Classification of Skin Disease Image using Texture and Color Features with Machine Learning Techniques

M. Kalaiyarivu

Research Scholar
Department of Computer and Information Science
Annamalai University
Chidambaram, India
Kalaiyarivu2016.c@gmail.com

Dr. N.J. Nalini

Associate Professor and Research Supervisor
Department of Computer Science and Engineering
Annamalai University
Chidambaram, India

Article Info
Page Number: 682 - 699
Publication Issue:
Vol 71 No. 3s2 (2022)

Article History
***Article Received**: 28 April 2022*
***Revised**: 15 May 2022*
***Accepted**: 20 June 2022*
***Publication**: 21 July 2022*

Abstract

Skin disease is a serious condition caused by DNA that can lead to death. This damaged DNA begins to develop uncontrollably in cells, and it now multiplies rapidly. Digitalized analysis of malignancy in skin ulcer images is being investigated. However, analysing these images is difficult due to various distracting variables such as reflections on the skin surface. Lesions of various shapes and sizes, as well as variations in color light. As a result, verification is done automatically. Recognizing skin disorder is important for pathologists to develop accuracy and skill early on. In this paper present a convolutional neural network model built from Deep Learning and compare it to certain machine learning tools for accurate classification of the following seven categories of skin diseases (Akiec,Bcc,Bkl, Mel, Nv, Df,Vasc). To get the graylevel features, we first used image resizing and then transformed RGB to Grayscale image. Second, apply a filter to remove undesired artifacts and noise. Finally, averaging the input images and extracting texture feature(GLCM,LBP) and color feature (variance, Entropy) characteristics aids in classification accuracy. The CNN model gives 84% outperformed other Machine Learning models such as DT, SVM, KNN, and LGBM.

Keywords: - Decision-Tree; Knearest neighbors; Light Gradient Boosting Machine; Support Vector Machine; Convolutional Neural Network.

1. Introduction

Human skin is one part of the most significant features of the body. Different styles of skin diseases will infect the skin readily and quickly. Skin disease is a major health issue that affects people all everywhere in the globe. Like this disease is large number of people affected nowadays. Manual diagnosis is difficult and time-consuming for doctors. The datasets are essentially divided into seven categories and also this is one of the imbalanced datasets . Because of the nature of data availability, datasets are inequitable. As a result, various methods are used to classify this data and investigate dataset issues. Due to the quantity and nature of datasets, traditional technology is time consuming, power intensive, and complex.

The objective of this project is to propose a skin disease image classification. It involves two processes: (i) Feature Extraction and (ii) Classification. The ability to extract feature vectors from the term "segmented image" is used to describe a image that has been divided into sections. The dataset collected from the internet and collected image divided into testing and training set. Texture feature (GLCM,LBP) and color feature (Color histogram,color space) are taken from the image of the lesion. The training data are helps to train the classifier using various type of models. Machine learning and deep learning have recently proven to be the most effective methods for classifying and detecting diverse objects in images. From dermatological images, which were used to categorize many types of skin problems. Several methods, including Deep learning and Machine learning are two types of artificial intelligence, have been employed to help in the classification and diagnosis of skin problems.

The traditional Machine Learning is a subcategory of Artificial Intelligence (AI), in which a machine may learn without needing to plan ahead of time. In this technology world, Machine Learning is finding applications in numerous sectors, with clinical diagnostics being the most recent addition[1]. Developing ML-based diagnostic tools is challenging, and choosing the correct decision-making algorithm to improve diagnostic accuracy is critical[1][2]. As a result, selecting the suitable feature and machine learning algorithm to obtain the best diagnosis accuracy is equally critical.

In any classification task requiring machine learning algorithms, the input feature chosen is crucial. We noticed that researchers employed various types of features in earlier studies, depending on the categorization goals. Color and Texture Features were used by the majority of the studies. The most prominent features used to visually define and identify skin diseases are color and texture information. Image information is critical for distinguishing one disease from another. A variety of techniques can be utilized extract these color properties, including Histograms include color histograms, color correlograms, and color descriptors, among others. [11, 19].

The texture information provides both the difficult visual patterns of skin lesions and spatially structured things like brightness, color, shape, and size. Image texture is largely determined by pixel intensity variation. To extract texture information from images, most researchers use approaches like GLCM and SIFT [3].

Deep Learning is another major learning technique that employs deep convolutional neural networks, which are simply neural networks with numerous hidden layers. In any neural network, there are three types of layers. Before extracting features from data and generating future predictions about new data, processing units output an input layer, also known as a hidden layer. Deep Learning refers to the process's hidden levels. In testing how appropriate the specific features are in blocking the modeling features and the common type item, instructions for learning from the data are specified by updating the probability weight assigned to the feature nodes in updating the system's performance is measured using a test set. The proposed work classification of skin disease is shown below Fig. 1.

1.1. *Contribution of this study*

- Automated dermoscopic image based system is proposed that efficiently classifies various type of skin disease images.

- Feature extraction was performed with several feature descriptors such as texture and color features each and every features having two types.

- Feature performance was evaluated with five types of classification methods.

