The Pharma Innovation Journal 2017; 6(10): 248-251

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating 2017: 5.03
TPI 2017; 6(10): 248-251

www.thepharmajournal.com
Received: 05-08-2017
Accepted: 06-09-2017

S Samundeswari
(a) Department of Computer Science and Engineering, Annamalai University Annamalai Nagar, Tamil Nadu, India
(b) Department of Computer Science and Engineering, Sri Sai Ram Engineering College, Chennai, Tamil Nadu, India

V Ramalingam
Department of Computer Science and Engineering, Annamalai University Annamalai Nagar, Tamil Nadu, India

B Latha
Department of Computer Science and Engineering, Sri Sai Ram Engineering College, Chennai, Tamil Nadu, India

S Palanivel
Department of Computer Science and Engineering, Annamalai University Annamalai Nagar, Tamil Nadu, India

Correspondence
S Samundeswari
(a) Department of Computer Science and Engineering, Annamalai University Annamalai Nagar, Tamil Nadu, India
(b) Department of Computer Science and Engineering, Sri Sai Ram Engineering College, Chennai, Tamil Nadu, India

Prediction of systemic lupus erythematosus based on cutaneous manifestations using machine learning models

S Samundeswari, V Ramalingam, B Latha and S Palanivel

Abstract
Systemic Lupus Erythematosus (SLE) or Lupus is an auto immune system in which the human insusceptible framework ends up plainly hyperactive and affects typical, solid tissues. Diagnosing the lupus is troublesome and the specialist may set aside long opportunity to analyze this intricate sickness precisely. This experimental study made utilization of images gathered from various hospitals in Tamilnadu. Images of 400 patients (200 SLE and 200 normal) are taken into this study. Machine learning models can offer a help to the doctor to predict the disease at the early stage early stage itself. Three classifiers, Naïve Bayes, Multilayer perceptron (MLP) and Random Forest are considered to classify the patients. Experimental results demonstrate that the MLP gives higher accuracy than different models for recognizing SLE.

Keywords: systemic lupus erythematosus (SLE), cutaneous manifestation, multilayer perceptron (MLP), naïve bayes (NB), random forest (RF), lupus

1. Introduction

Systemic Lupus Erythematosus (SLE) or Lupus is an immune system illness in which the human insusceptible framework winds up plainly hyperactive and affects ordinary, sound tissues. Systemic Lupus Erythematosus (SLE) is the immune system sicknesses in which body tissues are affected by its own particular insusceptible framework. SLE can influence various organs in the body like heart, lungs, liver, kidneys and central nervous system. At the point when contrasted with men, the infection seems nine times ordinarily in ladies; particularly it is more typical in ladies of youngster bearing ages. Sadly, the reason for lupus is flighty with times of ailment and health. This variety prompts analysis of SLE exceptionally difficult.

The greater part of the lupus patients have cutaneous signs. According to Gilliam classification, Lupus skin lesions are classified into lupus-specific (e.g., malar rash) and non-specific skin lesions (e.g., alopecia) [1]. Lupus specific is grouped into three noteworthy subtypes: systemic or acute cutaneous lupus erythematosus (ACLE), subacute cutaneous lupus erythematosus (SCLE) and chronic cutaneous lupus erythematosus (CCLE) [1-3]. Culture, Environment and hereditary qualities make fluctuation in frequency and infection seriousness among different racial and ethnic groups [5]. Laman *et al* [6] proposed the variance in the skin lesions from malar rash, discoid lupus to oral ulcer, alopecia, bullous lesions and so forth. The American College of Rheumatology (ACR) proposed 11 criteria as a characterization tool for recognizing the SLE in the patients [4]. Cutaneous lesions ought to fulfil for four of the 11 re-examined ACR criteria of SLE. Lupus-specific skin lesions can be effectively analyzed while lupus non-specific skin lesions are related with the failure of different organs and subsequently patients should be continuously monitored [7] for the severity of the disease. Therefore, a rheumatologist ought to exclusively assess the patients in view of the clinical and symptomatic criteria for viable diagnosis and treatment.

