

Development of Healthcare System Using Soft Computing Methods



Pallavi Mishra and Sachi Nandan Mohanty

Abstract Real-world problems associated with healthcare domain are dynamic and non-deterministic in nature. Traditional analytics models are no longer effective for diagnostic processes and frequently require a large amount of computation time for decision-making, which cannot be sustainable in a technology-driven world. Soft computing methods, such as neural networks, machine learning, fuzzy logic, not only deal with the existing environment, but are also highly self-adaptive in terms of predicting and classifying new diseases that appear in the existing environment. As a result, these techniques are extremely beneficial and effective for use in the healthcare system. In this paper, we proposed a weighted averaging ensemble technique to perform supervised multi-level classification of dermatological diseases. Our experimental results show that the proposed method achieved significant better results than individual classifier models.

Keywords Erythemato-squamous diseases · Soft computing techniques · Neural networks · Ensemble classifiers

1 Introduction

Dermatology is the medical specialty that deals with skin, hair, and nails in their broadest sense. Skin disease has become a major issue for people all over the world [1]. Psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris are the most common erythemato-squamous diseases (ESDs) [2]. Most of these ESDs have similar clinical symptoms such as erythema (skin rashes) and scaling (dryness) in the early stages of the disease, but as the disease progresses, it begins to develop other clinical features, making it difficult

P. Mishra (✉)

Department of Computer Science and Engineering, Faculty of Science and Technology (FST), IcfaiTech, IFHE, Hyderabad, India
e-mail: Pallavi.mishra@ifheindia.org

S. N. Mohanty

School of Computer Science and Engineering (SCOPE), VIT-AP University, Amaravati, AP, India
e-mail: Sachinandan09@gmail.com

for dermatologists to determine which class of disease the patient has [3]. Therefore, it is a critical and challenging task for the medical experts (dermatologists) to distinguish them clinically in order to prescribe the proper treatment and medications. In many cases, steroids and antifungal medications are commonly used to treat various types of skin problems. This type of treatment is harmful to society since it increases the negative effects of local drugs like steroids, which are only effective at suppressing diseases rather than curing them permanently. As a result, this real-world challenge has been a focus of research, with many researchers aiming to construct sophisticated models that will aid medical experts in correctly classifying and diagnosing skin disorders based on symptoms for the benefit of the healthcare domain.

Until now, skin diseases were diagnosed so on the basis of the proficient doctors' long-term experience as well as the outcomes of clinical and pathological tests. Automatic detection or diagnosis of many diseases based on a list of symptoms can vastly improve and speed up the diagnostic procedure. Researchers have already conducted extensive research on this problem for the detection and recognition of various types of dermatological diseases.

Skin diseases have become a common ailment afflicting individuals of all ages. The early-stage identification and recognition of the symptoms would aid in proper diagnosis of the sickness would aid in the design of effective health management [4]. There has been a lot of interest in employing expert systems to conduct medical diagnosis and prescribe successful therapies in recent years. The development of a medical expert system as a knowledge base employing various soft computing approaches such as naïve Bayes [5], neural networks [6–8], fuzzy systems [3], genetic models promises to be a viable way for diagnosis and possible therapy.

With the evolution of soft computing techniques [9] and their successful application in healthcare field, the researchers put their efforts in developing a robust model using soft computing techniques to examine how these complexities of skin diseases diagnosis can be handled effectively. Soft computing techniques form a baseline for the development of computational intelligence (i.e., to build a sense of reasoning power) as like human brain. It has various components like neural network, support vector machines, fuzzy computing, genetic algorithms, etc. The unique feature of soft computing approach is its adaptability to uncertainty in the data and holds good for imprecise and partial truth and approximation input data. Due to the fact that the symptoms of skin diseases are uncertain and imprecise, therefore, the application of soft computing techniques would be proved significant in developing a model for classification [9].

Many research works based on skin disease classification have mainly implemented classifier models such as neural networks, support vector machines [4], nearest neighbor, naïve Bayes and achieved good accuracy. However, the combination of two or more classifier models (i.e., ensemble concept) has been proved effective over individual models in terms of performance.