- Proposed method CNN achieves best performance of 84% accuracy obtained when compared to various Machine Learning models.

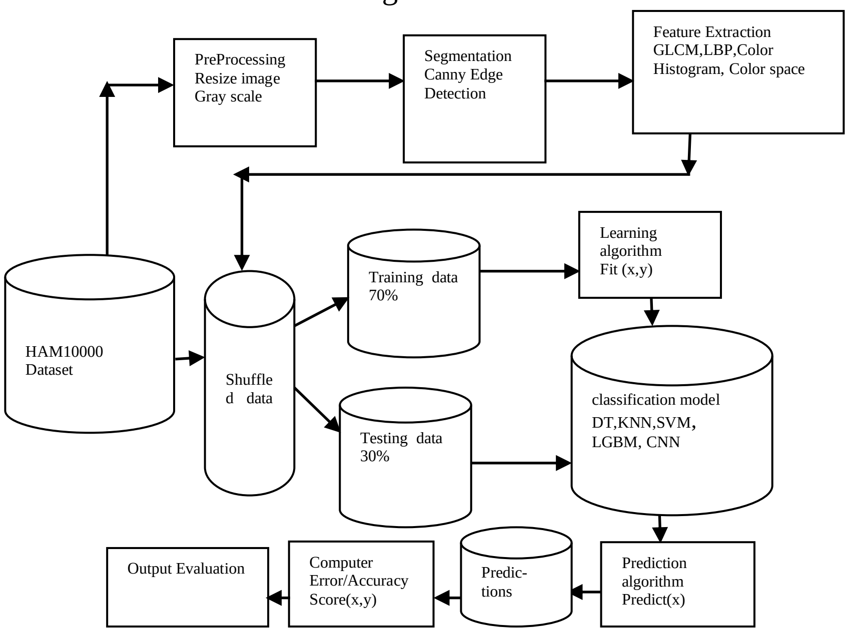


Fig. 1 Proposed work flow diagram

1.2 Organization

The procedure is organized as follows: section2 is review of previous work Dataset Collection detailed in section 3. Process of Pre-Processing the image. Section 4. Feature Extraction algorithm used in Section 5. In section 6 is Classification model used. After that Section 7 is demonstrate Experiments and results. This proposed work concluded by giving a summary of process and the feature work.

2. Litrature Review

Several image based on machine learning there have been experiments carried throughout the years to classify skin disease methods.. A survey was conducted prior to the start of this study to acquire data on skin disease diagnosis using a machine learning approach. Skin disease is a form of skin cancer that results in the formation of a malignant tumor on the skin. Skin cancer is identified using dermatological imaging cutaneous skin cancer. Skin cancer was detected using traditional machine learning based high-performance image, and the detection rate was high.

However, by extracting more features, the model's accuracy may be improved, and the model's sensitivity can be raised[1]. image processing processes to improve skin cancer diagnosis accuracy. However, they were unable to describe a precise model capable of accurately detecting cancer. Skin cancer diagnosis using a DL algorithm-based architecturedriven model Because model-driven architecture DL is so quick to develop, the model can virtually instantly predict the outcome. It had a greater detection rate for skin cancer[2]. To diagnose and classify skin cancer, the author suggested a lesion indexing network (LIN) based on DL. They were able to achieve decent results utilizing DL-based LIN by extracting additional features. However, segmentation performance must increase in order to improve the outcomes even further[4].

The author employed CNN to detect skin cancer from pigmented melanocytic lesions based on dermoscopic images. Skin tumors that were non-melanocytic and non-pigmented, on the other hand, were difficult to detect. It also got a decreased detection accuracy[5]. To detect skin lesions, DCNN employs three phases: first, colour modification improves contrast; second, a CNN approach extracts lesion borders; and third, transfer learning extracts deep characteristics. However, the beneficial thing good outcomes for some datasets, albeit results may vary per dataset. [26].

In a model based on CNN for detecting melanoma skin cancer, pre and post image processing were implemented for image development. By incorporating local and global contextual information, the model generated lesion regions. It achieved a decent prediction and classification performance. The execution time, is not provided, on the other hand, which could raise the value of the results[25].

The ResNet50 transfer learning model, which reduces the requirement for preprocessing steps and enhances manual selection accuracy. The classification reports are insufficient, but the results could be improved if the preprocessing phases are followed. One of the most well-known articles proposes using CNN to recognize and classify malignant growth. They claim that the accuracy prediction of their suggested CNN is superior to that of most neural systems. With a 96 percent accuracy rate, CNN can be a reasonable alternative for malignancy grouping classification, although the dataset for this method was quite limited at 1000 images[8].

To categorize melanoma and nevus, only 399 images were used. In this investigation, It is utilized a pretained deep neural network (DNN). This dataset is unnecessarily small for a system that is expected to classify sensitive client information. 92.1 % sensitivity, 95.18% specificity, and 93.64% accuracy are achieved with this methodology [26].

