International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 14, Number 1 (2018), pp. 29-39

http://www.ripublication.com

Machine Learning Approaches to Multi-Class Human Skin Disease Detection

Ms. Seema Kolkur [1], Dr. D.R. Kalbande [2], Dr. Vidya Kharkar [3]

[1]Department of Computer Engineering,
Thadomal Shahani Engineering College, Bandra, Mumbai, India.

[2]Department of Computer Engineering, Sardar Patel Institute of Technology,
Andheri, Mumbai, India.

[3]Department of Skin and VD., G.S.M.C & K.E.M Hosp., Parel,
Mumbai-400012, India.

Abstract

Skin is the largest organ of human body. Due to different environmental and personal factors a lot people are suffering from skin diseases. Many skin diseases need screening by expert dermatologist. Due to lack of medical facilities available, the patients may need to wait which in turn may increase chances of increasing severity and/or spreading the infections. To avoid this, early detection of these diseases is important. We developed a system which identifies the disease based on input symptoms. We have acquired symptoms data of 10 skin diseases with the help of expert doctor in the field. Different classifiers were trained on the symptom's data. We got 90% above accuracy.

Keywords: Machine learning, Classification, Neural Networks

1. INTRODUCTION

In India, skin diseases are increasing at a very fast rate. Different factors contributing to this growth are developing economy, differences in climate, illiteracy, personal hygiene issues, lack of awareness, social backwardness, dearth of primary health

centers in rural area, pollution and life style particularly in urban area etc [1].The prevalence of skin diseases in the general population has varied from 7.86% to 11.16% in various studies [2-3]. Due to discoloring and other problems with these diseases, the sufferers face many challenges physically and socially [4].Many skin diseases need screening by an expert dermatologist. Normally, manual diagnosis by experts is subjective and is based on the personal expertise in the area. To achieve objective accuracy of diagnosis, computer aided diagnosis is normally used.

With the advancement in computing technologies and digitization of medical field, computer aided diagnosis has become a reality. Machine learning is widely in the field of medical diagnosis.

2. RELATED WORK

Skin Disease Detection is a studied by many researchers in the field of Machine Learning and Artificial Intelligence. Skin Disease Detection is basically a classification task. Researchers have used different data mining, statistics, machine learning algorithms in past. Neural networks and Support Vector Machine(SVM) are widely used in literature. [5] gives good overview of different data mining techniques used for skin disease detection

In [6] authors used different data mining algorithms like Multilayer Perceptron (MLP), Naive Bayes and decision tree to predict the Erythemato Squamous skin Diseases(ESD) diseases. Naive Bayes classifier performed best with a classification accuracy of 97.4% in their case. [7-9] used extreme learning machine (ELM), Self-Organization Map and Genetic algorithms with Adaptive neuro fuzzy inference system to better identify ESD respectively. [10] applied SVM classifier for detection of leucoderma. In the works of [15,17,19,20] SVM was used for skin cancer detection. [11-14,16,18,21] shows use of neural networks in the detection of Scabies, Psoriasis and skin cancer. Use of KNN for Non-melanoma skin lesion classification is shown in [22]. In our previous work, we shown use of different algorithms with R for detection of ESD Disease[23].

3. METHODOLOGY

General methodology of the proposed system is depicted fig. 1. The system can be broadly categorized into data collection, preprocessing, feature extraction and classification modules.