In this paper, the proposed work is based on ensemble classifier which includes artificial neural network, K -nearest neighbor [10], and support vector machine for skin disease classification. The dataset collected from UCI repository has 366

Table 1 Class distribution of skin diseases

Class code	Class of diseases	No. of samples
1	Psoriasis	112
2	Seborrheic dermatitis	61
3	Lichen planus	72
4	Pityriasis rosea	49
5	Chronic dermatitis	52
6	Pityriasis rubra pilaris	20

instances with 6 target classes as shown in Table 1 [2, 3, 5] to prove insights regarding the number of patients suffering from different dermatological diseases. The dataset contains both clinical and biopsy, i.e., histopathological features as symptoms to classify the disease based on those symptoms. The challenging part lies in the fact that the earlier symptoms tend to vary and show symptoms of other type of skin diseases at subsequential stages, causing ambiguity for the experts in diagnosing the exact diseases at the early stage of the skin diseases.

The number of samples represents the total number of patients suffering from the corresponding class of diseases which again falls under either of 6 types of class codes.

2 Literature Review

Many studies on the diagnosis of erythematous-squamous diseases [3] have been conducted in depth by many researchers utilizing various models and approaches. The findings of the studies are summarized below.

The author [1] used bagging, AdaBoost, and gradient boosting classifier approaches to create an ensemble classifier using six different machine learning algorithms as separate classifiers for skin classification. When compared to individual classifiers, the model claims to achieve better results in ensemble mode.

The performance of CHAID decision trees [2] (DTs) for the analysis and diagnosis of ESDs is evaluated in this study. In terms of statistical accuracy, specificity, sensitivity, precision, true positive rate, true negative rate, and *F*-score, the prediction accuracies of these models are compared to the artificial neural network (ANN). Bagged ensemble surpasses other modeling methods in tests, according to the results.

A unique approach for the classification of erythematous-squamous diseases in an integrated fuzzy neural structure [3] (multi-input and multi-output) based on the Takagi-Sugeno-Kang (TSK)-type rule. The inference system's overall performance was empirically assessed in terms of classification accuracy, yielding a total accuracy of 98.37%. When compared to the results of other algorithms developed by other academics in the same domain, it was clear that the proposed algorithm was far superior.

An optimized support vector machine (SVM) was used to create an automatic erythematous-squamous illness identification system [4]. The swarm optimization algorithm, elephant herding optimization, was used to modify the parameters of support vector machines. When compared to the accuracy of other ways found in the literature, the prepared method excels and produces superior outcomes. Multi-layer feed-forward was used in this research. To categorize dermatology diseases, a neural network was deployed to a dermatology dataset retrieved from the UCI repository site. Working with a neural network that uses backpropagation for classification and prediction issues, algorithms offer the best results. It also has the ability to generalize and be applied to real-world issues.

On 366 occurrences of erythematous-squamous illnesses datasets, the study [5] employed naive Bayes, multi-layer perceptron, and J48 decision tree induction to create predictive data mining models. The baseline prediction performance of the classifiers was also evaluated using tenfold cross-validation and a set of performance criteria. The naive Bayes algorithm fared best, with a precision of 97.4%, followed by the multi-layer perceptron with a precision of 96.6%, and the J48 with a precision of 93.5%. Author [6] used soft computing techniques such as artificial neural networks and support vector machines to create models. The performance of ANN is more accurate for two hidden layers than SVM with RBF kernel, according to the results analysis.