The images used are mostly from the 2624 Universal Skin Imaging Coordinated Effort (ISIC). They apply transfer karnmy, sparse coding, deep residual scheme, and fully convolutional U-organization using AlexNet. A support vector machine is used to classify the extracted factures after they have been extracted using transfer learning. Melanoma and non-melanoma is divided into 93.1% sensitivity and 92.8% specificity. 73.9% accuracy, 73.8% sensitivity, and 74.3% specificity were obtained for more challenging segmentation between melanomas and atypical nevi [10].

Using AlexNet (transfer learning) extract features, create a multi-class classifier with ten labels. The researchers used 1300 images of 10 skin lesions and got an 81.8 % accuracy rate[11]. To assess the performance of CNN using detailed instructions on how to produce CNN using 12,378 open-source dermoscopic images, researchers used detailed instructions on how to prepare CNN. 100 images were used by 157 dermatologists from 12 different German college emergency hospitals. Standard sensitivity and specificity with dermoscopic images (which were the evaluation measurements used in their publication) were 74.1 % and 91.3 %, respectively, among dermatologists. [12].

Using CNN, investigate the utility of a programmed characterisation technique in the treatment of skin lesions. On 956 clinical images, the system achieved a score of 96 percent for Melanoma and 91 percent for Basal Cell Carcinoma [13]. Below Table 1 shows that HAM10000 MNIST dataset details:

S.No.	Disease type	Number of disease image
1	Nv	6705
2	Mel	1113
3	Bkl	1099
4	Bcc	514
5	Akiec	327
6	Vasc	142
7	Df	115

Tabel 1. Details of Dataset

3. Pre-Processing

This work discusses the phase of the proposed research, which includes proposed image enhancement techniques, filtration with a median filter, proposed pixel-wise interpolation and enhanced dual radon conversion based hair removal, including adaptive / global threshold and adaptive division processes. Pre-Processing is development of image enhancement, which eliminates and reduces unwanted noise and improves image quality, which is important for image processing. The Pre-Processing work involves three main types : image resize ,Gray scale conversion , Noise removal for image enhancement.

3.1 Resize the image

The Image resizing is one of the dimensions of images. Many image processing and Machine Learning techniques make scaling simple. This helps to decreases the pixels size in an image while also shortening the processing and computation time.

E.g. This can speed up the training model since the many size amount of pixels in a image at a given time, the more difficult it is for the model to solve the problem. Skin lesion original image size 600X400 pixel dimensions. So resized all the images are 224X224 dimensions. The image given as the input may not have the original size as required by the algorithm, so that the required image size is obtained.

3.2 Grayscale Conversion

There is brightness information in a grayscale image. Each pixel in this position image represents the amount or amount of length. The brightness gradient can be seen in a gray level scaling image, only the intensity of the light is measured image brightness is[0-255], '255' white and '0' black. Grayscale conversion converting a color image to grayscale. color images are very difficult to process and take a long time to create. Color(or RGB) image changed to grayscale.

3.3 Noise Removal

The process of detecting and removing unwanted noise from a digital image is called noise removal. It is difficult to determine which aspects of the image real and the effect of the noise. In our proposed system we are using median filter to eliminates unwanted noise. The mainly significant system for elimination of blur in images. It has the wide-ranging of restoration for finding the noisy image. Median filter is nonlinear filter, it leaves edges invariant. Median filteration is optimal because it reduces the average square error. Not only that, but it also reduces the overall average square error during reverse filteration and noise smoothing processes.

3.4 Segmentation

This feature is used to particular segmented ROI extract from the Pre-Processed image. It is a popular segmentation edge detection algorithm. This method approach with a threshold value of 0 was utilized for edge detection. This function when used together with the color feature provides excellent efficiency. The boundaries and inner areas of diseased cells are described by combination of edge and color features. As a result of the expert system, the retrieved features were stored in the database.

3.3.1 Edge detection procedure :

This segmentation process can be used to reduce the quantity of data value in an image and at the same time preserve the image structural properties for subsequent processing. This strategy has been employed using the following criteria in this proposed method.

Detection: the chance of correctly identifying true edge points should be increased. While the chance of misidentifying non-edge points should be minimized. It has the same effect as enhancing the noise ratio.

Localization: The edges have been identified possible to the actual edges.

Number of responses: There should be no more than one detected edge from the actual edge.

4. Feature Extraction

Feature extraction helps to remove the unwanted areas. It's may be feature extracted from texure, shape, color etc. and also get new ROI from the original image.

4.1 Texture Gray Level Co-occurrence Matrix (GLCM)

this is the most used tool for analysing lesion texture. At the same moment, we consider two pixels(or) Points, one reference and the other neighbour pixel. Before computing the GLCM, we establish a certain spatial connection between the reference and neighbour pixel. This work makes use of some glcm characteristics.