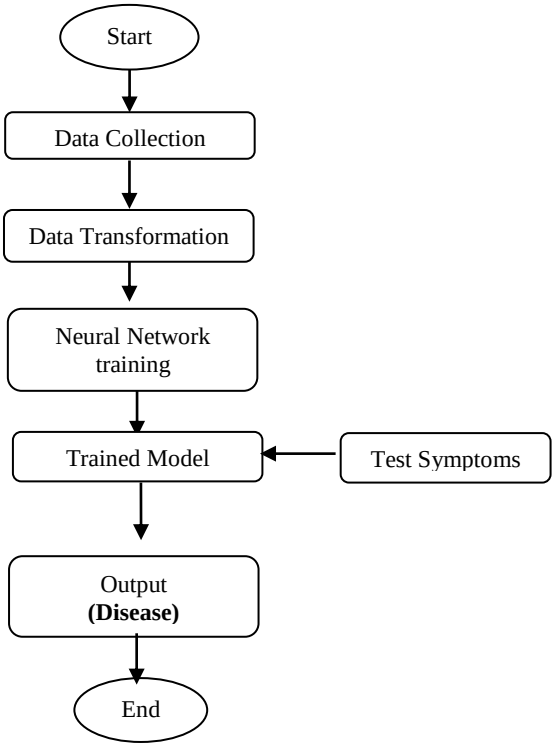

Fig.1 Methodology of Proposed Work

3.1 Data Collection

We have collected our dataset of symptoms and patient details related to ten common skin diseases from the department of Skin and VD from G. S. M. C. KEM Hospital, Parel Mumbai after rounds of dialogues with concerned doctor. Table 1 lists the diseases used in the study. Table 2 shows dataset description along with type of variable and possible range of values.

Table 1. List of Disease Names

Sr. No.	**Disease Name**
1.	Atopic Dermatitis
2.	Folliculitis
3.	Leprosy
4.	Lichen Planus
5.	Warts
6.	Herpes Zoster
7.	Pediculosis Captis
8.	Pityriasis Versicolor
9.	Seborheic Dermatitis
10.	Vitiligo

Table 2. Dataset Description

Sr. No.	**Symptoms**	**Type of Variable**	**Domain of Values**
1	Scaling	Categorical	{Thick, Thin}
2	Puss	Boolean	{Present, Absent}
3	Watering	Boolean	{Present, Absent}
4	Pain	Range	[1-10]
5	Itching	Range	[1-10]
6	Redness	Range	[1-10]
7	Skin Elevation	Categorical	{No Elevation, Papule (<1cm), Macule (>1cm)}
8	Hair Loss	Boolean	{Present, Absent}
9	Pigmentation	Categorical	{No Pigmentation, Hyper Pigmentation, Hypo pigmentation, De-pigmentation}
10	Affected Region	Categorical	{Scalp, Folds, Upper Extremity, Lower Extremity, Trunk, All Over}
11	Duration	Categorical	{less than 1 month, greater than 1month}
12	Age	Categorical	{Child, Adult}
13	Loss of Sensation	Boolean	{Present, Absent}

3.2 Data Transformation

All categorical and Boolean attributes are converted to numerical integers for ease of calculation.

3.3 Classification

The dataset was divided in training and validation dataset. We trained different classifiers on the training dataset [23].

3.3.1 Artificial Neural Networks(ANN)

ANN is one of the best classifier used in many applications of classification. Neural network is inspired by working of human brain. A neural network is a collection of a group of units, called neurons which are interconnected to each other. Every neuron, receives inputs, applies an activation function on sum of input and randomly initialized weights and yields output values [23, 24].

MultiLayer Perceptron model organizes all neurons in different layers such that every neuron in a layer is well connected to the neurons in the previous layer and the next layer with weighted connections. MLP model has an input layer, an output layer and 1

or more hidden layers. Nodes in the input layer correspond to dimensionality of the input data. Nodes in hidden layer apply activation function on sum obtained by multiplying inputs with node's weights and bias. Neurons in middle (hidden) layers use logistic activation function and the ones in the output layer use softmax function. The number of nodes in the output layer is equivalent to the number of possible output classes [25].

The Back Propagation Algorithm (BPA)is an iterative and supervised training algorithm used to train a multilayer neural network to fit to a labelled dataset. For every input record the network examines its output response. The error value, which is difference between to the known and calculated output is calculated. Based on the magnitude of this error, the connection weights are updated in the next cycle [26].

3.3.2 K Nearest Neighbors (KNN)

K nearest neighbors is an unsupervised non-parametric algorithm which classifies a new instance based on a similarity measure. Normally some distance function like Euclidean Distance is used as similarity measure. An input instance is classified to a class with maximum class label in its K nearest neighbors. KNN is used for many statistical and pattern recognition applications [23,27].

