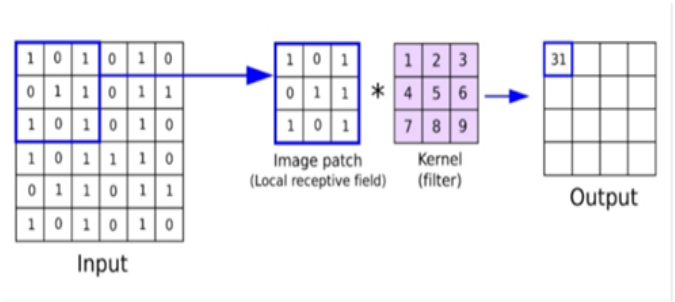


Fig 3:- Convolution Layer

*2)Pooling Layer :*

This is the second layer of the CNN. Thislayerhelp in decreasing the computational power by the means of dimensionality reduction. Further it does the task of or behaves like that of Principle Component Analysis.Eventually, this can be classified into two types :

Max Pooling – This operation select the maximum element from the image containing the most useful feature of the previous feature map. Average Pooling – This operation select the average element from the feature map covered by the features in the image.

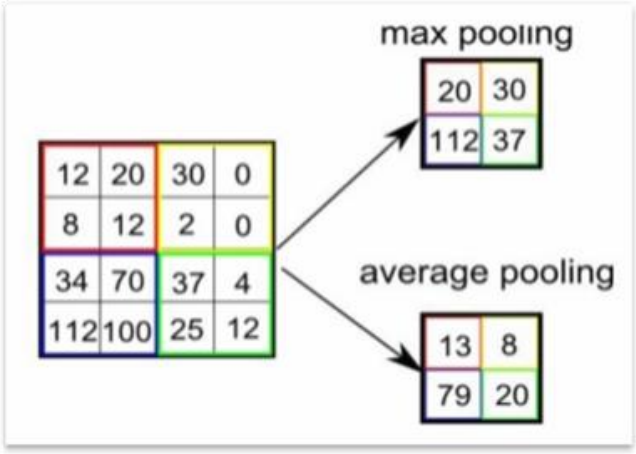


Fig 4:- Pooling layer

*3)Fully Connected Layer :*

This is the third layer of the CNN. The classification of the model architecture is done by the fully connected layer. It is the Non Linear function that convert into flattened one dimensional layer to create the model. The Softmax or Sigmoid are the activation function that are used to classify the output.

➢ *TRANSFER LEARNING :*

Transfer Learning is method to reuse the already trained model on new problem statement. We use Transfer Learning for the improvement in the model by transferring the knowledge from the existing model to the related task that has already been learned.

The machine uses the knowledge from pre-trained model where it reduces the resources and amount of the labeled data to trained the newmodel. We do not need to start the newmodel from the scratch. The new task will be related to the previously trained task reducing the time and human effort with the less consumption of the resources. This task involved the unlabelled data as well.

The main benefits of transfer learning for machine learning include:

- Removing the need for a large set of labeled training data for every new model.
- Improving the efficiency of machine learning model.
- Models can be trained within pre-trained model.
- Provides the generalized approach for the complex model.

✓ *How Transfer Learning Works*

The CNN aims at detecting the edges in the first layer whereas the Transfer learning will be task to the feature in the next layers. The next layersarere-trained by the means of transfer learning. It make the use of labeled and unlabeled data from the pre-trained model.

✓ *ResNet 50:*

ResNet-50 is the method in transfer learning which uses convolutional neural network this is 50 layers deep that is 48 convolutional layers, oneMaxPool layer, and one average pool layer. It was built and trained by Microsoft in2015. Thismodel is also trained on more than 1 million images from the ImageNet database. It can classify up to 1000 images and the network was trained on 224x224 pixels colored images. The main motivation behind this model was to improve the accuracy and reduce the error rate as the layers went on increasing it become deeper. Also, the ResNet model eliminates the Gradient Issues aimed.

Here is brief info about its size and performance of Resnet-50:
Size: 98 MB
Top-1: Accuracy: 74.9%
Top-5: Accuracy: 92.1%
Number of Parameters: 25,636,712