- ASM
- Entropy
- Energy
- Max probability

Angular Second Moment(ASM) : To calculate the homogeneity of the image.

$$= \sum_{l,m=0}^{l-1} p^{2}{}_{l,m} \quad (1)$$

Entropy: To measure grayscale randomness of intensity that exist in the lesion area.

$$c_e = -\sum_l \sum_m p_d[l,m]/np_d[l,m] \quad (2)$$

Energy: the glcm is the average of the squared elements. Homogeneity, commonly known as the angular second moment,

$$= \sqrt{ASM} \quad (3)$$

Maximum probability: It's refer to the biggest opening in the matrix, and according to the strongest
reaction. This may be the maximum in any of the matrices.

$$c_m = max_{l,m} pd[l,m] \quad (4)$$

4.2 Local Binary Patteren (LBP)

LBP is a widely used technique for face representation and classification on images. The most common method is to process each and every 3x3 window in the image to extract an LBP code. The processing implies using the mean and median, or the centre pixel, as thresholds to compare the centre pixel of that window to its surrounding pixels.

i_0	**i_1**	**i_2**
i_7	**i_c**	**i_3**
i_6	**i_5**	**i_4**

There are 36 unique LBPs for 256-gray-level image. So we have 36 features for this texture, we can represent texture as 36- dimensional point is 36- dimensional space.

4.3 Color Feature Extraction

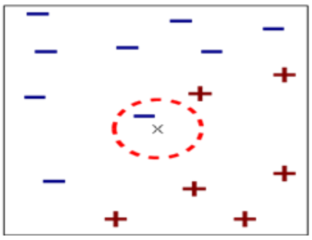
Color Feature-vector is considered to be another important feature for the classification of skin image analysis. When a specific area is affected the area of skin lesions effectively changes color. To identify and classify skin image corresponding color maps are created at different color intervals. RGB,LAB,HSV,HUE and OPP color spaces are used to create the 3-D Histogram[12].

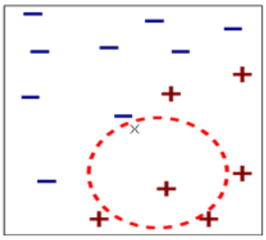
Color space is a computational model that includes color information by using three or four different color elements. Color utilized in a variety of applications in digital processing. image analysis, television broadcasting, and computer vision. For skin disease classification and identification, a variety of colour spaces are available. RGB-based colour features (RGB, normalised RGB), Hue-based color features (HSI, HSV, and HSL), and Luminance-based colour features (YCBCr, YIQ, and YUV)[5]. These models are further explained in the sections that follow. Color space selection is the most important step in skin colour modelling and classification.

5. Classificaion

5.1 K-Nearest Neighbor

Nearest Neighbors is segment of the most commonly used supervised Machine Learning approaches for classifying some set of images. It assigns a classification to a data point depending on how its neighbours behave. KNN keeps track of all available examples and characteristics new ones using similarity metrics. The number of adjacent neighbours who must be integrated in the majority voting procedure is denoted by the constraint K in KNN. A data point information is categorized by a majority of votes based on the value of neighbours. The KNN algorithm is based on similarity of features. The process of determining the correct value k is a process of parameter adjustment, which is critical for improved accuracy. Three different kinds of Knearest neighbors are demonstrate given below such as Fig. 2a., Fig.2b., Fig.2c.





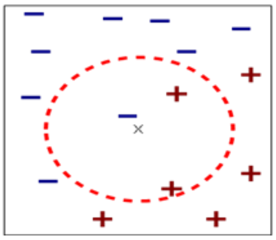


Fig.2a. one-knearest neighbor ***Fig. 2b. Two-knearest neighbor***

Fig. 2c. Three-knearest neighbor

The above process of a record x are data points with the smallest k distance to **x** Some frequently used distance functions.

Hamming Distance: $d(x,y) = \sum_{i=1}^{m} \left| \frac{x_i - y_i}{x_i + y_i} \right|$ (5)

Minkowsky Distance: $d(x,y) = [\sum_{i=1}^{m} |x_i - y_i|^r]^{\frac{1}{r}}$ (6)

Chebychev Distance: $d(x,y) = max_{i=1}^{m} |x_i - y_i|$ (7)

Euclidean Distance:

$$d\ (x,y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \quad (8)$$

Manhattan Distance

$$d(x,y) = \sum_{i=1}^{m} |x_i - y_i| \quad (9)$$

5.2 Decision Tree(DT)

This learning moves from interpretation about an object (described in the branches) to conclusions using a choice (as a predictive model). The object's target value (Indicated on the leaves). Decision Trees are a type of tree model classification that can take a single value for the target variable. Regression trees are result trees in which the target variables can take consecutive values (Typical values are real numbers). There are various distinctdecision-makingprocess.

.

- Random forest(Ensemble modeling is a combination of many trees)Based on Gini gain

- CHAID perfoms multi level splits when computing classification trees – Based on chi-square concept. Continuous independent variables should binned into finite number of bins to create categories.