3.3.3 Decision Tree

Decision tree is a supervised learning method used for constructing prediction model from given data. The tree is constructed in top-down recursive manner by partitioning the data space based on some splitting criteria. A model is constructed for each partition successively [23,27].

3.3.4 Support Vector Machine (SVM)

SVM is based on the concept of discovery of hyperplanes that can find boundaries for different groups of objects. A decision plane separates all the objects in n-dimensional space based on their class labels. To construct an optimal hyperplane, SVM minimizes an error function in every iteration during training. It is used as classifier that classifies input data by constructing such hyperplanes [23]. SVM can be used for both classification and regression [28].

3.3.5 Random Forest

Random forest is an ensemble learning technique. It constructs many decision trees during training. To classify a new object, it is given as an input to each of the trees in the forest. Each tree generates output classification. The forest chooses the class with majority votes [23, 29].

4. RESULTS

We trained individual classifiers mentioned above on training symptoms dataset. A neural network is trained on a dataset to predict possible disease as an output in R with sigmoid activation function [28]. The architecture of neural network is 13-0-10 with 13 input neurons, 10 output neurons and no neurons hidden layer as shown in fig 2 which becomes a simple logistic regression.

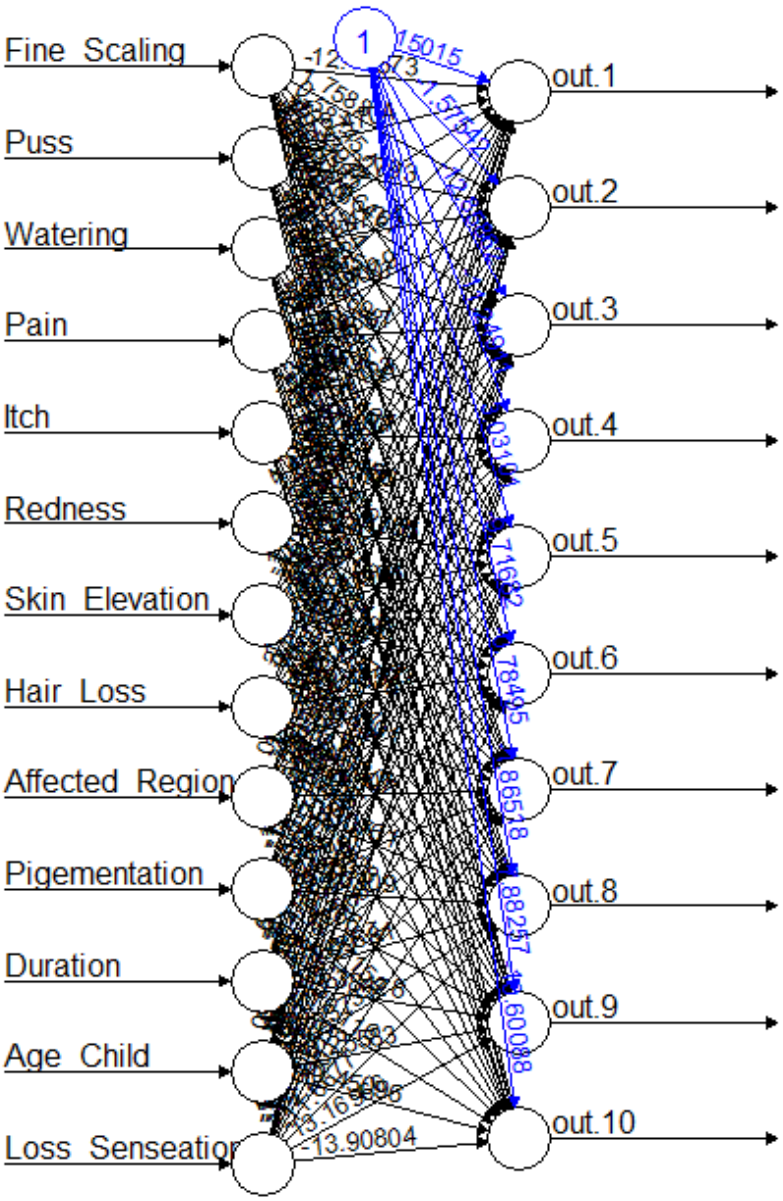

Fig. 2. Neural Network Architecture

The trained network is tested for validation on one-third of the database with 100% of accuracy. Fig. 3 shows confusion matrix for the same. All the non-diagonal entries are zero.