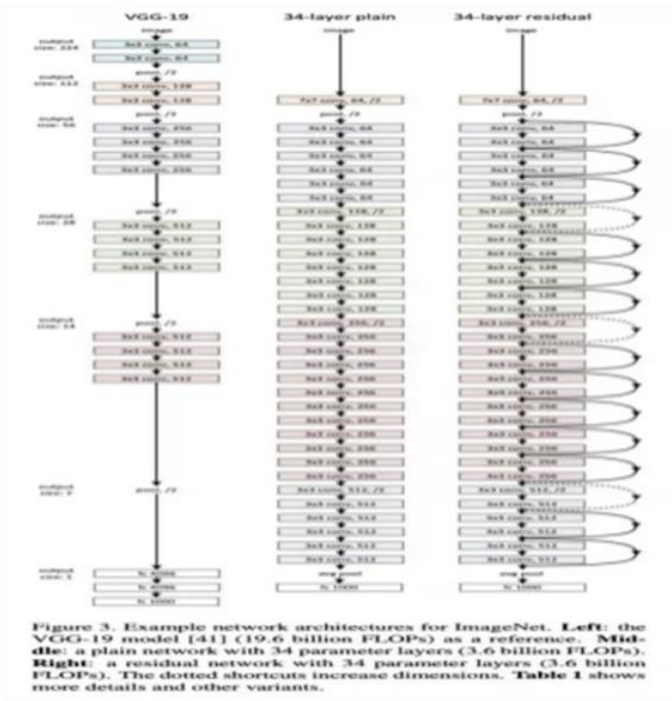


Figure 3. Example network architectures for ImageNet. **Left**: the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle**: a plain network with 34 parameter layers (3.6 billion FLOPs). **Right**: a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.

Fig 5 ResNet 50

## IV. MODEL EVALUATION

Least the error rate improve the quality of the model trained. The Model is evaluated on the basis of the least error and the most accuracy where as the accuracy can be classified as the validation and test accuracy. The validation dataset is different from the trained dataset and is used for selected parameters.

With these parameters we can use confusion matrix to display the relationship between accuracy, error rate, precision and recall.

## V. DISCUSSION

The paper summarizes most of the algorithm applicable for the detection the skin lesions by different method. Below is the comparison table between skin lesions detection algorithm with respect to the given parameter:

ISSN No:-2456-2165

Table1 : Classifier used in various model.

| Dataset used | Classifier Used | Accuracy in percentage |
|---|---|---|
| Dermnet | CNN | 98.6 to 99.04 |
| Ham10K | CNN | 95.18 |
| Dermnet | Vgg16 | 72.5 |
| Dermnet | Vgg19 | 73.5 |
| Dermnet | GoogleNet | 70.8 |
| Ham10K | ResNet | 90.5 |

## VI. CONCLUSION

In this paper we have presented a widespread literature survey on the detection of various skin diseases and lesions. The main objective of the algorithm is to detect the skin lesion by the automated system in the real time.

The transfer learning model is used over here will provide the greater accuracy with the speed and the performance for the early stage detection of the various dermatoscopic skin related diseases.We are expecting the higher accuracy while using the transfer learning model in our automated system to detect the skin lesions

## REFERENCES

[1]. Tanvi Goswami Vipul K. Dabhi Literature review Harshadkumar B. Prajapati,Skin Disease Classification from Image - A Survey ,Published in: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)

[2]. T.Shanthi,R.S.Sabeenian,R.Anand,2020, Automatic diagnosis of skin diseases using convolutional neural network

[3]. Farhat Afza , Muhammad Sharif , Muhammad Attique Khan , Usman Tariq , Hwan-Seung Yong and Jaehyuk Cha,(21 January 2022 )Multiclass Skin Lesion Classification Using Hybrid Deep Features Selection and Extreme Learning Machine.

[4]. Anastasia Butskova ,Skin Lesions :HAM10000 dataset and BIODS220 PROJECT

[5]. Bhuvaneshwari Shetty,Roshan Fernandes,Anisha P Rodrigues ,Kuruva Lakshmanna ,(Oct , 2022 )Skin Lesion Classification of Dermoscopic Images using Machine Learning and Convolutional Neural Network .

[6]. Xinyuan Zhang, Shiqi Wang,Jie Liu, and Cui Tao,(2018)Towards improving diagnosis of skin diseases by combining deep neural networks and human knowledge.

[7]. Rehman,S.H. Khan, S.D. Rizvi,Z. Abbas and A. Zafarthey,2018 ,Classification of skin lesions by interference of segmentation and convolution neural network.

[8]. I. Zaqout ,2019,Diagnosis of skin lesions based on thermoscopic images using image processing TECHNIQUES.

[9]. Andre Esteva , Brett Kuprel , Roberto A Novoa , Justin Ko , Susan M Swetter , Helen M Blau , Sebastian Thrun , Dermatologist-level classification of skin cancer with deep neural networks, - 2017

[10]. S. Ranjana ,R. Manimegala, K.Priya,(2020),Deep Learning System for Skin Disorder Segmentation using Neural Network

[11]. Nawal Soliman ,ALKhalifi, ALEnezi,2019 ,A Method of Skin Disease Detection Using Image Processing and Machine Learning

[12]. Li-sheng Wei, Quan Gan, and Tao Ji, Herpes Dermatitis Psoriasis,2018,Skin Disease Recognition Method Based on Image Colorand Texture Features.

[13]. I. Ketut Eddy Purnama, Arta Kusuma Hernanda, Anak Agung Putri Ratna, Ingrid Nurtanio, Afif Nurul Hidayati, Mauridhi Hery Purnomo, Supeno Mardi Susiki Nugroho, Reza Fuad Rachmadi ,(Nov 2019)Disease Classification based on Dermoscopic Skin Images Using Convolutional Neural Network in Teledermatology System

[14]. T. Shanthi, R. S. Sabeenian, and R. Anand,( 2020.),Automatic diagnosis of skin diseases using convolution neural network.