5.3 Support Vector Machine

For a linearly separable binary set, use SVM. Assume we have two features, x1 and x2, and we wish to categorise all of these items. The model of the svm is to create a hyperplane; here in this case, the hyperplane is this green line. The objective is to create a hyper plane that divides all training vectors into two groups. Two types of Hyperplane graph representation in Fig(3a), Fig(3b)

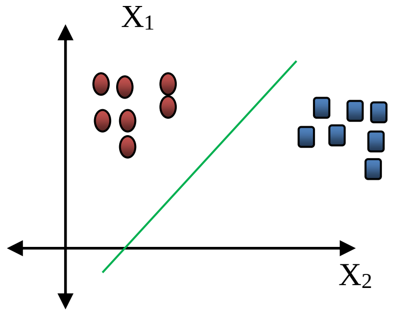


Fig. 3(a). Hyperplane1

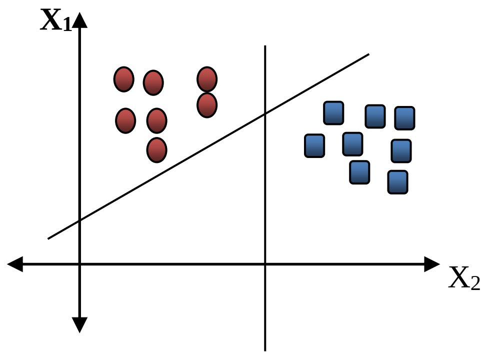


Fig. 3(b). Hyperplane2

We show here two different hyperplanes that can accurately categorize all the events in this feature set but the best choice is the hyper plane is greatest option which leaves the maximum margin from both.

$$g(\vec{x}) \geq 1, \forall \vec{x} \in class1$$
$$g(\vec{x}) \leq -1, \forall \vec{x} \in class2$$
$$z = \frac{|g(\vec{x})|}{||\vec{w}||} = \frac{1}{||\vec{w}||} \quad (10)$$

The total margin is computed by

$$\frac{1}{||\vec{w}||} + \frac{1}{||\vec{w}||} = \frac{2}{||\vec{w}||} \quad (11)$$

The goal is to maximize separability by lowering this term. Minimizing w is a problem of nonlinear optimization that can be solved using langrange multipliers and the karush-kuhn-Tulker(KKT) criteria.

$$\vec{w} = \sum_{i=0}^{N} \lambda_i \, y_i \vec{x_i}$$
$$\sum_{i=0}^{N} \lambda_i \, y_i = 0 \quad (12)$$

Omega will be the outcome of this addition, and we will also have this additional regulation. The dividing line between the two groups that will enhance their separability.

5.4 Light Gradient Boosting Machine (LGBM)

The LGBM (Light Gradient Boosting Machine) is an extremely effective gradient boosting machine. method and library that came out in 2017. Main Contributions is more focus on under-trained data points. Exclusive feature Bundling (EFB) : Efficient representation for sparse features such as one-hot-encoded features. Gradient –based one-side sampling Idea:

- we can put more focus(importance) on under – trained.
- data points when building our trees.
- Small gradients: means small error, the data point is learned well (Not important).
- Large gradients : means large error, the data points is not learned well (important) Sort the data according to gradient value.
- Keep the top ax 100% of the data samples(large gradients).
- Randomly sample bx100% from rest of the data(small gradients).
- Amplify small gradients by multiplyting $\frac{1-d}{b}$ when calculating information gain.

Exclusive feature combilation bundling(EFB):

A nice way to reduce number of features by talking advantage of sparsity of large datasets. The main observation is many characteristics are never non-zero. at the same time. Therefore , we can merge(bundle) them and save space by doing that below example,

Feature1	**Feature2**	**Feature3**
0	**1**	**2+1**
0	**0**	**0**
1	**3**	**0**
0	**0**	**3+1**
5	**0**	**0**
0	**0**	**7+1**

We cannot merge feature1 and feature2 there is overlap. We will add+1 to the non-zero feature-3 rows so that feature3 and feature1 have different ranges. No overlap between feature 1 and feature 3. We can merge them.

Feature1+Feature2	Feature3
3	1
0	0
1	3
4	0
2	0
8	0

sample query in a bundle: If the value is larger than 2. It refers to feature3

5.6 *Convolutional Neural Network*

It is a part of machine learning and its uses neural networks, to imitate human-like decision-making, neural networks were used.. CNN learning algorithm uses these networks will have three different kinds of layers such as an input neurons, a hidden units, and an output neurons of preprocessing units to extract features from data and make predictions about new data. Deep learning is hidden layers of processing and learning algorithms refers to gain knowledge from data by modeling features and updating probability weight assigned to attribute nodes in testing how appropriate particular features are in determining the common kinds of item. Below diagram Fig. 4. represent working process of convolutional Neural Network system.

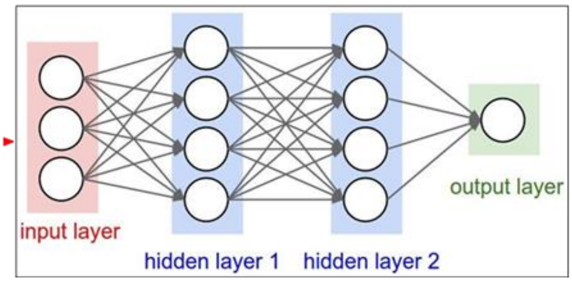


Fig. 4. CNN System Architecture

6. Experimental Results and Discussions

The train data and validation data are pre-processed in this step. The purpose of the preprocessing data procedure is to normalize the size of each and every set of data or image. We used dermoscopic images from the MNIST HAM10000 dataset in this research, with each image having a dimension of 600x450 pixel values in RGB format. All neurons had their images scaled to 224x224 pixels. The RGB levels are converted to a scale of -1 to 1.