		Actual Labels					
		1	**4**	**5**	**7**	**8**	**9**
Predicted Labels	**1**	8	0	0	0	0	0
	4	0	4	0	0	0	0
	5	0	0	15	0	0	0
	7	0	0	0	5	0	0
	8	0	0	0	0	3	0
	9	0	0	0	0	0	10

Fig. 3. Confusion Matrix for validation Test

Fig. 4 shows decision tree on the training dataset. All leaf nodes represent the disease class. As can be seen highest importance is given to redness of skin lesion being a root node of the tree.

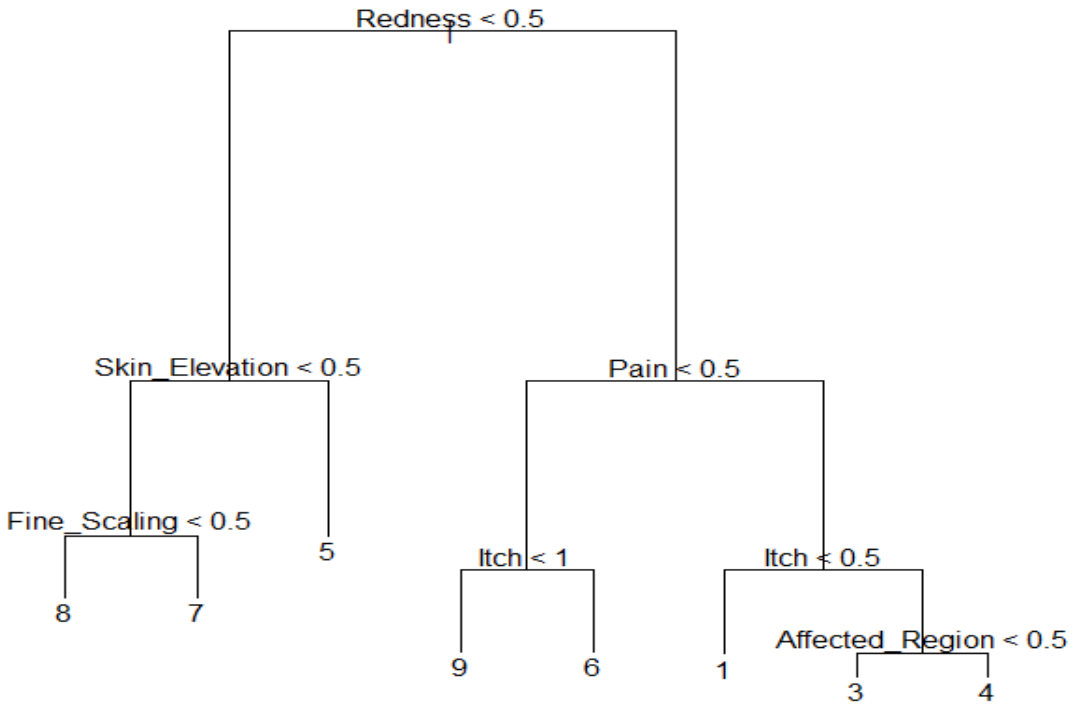

Fig. 4. Decision tree on training dataset

Performance of a classifier model is measured in terms of precision, recall and accuracy. Accuracies of these classifiers on validation dataset which are calculated by equation 1 are summarized in table 3 [27].

Table 4. Accuracy of Classifiers

Sr. No.	Classifier Model	Accuracy
1.	ANN	100%
2.	KNN	98&
3.	SVM	99%
4.	Decision Tree	97.7%
5.	Random Forest	100%

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (1)$$

We also developed a simple GUI which can be used for detection of disease for new patient as shown in fig. 5. The system is verified successfully by concerned doctor in the hospital.