Color Histogram is broadly used to extract color features of an image [8]. It is used to discover the frequency distribution of the color bins in an image. Pixels having same intensity are counted and stored. As color histogram can be easily calculated, it is vital for analyzing the images [10]. Nanni *et al*. [11] proposed combination of different surface descriptors utilizing GLCM and have been tried for various classification of medical images. Gabor filters, local binary patterns (LBP), scale-invariant feature transform (SIFT), histogram of oriented gradients (HOG), speeded up robust feature (SURF) techniques are also used to extract the texture features. The Gray - Level Co-Occurrence Matrix (GLCM) feature descriptors are extracted by finding out the relationship between the two neighbouring pixels [9].

2. Materials and Methods

Patients who fulfilled the revised 2012 SLICC ACR characterization criteria for SLE are selected and their images are collected from different hospitals in Tamilnadu. A sample of 400 images (200 SLE and 200 Non-SLE) is taken for this study. These 400 image samples are used to classify the images into SLE and Non-SLE. Some of the cutaneous manifestations of SLE images are shown in figure 1. Pre-processing is the first step in the analysis of the images. The undesirable noise present in the images should be removed so as to enhance the effectiveness of our framework. Hence Medium filter, a nonlinear digital filtering technique is used to remove the undesirable noise.

Fig 1: Image samples of cutaneous manifestation of SLE patients

2.1. Methodology

The objective of the paper is to analyze and compare the performance of the classifiers Naïve Bayes, MLP and Random Forest by extracting the color histogram and texture features in HSV color space from the various images. The general methodology overview of the proposed approach is outlined in figure 2. The proposed framework comprises of three phases: (I) Pre-processing, to remove the noise using Median filtering (ii) Feature extraction to extract the color histogram and texture features in HSV color space (iii) classifiers, Naïve Bayes, MLP and Random Forest to classify the images into SLE and Non-SLE. The performances of the classifiers are assessed by computing the accuracy, sensitivity and specificity.

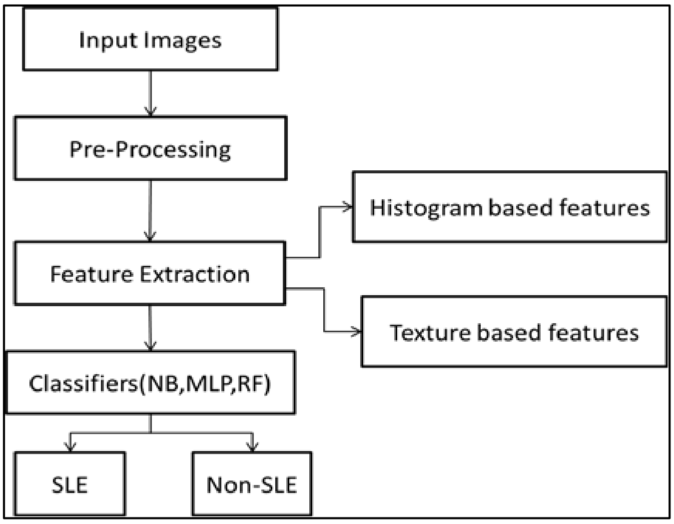

Fig 2: Proposed methodology to classify the images.

2.2 Feature Extraction

Feature extraction is one of the important steps for analysis process. This technique is used to extract a set of highly informative features from the images. In this proposed work, color histogram based features and the texture based features are used to classify the images. The color histogram features of mean, standard deviation, skewness, kurtosis, and energy are computed based on the probability distribution of the histogram intensity levels. 5 histogram features are extracted for each channel of HSV color space. The Gray-Level Co-Occurrence Matrix (GLCM) method is used to extract the texture based features. The GLCM function calculates the spatial association of the pair of pixels with specific values and creates a co-occurance matrix. The texture features are extracted from this matrix. The images are decomposed into three channels of HSV color space thereby obtaining three different images to extract the texture features of the color images. Then the features are extracted for each channel of HSV color space. In our proposed work, 8 features namely, autocorrelation, contrast, cluster prominence, dissimilarity, energy, entropy, sum average and sum variance are computed. Hence a total of 39 features (5 Histogram features and 8 texture features for each channel of the color space) are extracted from the images and given as an input to the classifiers. The extracted features are shown in Table 1.