Segmentation process is very important to any other Machine Learning applications for needed region extracted from the original image . This process helps to feature extraction techniques. The Fig. 5 shown below for sample segmented results.

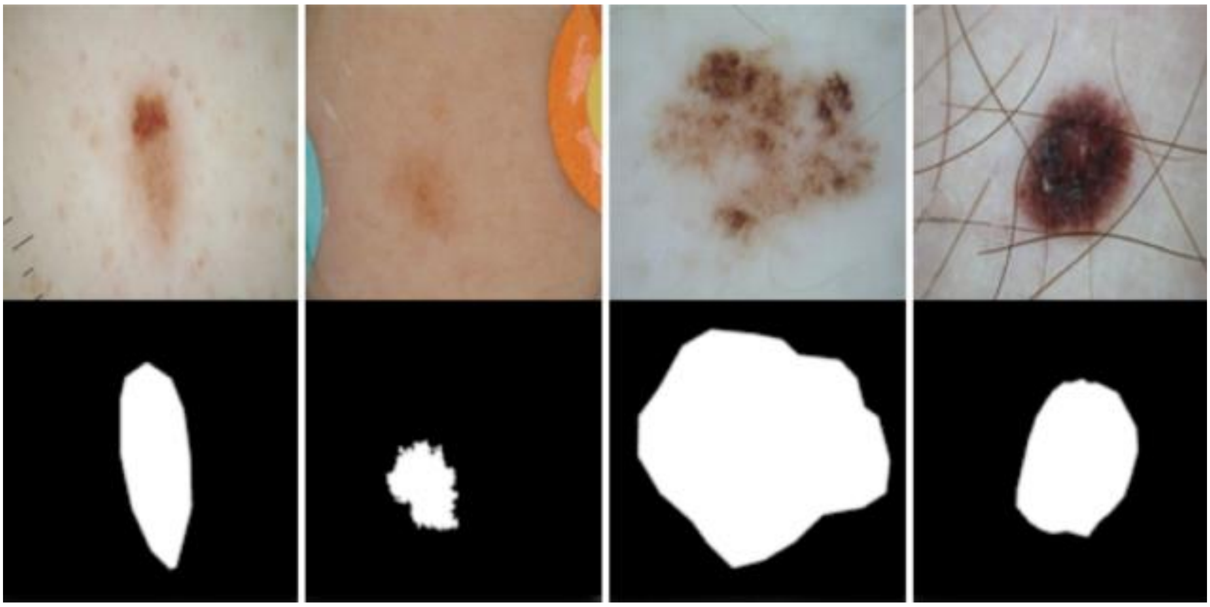


Fig. 5 Sample segmented results

The pre-trained CNN model was retrained using the training model. The training procedure is carried out in the Jupyter editor Notebook environment, which is based on cloud computing. The following will be explained in the methodology of training procedure desire to resolve the batch size, training, and epochs: The package size is determined by the number of data trains to be distributed each iteration of the neural network. The epoch size is set to 20 at this time, that is, each step sends 20data to the neural network. The number of sample data in that formula represents the initial size of the data train before the data multiplication process.

A group of sets training steps is referred to as an era. For some epochs, the Neural Network required training and errors were found to be near to zero. In this study, the epoch was employed 30 times in the training model procedure. The CNN architecture's layer values were retrieved, followed by the feature values of the remaining four layers, and supplied as input to the models.

The Dropout layer and the Dense layer are the next two layers to be added. To avoid the completely linked layer from over-fitting, the exit layer is introduced. For operation, a dense layer is used. Softmax activation is a feature of the dense layer.

To meet the categorization requirements, fine tuning is the process of removing specific layers and adding new layers to the model. The initial CNN sample only had one layer. The entire layer freezes, except for the 23 layers and the weight value. These are the layers that will be combined to form the train layer. After that two layers were added to the dropout to prevent over-fitting. The first dense layers was added using the softmax activation function to clarify the final output value of the first node to 128. Second, using softmax and set the output value of the second node to 7 neurons. About the Table 2 below explain the parameter of the Convolutional Neural Network layers.

Convolution layer	Kernel size: 3 X 3
	No. of kernels : 16
	Sride: 1x1
	Activation function1: ReLU
Pooling layer	Pooling method: Max pooling1
	Filter size: 2 x 2
	Stride1: 1 x 1 matrix, padding1: "same"
Fully Connected layer	Activation function1: Softmax
Others	Model architecture1: Sequential
	Optimizer1: Adam
	Loss function1 : Categorical_crossentropy1
	Epochs: 30