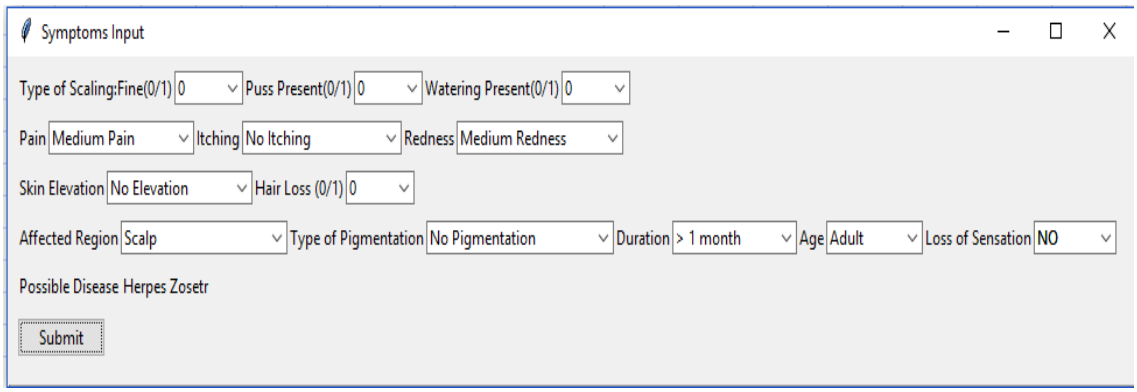

Fig. 5 GUI in Testing Phase

5. CONCLUSION

In this paper, we discussed different classifiers for skin disease detection. Due to digitalization and improvement in technology, more and more detailed data is becoming available. Medical data mining can be used for diagnosis, and decision making, etc. We collected original data related to 10 common skin diseases and applied five different classifiers for classification of an input record to one of these 10 diseases. Results indicate these classifiers do well in identification of disease. This system can help new medical practitioners to correctly identify disease in case of doubt.

REFERENCES

[1] Jain S, Barambhe M S, Jain J, Jajoo U N, Pandey N. Prevalence of skin diseases in rural Central India: A community-based, cross-sectional, observational study. J Mahatma Gandhi Inst Med Sci 2016; 21:111-5

[2] Grover S, Ranyal RK, Bedi MK. A cross section of skin diseases in rural Allahabad. Indian J Dermatol 2008; 53:179-81

[3] Rao GS, Kumar SS, Sandhya. Pattern of skin diseases in an Indian village. Indian J Med Sci 2003; 57:108-10.

[4] G. K. Jana, A. Gupta, A. Das, R. Tripathy, and P. Sahoo. Herbal treatment to skin diseases: A global approach. Drug Invention Today, 2(8):381–384, August 2010.

[5] E. Barati at el, A Survey on Utilization of Data Mining Approaches for Dermatological (Skin) Diseases Prediction, Journal of Selected Areas in Health Informatics (JSHI): March Edition, 2011

[6] Kwetishe Danjuma, Adenike Osofisan, Evaluation of Predictive Data Mining Algorithms in Erythemato-Squamous Disease Diagnosis, IJCSI International Journal of Computer Science Issues, 11(6), 85-94 (2014)arXiv:1501.00607v1 [cs.LG]

[7] Sunday Olusanya Olatunji and Hossain Arif, Identification Of Erythemato-Squamous Skin Diseases Using Extreme Learning Machine And Artificial Neural Network, ICTACT Journal On Soft Computing, October 2013, Volume: 04, Issue: 01

[8] Haryanto et al, The Erythemato-Squamous Dermatology Diseases Severity Determination using Self-Organization Map, IPTEK, Journal of Proceeding Series, Vol. 1, 2014 pp 279-284

[9] Lejla et al, GA-ANFIS Expert System Prototype forPrediction of Dermatological Diseases, Europian Federation of Medical Informatics(EFMI) 2015, doi:10.3233/978-1-61499-512-8-622

