Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, No. 2, December : 2023

AUTOMATIC IDENTIFICATION OF SKIN DISEASE USING KNN

Ms. Hiral S. Padhariya, Teaching Assistant, Dept. Of Information And Communication Technology, Veer Narmad South Gujarat University, Surat.
Dr. Shailesh Chaudhari, Associate Professor, Dept. Of ICT, VNSGU, Surat.

Abstract
Skin cancer is deadly form of cancer caused in almost entire world. Mostly skin cancer is caused by Ultraviolet rays from the sun which damage the DNA in the skin cells. Some of the artificial sources which can also be cause of skin cancer are sunlamps and tanning beds. As skin cancer is second death causing diseases in the entire world and it can only be cure at the initial stage, so it is recommended to detect skin cancer at the early stage. In this paper, we propose a model for classification of skin diseases like Actinic Keratosis(akiec), Basal Cell Carcinoma(bcc), BenignKeratosis-likelesions(bkl), Dermatofibroma(df), Melanoma(mel), Melanocytic Nevi(nv) and Vascular lesions(vasc) for HAM 10000 dataset with the use of most efficient machine learning techniques as K-Nearest Neighbor(KNN) classifier are applied by using extracted features for different classification of skin images. KNN is used because it gives us efficient and high accurate results. The experiment results have shown that the proposed algorithm is very promising in the classification of skin diseases.
Keywords: Machine Learning, Image Processing, Supervised Learning, K-Nearest Neighbor.

I. INTRODUCTION

Skin disease is considered to be most common type of disease which is being faced by many of people considering of all age groups. Due to some kind of bacteria, fungal infection, viruses etc.. skin disease may happen, sometimes due to such type of infection it may lead to skin cancer. Therefore, it is important to diagnosed disease at the early stage. Treatment of such a disease may take time and causes more financial cost to the patient. Also, there are many such other types of skin disease which may be caused due to pollution or sunburn. About 70% of people across the various countries of world are suffering from various ski problems in their day-to-day life. Talking about the death ratio over 10 million of the people die due to low income [21].

In most of the cases patient doesn't know the type of disease and the stage of that disease some skin disease shows the symptoms several months later which causes more damage to the skin.one of the reason behind this is due to lack of the public awareness regarding such medical illness. Also, the test done for identifying the disease are very expensive. The tools which are required for early detection of skin disease are still not available globally in the various countries. Some advancements have been done in medical which make possible to diagnose disease with least time and more accuracy.

Some of the common skin disease in south Africa are acne, eczema, fungal infection and skin tumors. As with advancements in the technology using laser and photonics-based technology skin disease are identified, but it is too costly and also takes a long time for getting results. Between this period disease may spread more and can cause damage to the skin cells. So, there should be ML techniques use to overcome such problems [22].

Machine learning, which is a branch of AI is being used to develop better diagnostic tools in healthcare sector. ML algorithms are applied with huge amount of data for detection of various skin disease providing better accuracy as compared to the traditional approach. ML models learn from various data called as training data without providing any explicit instructions. Due to the tremendous progress in the ML field algorithms such as KNN, decision-tree, Random Forest, clustering they provide a solution for such an existing problem [23].

As techniques of Image Processing and ML algorithms plays a vital role in for the classification of skin disease, many researchers have proposed various model for identifying the disease but the accuracy found is not convenient also these models have been able for detecting only for one or two different types of skin diseases.

This research paper proposes KNN supervised machine learning algorithms which is able to identify various skin disease like actinic keratosis, basal cell carcinoma, benign keratosis like lesions, dermatofibroma, melanoma, melanocytic nevi and vascular lesions which uses HAM10000 dataset. HAM10000 is as dataset which consists 10015 training images for the detection of various skin lesions. It consists of collection of categories such as Actinic Keratoses (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (vasc), Also for feature extraction some of the image processing techniques are applied such as edge detection, gray scale, normalization of images etc..[24].