Table 1: Features extracted for classification of images

Histogram based features	Texture features
Mean	Autocorrelation
Variance	Contrast
Skewness	Cluster Prominence
Kurtosis	Dissimilarity
Energy	Energy
	Entropy
	Sum average
	Sum variance

2.3 Classification Techniques

Classifiers, Naïve Bayes, Multilayer Perceptron and Random Forest are used to classify the images using the features descriptors. Naïve Bayes classifier is simple classifier which is based on Bayes Theorem of conditional probability and strong independence assumptions. This classifier emphasizes on measure of probability that whether the document A belongs to class B or not. It is based on the assumption that the occurrence or non occurrence of a particular attribute is unrelated to the occurrence or non occurrence of a particular attribute [14]. MLP neural networks [15] comprise of three layers of nodes. It has (i) input layer, (ii) output layer, and (iii) hidden layer and each of which contains input node (s) otherwise called as sensory nodes, output node (s) called as responding nodes, and hidden node (s), respectively. The hidden and output nodes are the processing elements which contains the activation function that is used to obtain the output. Random Forest [15] uses decision tree as base classifier. Random Forest generates multiple decision trees; the randomization is present in two ways: (1) random sampling of data for bootstrap samples as it is done in bagging and (2) random selection of input features for generating individual base decision trees. Strength of individual decision tree classifier and correlation among base trees are key issues which decide generalization error of a Random Forest classifier.

3. Experimental Results

In this proposed work, images of both SLE and normal patients are collected. These images are used to analyze and evaluate the performance of the classifiers for identifying the lupus. Feature descriptors are extracted from the images based on GLCM and color histogram in HSV color space. The classifiers, Naïve Bayes, MLP and Random Forest are trained and tested to produce the required output for the given features. Precision, recall, specificity and accuracy are the performance indicators used to evaluate the performance of the above three classifiers for the classification of the images into SLE and Non-SLE.

Precision = TP/ (TP+FP) (1)

Recall = TP/ (TP+FN) (2)

Specificity = TN/ (TN+FP) (3)

Accuracy = (TP+TN)/ (TP+TN+FP+FN) (4)

Where TP (True Positive) represents the number of correctly classified SLE images, FP (False Positive) represents the number of misclassified SLE images, FN (False Negative) represents the number of misclassified Non-SLE images, and TN (True Negative) represents the number of correctly classified Non-SLE images.

3.1 Performance Evaluation

Table 2 outlines the classifiers with corresponding precision and recall values. Figure 3 is the graphical representation of Table 2. Figure 3 shows that MLP performed better than the other classifiers Naïve Bayes and Random Forest.

Table 2: Classifiers with corresponding Precision and Recall values

Classifier	Precision (%)		Recall (%)	
	SLE	NON-SLE	SLE	NON-SLE
NB	75.4	76.1	76.5	75
MLP	0.857	0.868	0.870	0.855
RF	82.3	85.3	86	81.5

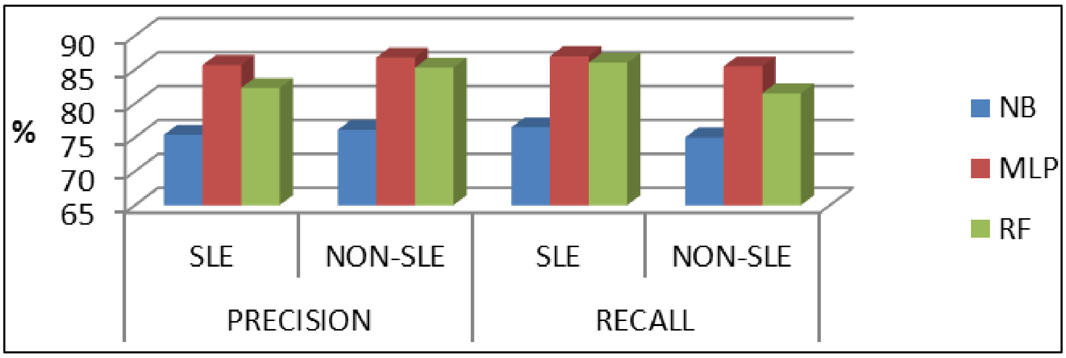

Fig 3: Classifiers vs. Precision and Recall

The performance measures (sensitivity, specificity and accuracy) of various classifiers are shown in Table 3. From Table 3, one can infer that again MLP outperforms Naïve Bayes and Random Forest in all the performance descriptors, namely, sensitivity (87%), specificity (85.5%) and accuracy (86.25%).