To categorize dermatology illnesses, a multi-layer feed-forward neural network [7] was deployed to a dermatology dataset retrieved from the UCI repository site. For classification and prediction tasks, an artificial neural network using a back-propagation algorithm delivers the best results. The author [8] used ANN model for diagnosing skin illness and achieved 90% accuracy. Rambhajani et al. [9] has surveyed on various machine learning algorithms on dermatology dataset and by comparing past studies, the author concluded that nearest neighbor and naive bayes algorithms performed well with 99.25% accuracy. The paper [10] describes a unique approach for erythematous-squamous illness differential diagnosis. Fuzzy weighted preprocessing, nearest neighbor (KNN)-based weighted preprocessing, and a decision tree classifier are used in the proposed method. Using 20-fold cross-validation, the used decision tree classifier, fuzzy weighted pre-processing decision tree classifier, and KNN-based weighted preprocessing decision tree classifier achieved classification accuracies of 86.18, 97.57, and 99.00%, respectively. Vijaya [11] focuses on non-melanoma skin cancer and classifies types using support vector machine (SVM) to predict disease types effectively. Preprocessed skin lesion photographs are used to extract chrominance and texture attributes, and a training dataset is created.

Voting feature interval (VFI5) is a new classification method created and applied to the problem of differential diagnosis of erythematous-squamous disorders. The domain comprises patient records with a known diagnosis [12]. The VFI5 classifier learns how to distinguish a novel case in the domain given a training collection of such records. VFI5 is a notion represented by feature intervals on each feature dimension individually. The VFI5 method uses real-valued voting to classify data.

Each feature participates in the voting process equally, and the predicted class is determined by the class that obtains the most votes. The VFI5 classifier's performance is empirically assessed in terms of classification accuracy and running time.

3 Dataset Description

The dataset used in this study is taken from UCI machine repository Güvenira and Demiröza [12]. This dataset is labeled (i.e., target values are present) and intended to classify various erythemato-squamous diseases [5]. There are 35 variables in this dataset, 34 of which are linear and 1 of which is nominal [1, 2]. Since the classes of the skin disease as mentioned in Table 1 are liable to share similar clinical features of scaling and erythema, thus identifying and diagnosing erythemato-squamous ailments can be challenging in dermatology. The six forms of skin illness include psoriasis, seborrhea, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra [2]. One of the most prevalent therapies for these disorders is biopsy (i.e., tissue examination) [5]. In the early stages of a disease, a disease may have the characteristics of another class of disease, which is another challenge that dermatologists face when diagnosing these diseases. Patients were first analyzed using 12 clinical features, then 22 histological traits were assessed using skin disease samples [2]. The samples were examined under a microscope to identify histological features. If any diseases are discovered in the family, the family history attribute in the domain dataset is set to 1 (one), and if none are discovered, the value is set to 0 (zero). The patient's age is used to determine age characteristics [2]. Every additional clinical and histological trait was assigned a numerical value between 0 and 3 (0 = no features; 1, 2 = comparative middle values; 3 = highest value). The dataset of erythemato-squamous diseases has 366 occurrences and 35 attributes. Table 2 [3] summarizes the contents of the aforementioned attributes.

4 Proposed Work

The concepts of different classifiers used in the proposed work need to be explained to acquire insight knowledge of functioning of models (classifiers) for their efficient implementation in the classification of dermatological diseases.

4.1 Artificial Neural Networks

Artificial neural network (ANN) [7] is a mathematical model that consists of number of artificial neurons interconnected each having associated with weights that can be

Table 2 Summary of erythematous-squamous disease dataset

Classes	Clinical attributes	Histopathological attributes
C1: psoriasis	f1: erythema	12: melanin incontinence
C2: seborrheic dermatitis	f2: scaling	13: eosinophils in the infiltrate
C3: lichen planus	f3: definite borders	14: PNL infiltrate
C4: pityriasis rosea	f4: itching	15: fibrosis of the papillary dermis
C5: chronic dermatitis	f5: Koebner phenomenon	16: exocytosis
C6: pityriasis rubra pilaris	f6: polygonal papules	17: acanthosis
	f7: follicular papules	18: hyperkeratosis
	f8: oral mucosal	19: parakeratosis
	f9: knee and elbow	20: clubbing of the rete ridges
	f10: scalp involvement	21: elongation of the rete ridges
	f11: family history	22: thinning of the suprapapillary epidermis
	f18: hyperkeratosis	23: spongiform pustule
	f34: age	24: munro microabcess
		25: focal hypergranulosis
		26: disappearance of the granular layer
		27: vacuolization and damage of basal layer
		28: spongiosis
		29: saw-tooth appearance of retes
		30: follicular horn plug
		31: perifollicular parakeratosis
		32: inflammatory mononuclear infiltrate
		33: band-like infiltrate

used for both classification and regression. The concept of neural network has been derived from the functioning and mechanism of human brain.