II. Literature Review

This section presents the survey of the various related work done by the various researcher. In [1] the author presents the system which is limited to classify three different types of skin disease. The data processing includes image acquisition, preprocessing of noise removal, feature extraction and noise removal. Here five classification algorithms were used namely Naïve Bayes, AdaBoost, BayesNet, J48 and MLP. In [2] the author proposed a system for image-based techniques without need of doctor. The system contains 2 stages. In First stage disease is detected by using color image processing technique. For identification of the skin disease K-mean clustering algorithm have been used and second stage consists of identification of the skin disease using artificial neural network. For first stage average accuracy is 95.99% and for second stage average accuracy is 94.016%. In [3] the author proposed system which uses feature extraction method for detection of skin disease and accuracy lies upto 90%. In [4] the author has been focused on the various segmentation technique for detection of melanoma disease using image processing techniques. In [5] the work proposed by the author develop the tool for detecting melanoma disease using specialized algorithm. In [6] the author proposed classification of skin disease such as Melanoma, Basal cell carcinoma, Nevus and Seborrheic keratosis using Support Vector Machine (SVM) which gives best accuracy from other technique. In [7] the author proposed a system that automatically identifies and detects eczema skin disease using Support Vector Machine (SVM). In [8] the author proposed new approach for detecting skin disease using computer vision and machine learning as combination which is tested for six different skin disease and accuracy is 95%. In [9] the author has used the image analysis technique for noise removal and feature extraction and for removing noise image is fed to the classifier for further processing. In [10,11,12] the author has used the various machine learning algorithms, also image processing technique have used for pre-processing task. In [13] the author creates an artificial neural network for predicting skin disease with accuracy 90%. In [14] the author proposed system which used various different pre-processing algorithm and feed forward back propagation artificial neural network is used for training and testing purpose. In [15] the author used matlab tools for image processing. In [16] the author proposed automatic skin detection and recognition with accuracy 98%. In [17,18,19] the author compared various different features for different types of algorithms. In [20] the author uses and elaborates back propagation neural network which provides accuracy rate upto 91.2%.

III. Methodology

The Methodology of the proposed system for classification of skin disease is described. It will be able for detection of different types of skin disease such as Actinic Keratoses (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (vasc). The entire block diagram is represented as shown in Figure 1.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, No. 2, December : 2023

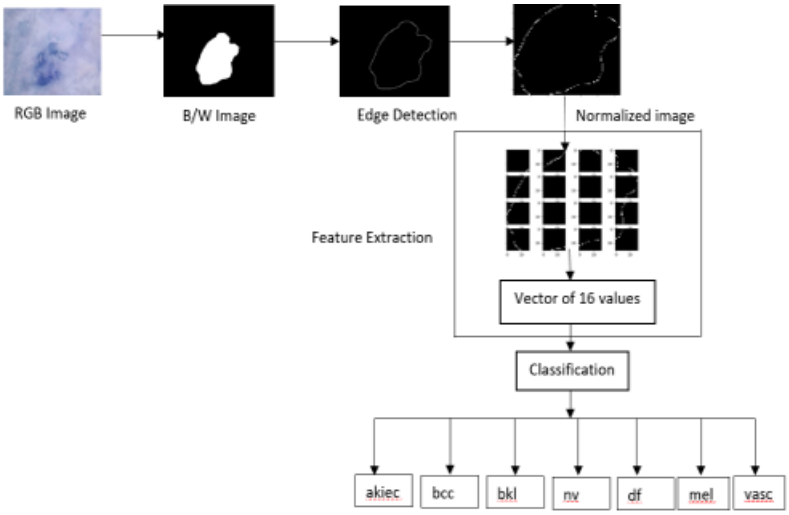

Figure 1: Architecture of Methodology used

3.1 Image Acquisition:

We used the public dataset HAM10000 which contains 10015 images having .jpg extension. Initially we read the color image from the dataset and we convert the RGB image to black and white using Binary Segmentation mask algorithm and store in .png extension.

3.2 Edge Detection:

For Edge Detection many algorithms are available, we have used Canny edge detection algorithm in the present work. The main stages for Canny edge detection algorithm are:

- Filtering noise using Gaussian blur algorithm.
- Finding strength and direction of edges.
- Isolating edges and thin to one-pixel using non-maximum suppression.
- Hysteresis to isolate best edges.

3.3 Image Normalization:

For uniformation of all the images, we normalized the image into 64X64 size. For that we create a bounding box after surrounding the edge image in the previous stage.

3.4 Feature Extraction:

For extracting the features from the normalized image we divide into 4X4(equal size of 16X16 pixel) for non-overlapping 16 regions and further feature pixel density ratio is calculated by number of white pixels divided by number of black pixel for the specific region same process is repeated for the all 16 regions of the input image and thereby creating a feature vector of 16 value.

3.5 KNN classifier:

With the daily problems noted in medical sector, so for solving such a problem faced by the dermatologist machine learning techniques are used. A wide range of various techniques and algorithm are there such as K-Nearest Neighbor (KNN), Naïve Bayes, Support Vector Machine, Random Forest, Decision Tree and K-mean clustering. Among this algorithm the most frequently used algorithm is as K-Nearest Neighbor as K-Nearest Neighbor algorithm is one of the simplest learning algorithms of the supervised machine learning technique. This algorithm is known as lazy algorithm as it does not immediately learn from the training data rather at time of classification it stores dataset and perform action on dataset [25].

Since K-Nearest Neighbor algorithm works well mostly for the classification problems. At training phase K-Nearest Neighbor algorithm stores dataset and when a new dataset is given, it classifies data into category that is similar to the new data. In the proposed work we used K-Nearest Neighbor at initial stage for better and significant improvement for identification of skin disease this model works well.