Table 3: Classifiers versus various performance measures

Classifier	Sensitivity (%)	Specificity (%)	Accuracy Value (%)
NB	76.50	75.00	75.75
MLP	87.00	85.50	86.25
SVM	86.00	81.50	83.75

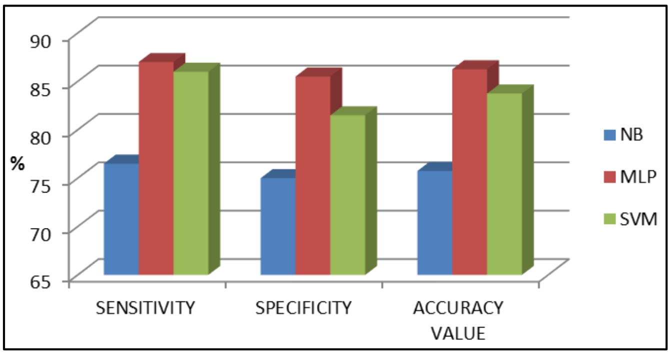

Fig 4: Performance measures (sensitivity, specificity and accuracy) by various classifiers

Figure 4 shows the graphical representation of Table 3. Naïve Bayes classifier shows poor performance when compared with MLP and Random Forest.

4. Conclusion

Prediction of Systemic Lupus Erythematosus based on cutaneous manifestations is proposed. Cutaneous manifestations give imperative information to the rheumatologists to analyze whether the patient has lupus or not. As Lupus affects all the major parts of the body such as lungs, heart, kidneys and central nervous system, it is very important to diagnose the lupus at the early stage to ensure speedy treatment. The diagnosing methodology use different classifier algorithms for the classification of SLE from a group of patients. MLP shows the maximum performance than other classifiers Naïve Bayes and Random Forest. MLP yields maximum of 86.25% as accuracy. If precision, recall, accuracy are considered together, one can conclude that MLP is the better choice than other classifiers, namely, Naïve Bayes and Random Forest. Further this decision is supported by Rheumatologist also.

5. References

1. Gilliam JN, Sontheimer RD. Skin manifestations of SLE. Clin Rheum Dis. 1982; 8:207-18 [PubMed].
2. A Vadivel, AK Majumdar, S Sural. Characteristics of Weighted Feature Vector in Content-Based Image Retrieval Applications, International Conference on Intelligent Sensing and Information Processing, 2004, 127-132.
3. Font J, Bosch X, Ingelmo M, Herrero C, Bielsa I, Mascaró JM. Acquired icthyosis in a patient with SLED, Arch Dermatology. 1990; 126:829 [PubMed].
4. Tan EM, Cohen AS, Fries JF, *et al*. The revised criteria for the classification of systemic lupus erythematosus, Arthritis Rheum. 1982; 25:1271-7.
5. McCauliffe DP. Cutaneous lupus erythematosus. Semin Cutan Med Surg. 2001; 20:14-26 [PubMed].
6. Laman SD, Provost TT. Cutaneous manifestation of SLE. Rheum Dis Clin North Am. 1994, 20:195

[PubMed].
7. Zeevi R, Vojvodi D, Risti B, Pavlović MD, Stefanović D, Karadaglić D. Skin lesions: An indicator of disease activity in SLE? Lupus. 2001; 10:364-7. [PubMed]
8. Zhao Q, Yang J, Liu H. Stone Images Retrieval Based on Color Histogram, in IEEE International Conference on Image Analysis and Signal Processing, 2009, 157-161.
9. Hafner J, Sawhney H, Equitz W, Flickner M, Niblack W. Efficient color histogram indexing for quadratic form distance functions, IEEE Transactions of Pattern Analysis and Machine Intelligence. 1995; 17(7):729-736.
10. Han J, Ma KK. Fuzzy Color Histogram and Its Use in Color Image Retrieval, IEEE Transactions on Image Processing. 2002; 11(8):944-952.
11. Nanni L, Brahnam S, Ghidoni S, Menegatti E, Barrier T. Different Approaches for Extracting Information from the Co-Occurrence Matrix. PLoS One. 2013; 8:12.
12. Gong R, Wang H. Steganalysis for GIF images based on colors-gradient co-occurrence matrix. Optics Communications. 2012; 285(24):4961-4965.
13. Burges CJC. A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery. 1998; 2(2):121-167.
14. Qiang Ye, Ziqiong Zhang, Rob Law. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, Expert Systems with Applications. 2009; 36:6527-6535.
15. Haykin S. Neural Networks: A Comprehensive Foundation, Englewood Cliffs, NJ: Prentice-Hall, 1995.