Table 2. Perfomance of Convolutional Neural Network

6.1 Comparision of performance measures for skin disease classification

the report analyzes for image classification with convolutional neural network and compares them to various Machine Learning models. This paper various algorithm used for skin disease classification. Deep Learning models CNN is more accurate than the traditional Machine Learning . It has an 84% accuracy. The first stage training data is approximately 70%. The training data set will undoubtedly improve this. Seven skin diseases were primary tested here, with the possibility of expanding further in the future. A large data set can increase the accuracy level. The table 3 signified that the presented CNN model has attained better classification performance over the compared with four type of Machine Learning classification models. Firstly, the CNN model has obtained effective classification with the precision of 84%, recall of 84% and F1-score of 84%. Figure 6 offers a comprehensive results analysis with the recent state of art methods such as GLCM. From this study, feature extraction and proper segmentation technique are very important for any machine learning based applications, because any clinical image can be categorized very accurately based on the segmented image and their feature values.

	precision	recall	f1-score	support
0	0.69	0.55	0.61	33
1	0.68	0.71	0.69	51
2	0.70	0.65	0.68	110
3	0.71	0.83	0.77	12
4	0.91	0.94	0.93	671
5	0.66	0.57	0.61	111
6	1.00	0.93	0.96	14
accuracy			0.84	1002
macro avg	0.76	0.74	0.75	1002
weighted avg	0.84	0.84	0.84	1002

Table 3. Classification Report for CNN

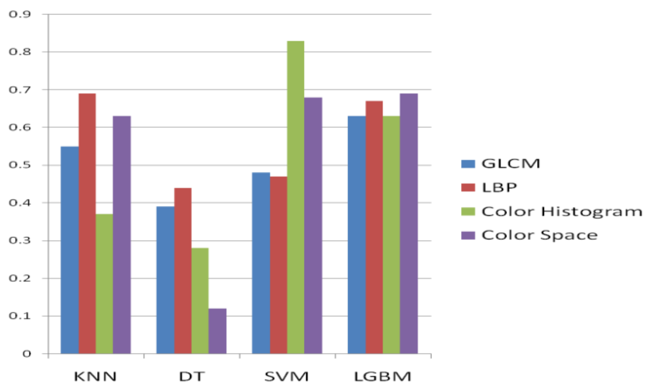


Fig 6. Comparision of Machine Learning algorithms with their Feature Extraction using Bar diagram

6.2 Sample Classification report for GLCM with Machine Learning algorithms

(7809, 10000)
(1949, 10000)

	precision	recall	f1-score	support
0	0.03	0.03	0.03	65
1	0.03	0.02	0.02	102
2	0.19	0.20	0.20	219
4	0.70	0.81	0.75	1341
5	0.12	0.02	0.04	222
accuracy			0.58	1949
macro avg	0.21	0.22	0.21	1949
weighted avg	0.52	0.58	0.54	1949

Classification Report for DT

	precision	recall	f1-score	support
0	0.04	0.06	0.05	65
1	0.04	0.07	0.05	102
2	0.14	0.21	0.17	219
4	0.69	0.50	0.58	1341
5	0.12	0.20	0.15	222
accuracy			0.39	1949
macro avg	0.21	0.21	0.20	1949
weighted avg	0.51	0.39	0.44	1949

0.3930220625962032

Classification Report for KNN

	precision	recall	f1-score	support
0	0.06	0.15	0.09	65
1	0.08	0.14	0.10	102
2	0.10	0.11	0.10	219
4	0.71	0.65	0.68	1341
5	0.13	0.09	0.11	222
accuracy			0.48	1949
macro avg	0.22	0.23	0.22	1949
weighted avg	0.52	0.48	0.50	1949

Classification Report for SVM

	precision	recall	f1-score	support
0	0.12	0.02	0.03	65
1	0.00	0.00	0.00	102
2	0.17	0.18	0.18	219
4	0.74	0.86	0.79	1341
5	0.23	0.14	0.18	222
accuracy			0.63	1949
macro avg	0.25	0.24	0.24	1949
eighted avg	0.56	0.63	0.59	1949

Classification Report for LGBM

7. Conclusion and Future work

This work is able to classify 7 type of skin diseases efficiently. The CNNs output accuracy 84% gives better than Machine Learning accuracy. Convolutional Neural Network feature on their own. This study used large amount of dataset, so a large dataset can increase with advanced right computational algorithms to gives better accuracy. This system helps to match the accuracy of skin disease image thus enhancing the better quality specification in the medical and research fields. In next, future work implementation by using advanced techniques and small amount of data with suitable segmentation process will be use for increasing Machine Learning accuracy.