The working of K-Nearest Neighbor is described as below:

- First selecting number of K of neighbors.

- Calculating Euclidean distance for K number of neighbors.
- Considering KNN as calculated by Euclidean distance.
- Counting the data points for each category according for the K neighbor.
- The maximum number of the neighbor that is obtain is assign to the new data point.
- Model is ready to use.

3.6 Model Implementation:

In the proposed model, we divide the dataset into 80:20 that is 80% is given for the training data and 20% is given for testing the data to. After train test split ratio, the model is trained and predict the image using the test data and for summarizing better performance of algorithm we create a confusion matrix. The proposed model is able to identify the different type of skin disease such as akiec, bcc, bkl, nv, mel, df and vasc.

IV. Result and Discussion

The system is implemented in google colab. We use a platform Intel core i7 processor with 8GB RAM.

4.1 Dataset Preparation:

In this proposed system we used a public dataset HAM10000 for model preparation. Table 1. shows a summary of the dataset used.

Table 1. Summary of HAM10000 dataset

Dataset	Total Image	Akiec	Bcc	Bkl	Df	Mel	Nv	Vasc
HAM1000	10015	327	541	1099	155	1113	6705	142

Initially input images are preprocessed and features are extracted using region wise pixel density ratio. Finally, we have used supervised machine learning algorithm K-Nearest Neighbor classifier for classification of various skin disease. In the proposed system, we have taken 100 images from each category of public HAM10000 dataset and trained with the KNN model.

As mention in Figure 2. And Figure 3. It shows some sample images of different category. Figure 2. Shows RGB images and Figure 3. Shows the black and white images from public HAM10000 dataset.

Figure 2: RGB Image

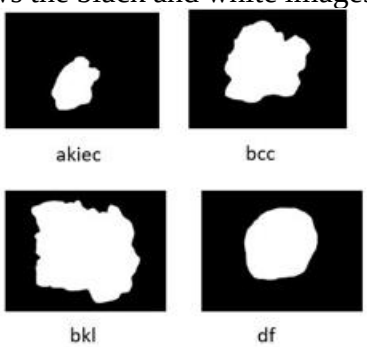

Figure 3: B/W Image

4.2 Experimental Analysis:

This section presents the experimental analysis of the proposed work in which we have used supervised machine learning algorithm K-Nearest Neighbor which is being used for identification of skin disease. The results show the confusion matrix which represents the number of correct classification and how many are wrongly classify.

The confusion matrix for the K-Nearest Neighbor model is shown in Table 2. The diagonal of the matrix represents about the accuracy of the algorithm. In case of the highest correct answers are in 0th prediction and 0th label, where X-axis depicts the Prediction and Y-axis depicts the Labels. The accuracy that we found using the K-Nearest Neighbor model for 7 different skin disease is 97%.

Table 2. Confusion Matrix

Prediction Class							
Metrics	**Akiec(0)**	**Bcc(1)**	**Bkl(2)**	**Df(3)**	**Mel(4)**	**Nv(5)**	**Vasc(6)**
Akiec(0)	25	0	0	0	0	1	0
Bcc(1)	0	22	0	1	0	0	0
Bkl(2)	0	0	19	0	0	0	0
Df(3)	0	0	0	19	0	0	0
Mel(4)	0	0	0	0	14	0	0
Nv(5)	0	0	2	0	0	17	0
Vasc(6)	0	0	0	0	0	0	20

V. Conclusion

Machine Learning have been widely used for the very accurate learning method of identifying and classification of images. In this work, we present the supervised machine learning algorithm K-Nearest Neighbor for classification of the skin disease and public dataset HAM1000 used for the proposed work which consists of 7 different types of skin disease image. Further for feature extraction pixel density ratio is being used and image is being normalized into 64X64 size and this image is being divided into 16 non-overlapping regions. The accuracy for K-Nearest Neighbor model is 97%. Also, for the different disease these model can be used and for various different medical diagnosis such as X-ray, MRI etc. this model can be used.

For Future enhancement the present work can be extended by increasing the dataset size by adding more skin disease images. Further the accuracy can be increased by implementing advanced machine learning algorithm like deep learning.

REFERENCES

[1] D.A. Okuboyejo, O.O. Olugbara and S.A. Odunaike (2013), "Automating skin disease diagnosis using image classification", In Procedding of the world congress on engineering and computer science, Vol. 2, pp. 850-854.

[2] Arifin, S., Kibria, G., Firoze, A., Amini, A., & Yan, H (2012), "Dermatological Disease Diagnosis Using Color-Skin Images", International Conference on Machine Learning and Cybernetics.

[3] Yasir, R., Rahman, A., & Ahmed (2014), "Dermatological Disease Detection using Image Processing and Artificial Neural Network", International Conference on Electrical and Computer Engineering.