[10] Archana Singh, Ajay Rana and Shipra Srivastava, Comparative Analysis and Automatic detection of various skin diseases and Implementation of leucoderma using support vector machine, International Journal of Computer Applications (0975 – 8887) Volume 48– No.24, June 2012

[11] Dr. Trupti, P. Shah, Pooja J. Shah, Connectionist Expert System for Medical Diagnosis using ANN– A case study of skin disease Scabies, International Journal of Advanced Research in Computer Science and Software Engineering 3(8), August - 2013, pp. 227-230

[12] Nidhal K. Abbadi, Nizar Saadi Dahir, Muhsin A. Al-Dhalimi and Hind Restom, Psoriasis Detection Using Skin Color and Texture Features, Journal of Computer Science 6 (6): 48-652, 2010

[13] Priyanka s. Biradar,S. N.Patil, Psoriasis Detection Using Color and Texture Feature, International Journal of Advanced Computing and Electronics Technology (IJACET) , Volume-2, Issue-4, 2015 154, pp 153-156

[14] Sheha A.M, Mabrouk S.M and Sharawy, Automatic Detection of Melanoma Skin Cancer using Texture Analysis, International Journal of Computer Applications (0975 – 8887) Volume 42– No.20, March 2012

[15] V. Jeya Ramya, J. Navarajan, R. Prathipa and L. Ashok Kumar, DETECTION OF MELANOMA SKIN CANCER USING DIGITAL CAMERA IMAGES, ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 7, APRIL 2015,pp 3082-3085.

[16] Sonali Jadhav, D. K. Kamat, Analysis and Detection of Skin Cancer, IOSR Journal of Electronics and Communication Engineering, Volume 9, Issue 4, Ver. I (Jul - Aug. 2014), PP 50-54

[17] Idris Nayaz Ahmed Chaya P, Segmentation and Classification of Skin Cancer Images, International Journal of Advanced Research in Computer Science and Software EngineeringVolume 4, Issue 5, May 2014

[18] Dr. J. Abdul Jaleel, Sibi Salim, Aswin.R.B, Diagnosis and Detection of Skin Cancer Using Artificial Intelligence, International Journal of Engineering and Innovative Technology, Volume 3, Issue 2, August 2013, pp311-315

[19] Bethanney Janney, G Rohini, and J Pauline, Detection of Malignancy of Skin Lesions Based on Local Texture And Fractal Analysis, Research Journal of Pharmaceutical, Biological and Chemical Sciences, pp 408-412

[20] Rashi Goel, Saranjeet Singh, Skin Cancer Detection Using GLCM Matrix Analysis and SVM Classifier, International Journal of Applied Engineering and Technology Vol. 5 (1) January-March2015, pp.6-11

[21] Goel Rashi, Saranjeet Singh, Skin Cancer Detection using GLCM Matrix Analysis and Back Propagation Neural Network Classifier, *International Journal of Computer Applications* 112.9 (2015)

[22] Ballerini, Lucia, et al. Non-melanoma skin lesion classification using colour image data in a hierarchical K-NN classifier, 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI) IEEE, 2012, pp 358-361

[23] Seema Kolkur, D. R. Kalbande. Review of Machine Learning Algorithms in R software for Diagnosis of ESD Diseases, Proceedings of the ACM Symposium on Women in Research 2016 - WIR '16, 2016

[24] N.Satyanandam, Dr. Ch. Satyanarayana, Md.Riyazuddin, Amjan.ShaikData mining machine learning approaches and medical diagnosis system: A survey, International journal of computer and organization trends,(3),53-60.(2012)

[25] Bakpo, F. S. and Kabari, L. G (2011). Diagnosing Skin Diseases Using an Artificial Neural Network, Artificial Neural Networks - Methodological

Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), ISBN: 978-953-307-243-2

[26] Dr. Jacek M. Zurada (1992), Introduction to Artificial Neural Systems, West Publishing Co. ISBN:0-314-93391-3. 1992

[27] Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd edition.

[28] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani(2013) , An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)

[29] https://documents.software.dell.com/statistics/textbook/support-vector-machines-svm