References

1. Sourav kumar patnaik,Mansher Singh Sidhu"Automated Skin Disease Identification using Deep Learning Algorithm", Biomedical & Pharmacology Journal, Vol. 11(3), p.1429-1436 September 2018.
2. Jainesh Rathod, Vishal Waghmode, "Diagnosis of skin diseases using Convolutional Neural Networks", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA 2018).
3. Nazia Hameed a.b.* , Antesar M .Shabut c "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques" , Expert systems with Applications , September 2019.
4. Yeong Chan Lee1, Sang-Hyuk Jung1,"WonDerM: Skin Lesion Classification with Fine-tuned Neural Networks Yeong ", Samsung Advanced Institute of Health Sciences and Technology, 2018
5. Ms. Seema Kolkur 1, Dr. D.R. Kalbande 2,"Machine Learning Approaches to Multi-Class Human Skin Disease Detection", International journal of Computational intelligence Research",vol 14,PP.29-39 ,2018.
6. Kouheri Shimizu*, Hitoshi Iyatomi , "four classification of skin lesions with task Decomposition strategy",IEEE transaction on Bio-medical engineering, vol 62,jan-2015
7. Ramon.c,Maron,"Systematic outperformance of 112 dermatologist in multiclass skin cancer image classification by convolutional neural networks", European journal of Cancer, vol 119,September 2019, pages 57-65.
8. Archana Ajith, Vrinda Goe,"Digital Dermatology", International conference on Intelligent computing and control system ICICCS 2017.
9. Silvio M. Pereira, Macro A.C.Frade,"Classification of color images of Dermatological Ulcers", IEEE journal of Bio-medical and Health informatics, vol-17.
10. V. B. Kumar, S. S. Kumar, and V. Saboo, "Dermatological disease detection using image processing and machine learning," in Artificial Intelligence and Pattern Recognition (AIPR), International Conference on. IEEE, 2016, pp. 1–6.
11. R. Yasir, M. A. Rahman, and N. Ahmed, "Dermatological diseasedetection using image processing and artificial neural network," in 8th International Conference on Electrical and Computer Engineering:Advancing Technology for a Better Tomorrow, ICECE. Pan Pacific Sonarga on Dhaka, 2014, pp. 687–690.
12. A. Amarathunga, E. Ellawala, G. Abeysekara, and C. Amalraj, "Expert system for diagnosis of skin diseases," International Journal of Scientific & Technology Research, vol. 4, no. 01, pp. 174–178, 2015.
13. M. A. Sheha, M. S. Mabrouk, and A. Sharawy, "Automatic detection of melanoma skin cancer using texture analysis," International Journal of Computer Applications, vol. 42, no. 20, pp. 22–26, 2012.
14. M. S. Arifin, M. G. Kibria, A. Firoze, M. A. Amini, and H. Yan,"Dermatological disease diagnosis using color-skin images," in Machine Learning and Cybernetics (ICMLC), 2012 International Conference on, vol. 5. IEEE, 2012, pp. 1675–1680.

15. I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, “Med-node: a computer-assisted melanoma diagnosis system using nondermoscopic images,” Expert systems with applications, vol. 42, no. 19, pp. 6578–6585, 2015.
16. V. K. Shrivastava, N. D. Londhe, R. S. Sonawane, and J. S. Suri, “Reliable and accurate psoriasis disease classification in dermatology images using comprehensive feature space in machine learning paradigm,” Expert Systems with Applications, vol. 42, no. 15-16, pp. 6184–6195, 2015.
17. Y. W. M. Yusof, H. Hashim, K. A. Sulaiman, S. Subahir, N. E. Abdullah, and F. N. Osman, “Plaque lesion classification fuzzy model based on various color models.”
18. A. F. Jerant, J. T. Johnson, C. Sheridan and T. J. Caffrey, “Early detection and treatment of skin cancer,” *American family physician*, no. 2, pp. 357-386, 2000.
19. K. Korotkov and R. Garcia, “Computerized analysis of pigmented skin lesions: A review,” Artificial intel. in medicine, pp. 69-90, 2012.
20. M. H. Jafari, S. Samavi, S. M. R. Soroushmehr, H. Mohaghegh, N. Karimi, K. Najarian, Set of Descriptors for Skin Cancer Diagnosis Using Non-Dermoscopic Color Images,” Image Processing (ICIP), 2016 IEEE International Conference on, Arizona, 2016.
21. V. B. Kumar, S. S. Kumar, and V. Saboo, “Dermatological disease detection using image processing and machine learning,” in Artificial Intelligence and Pattern Recognition (AIPR), International Conference on. IEEE, 2016, pp. 1–6.
22. A. R. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari and F. Kolahdouzan, “Segmentation of Dermoscopy Images Using Wavelet Networks,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1134- 1141, 2013.
23. Sourav kumar patnaik,Mansher Singh Sidhu”Automated Skin Disease Identification using Deep Learning Algorithm”, Biomedical & Pharmacology Journal, Vol. 11(3), p.1429-1436 September 2018.
24. Nazia Hameed a.b.* , Antesar M .Shabut c “Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques” , Expert systems with Applications , September 2019.
25. Ankit Chopade, “Seven Class Classification of Skin Lesions by using Incremental Convolutional Neural Network in Python”, Proceedings of the International Conference of Advances in Electronics, Electrical & Computational Intelligence(ICAEEC)2019.
26. [26]. [Md ShahinAli, Md SiponMiah, JahurullHaque Md Mahbubur Rahman, Md Khairullslam, “An Enhanced technique of skin cancer classification using Deep Convolutional Neural Network with transfer learning models”. Machine Learning with Applications Volume 5, 15 September 2021, 100036
27. [27]. Nabeela Kausar, Abdul Hameed “Multiclass skin cancer classification using ensemble of fine-tuned deep learning models” by Applied sciences Volume 11 issue 2.