[4] Santy, A., & Joseph (2015), "Segmentation Methods for Computer Aided Melanoma Detection", Global Conference on Communication Technologies.

[5] Zeljkovic, V., Druzgalski, C., Bojic-Minic, S., Tameze, C., & Mayorga, P (2015), "Supplemental Melanoma Diagnosis for Darker Skin Complexion Gradients", Pan American Health Care Exchanges.

[6] Suganya R (2016), "An Automated Computer Aided Diagnosis of Skin Lesions Detection and Classification for Dermoscopy Images", International Conference on Recent Trends in Information Technology.

[7] Alam, N., Munia, T., Tavakolian, K., Vasefi, V., MacKinnon, N., & Fazel-Rezai, R (2016), "Automatic Detection and Severity Measurement of Eczema Using Image Processing", IEEE.

[8] Kumar, V., Kumar, S., & Saboo (2016), "Dermatological Disease Detection Using Image Processing and Machine Learning", IEEE.

[9] Abrham debasu mengistu, dagnachew melesew Alemayehu (2015), "Computer vision for skin cancer diagnosis and recognition using rbf and som", International journal of image processing (IJIP), volume (9): issue (6).

[10] S.S. Mane1, s.v. Shinde (2017), "Different techniques for skin cancer detection using dermoscopy images", International journal of computer sciences and engineering vol.5(12), e-issn: 2347-2693.
[11] Yuexiang li and linlin shen (2018), "Skin lesion analysis towards melanoma detection using deep learning network", International Journal of Engineering Research & Technology (IJERT).
[12] Muhammad imran razzak, saeeda naz and ahmad zaib (2018), "Deep learning for medical image processing: overview, challenges and future", International Journal of Engineering Research & Technology (IJERT).
[13] L. G. Kabari and F. S. Bakpo, Member (2009), "Diagnosing Skin Diseases Using an Artificial Neural Network", IEEE.
[14] M. Shamsul Arifin, M. Golam Kibria, Adnan Firoze, M. Ashraful Amin, Hong Yan (2012), "Dermatological Disease Diagnosis Using Colour-skin Images", Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian.
[15] Hadzli Hashim, Rozita Jailani, and Mohd Nasir Taib (2002), "A Visual Record of Medical Skin Disease Imaging Using MATLAB Tools", Student Conference on Research and Development Proceedings.
[16] Chuan-Yu Chang, Heng-Yi Liao (2011), "Automatic Facial Skin Defects Detection and Recognition System", Fifth International Conference on Genetic and Evolutionary Computing.
[17] Rozita Jailanil, Mohd Nasir Taib and Saadiih Sulaiman (2003), "Colour Space for Psoriasis Skin Diseases Analysis", Asian Conference on Sensors.
[18] Shang Keke, Zhou Peng, Li Guohui (2010), "Study on Skin Colour Image Segmentation Used by Fuzzy-c-means arithmetic", Seventh International Conference on Fuzzy Systems and Knowledge Discovery.
[19] Nibaran Das, Anabik Pal, Sanjoy Mazumder, Somenath Sarkar, Dwijendranath Gangopadhyay, Mita Nasipuri (2013), "An SVM based skin disease identification using Local Binary Patterns", Third International Conference on Advances in Computing and Communications.
[20] Shuzlina Abdul-Rahman, Ahmad Khairil Norhan, Marina Yusoff, Azlinah Mohamed, Sofianita Mutalib (2012), "Dermatology Diagnosis with Feature Selection Methods and Artificial Neural Network", IEEE EMBS International Conference on Biomedical Engineering and Sciences, Langkawi.
[21] Ibrahim Mahmood Ibrahim, Adnan Mohsin Abdulazeez (2021), "The Role of Machine Learning Algorithms for Diagnosing Diseases", Journal Of Applied Science And Technology Trends.
[22] H. R. Mhaske, D. A. Phalke (2013), "Melanoma Skin Cancer Detection and Classification based on Supervised and Unsupervised Learning", International conference on Circuits, Controls and Communications (CCUBE).
[23] Martin Kreutz, Maik Anschiitz, Stefan Gehlen, Thorsten Griinendick, Klaus Hoffmann (2001), "Automated Diagnosis of Skin Cancer Using Digital Image Processing and Mixture-of-Experts", Bildverarbeitung für die Medizin.
[24] Ardan Adi Nugroho, Isnandar Slamet, Sugiyanto (2019), "Skin Cancer Identification System Of HAM10000 Skin Cancer Dataset Using Convolutional Neural Network", AIP Conference Proceedings 2202, 020039.
[25] Songbo Tan (2006), "An effective refinement strategy for KNN text classifier", Expert Systems with Applications Volume 30, Issue 2, Pages 290-298.