It functions with three layers, namely (i) input layer; that receives the input data, (ii) hidden layer; this layer is computational or processing layer, and (iii) output layer; it provides the final output [8] (Fig. 1).

The input layer transmits the information to the node of hidden layer and output layer, where actual processing takes place. The training phrase of the model starts by first assigning random values to weights for interlink between input variable to hidden node and hidden node to output node in feed-forward neural network. The output of each neuron is calculated by linear summation of inputs value and its corresponding weights and finally performing a linear or nonlinear activation function (i.e., sigmoid, tanh, ReLU, etc.) to control the output of model.

The activation functions on the output layer are chosen based on the nature of classification or prediction problem, i.e., either binary classification or multi-variate classification task. In case of neural networks, the architecture is drawn so as to apply activation functions on only hidden layer(s) and output layer (Fig. 2).

Different training algorithms are available such as Levenberg–Marquardt algorithm, which is one of the best-known algorithms in artificial neural network. The feed-forward neural network is used for training phrase where output of all nodes is calculated. Finally, output layer gives the overall output of the neural network.

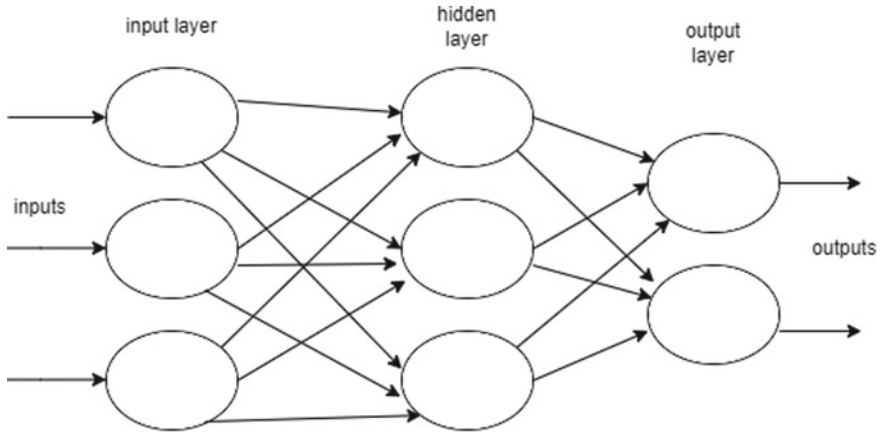
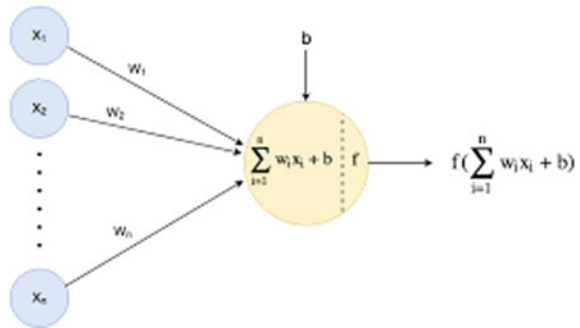


Fig. 1 Feed-forward neural network [6]

Fig. 2 Activation functions in neural networks [7]



Mathematically, the output is calculated as:

$$y = \text{activation function} \left(\sum_{i=0}^n x_i \cdot w_i + \text{bias} \right) \tag{1}$$

where x_i : i th input value; w_i : weight of the i th input

The difference between the actual output and the obtained output is the error of the model. In order to minimize this error, gradient descent technique is used to update the weights till the error gets minimized. The error is calculated using minimum square error as loss function

$$\text{Error} = \min(\text{actual output} - \text{obtained output})^2 \tag{2}$$

The process of error minimization is possible using backpropagation approach. In this paper, error calculation is done using mean square error (MSE).

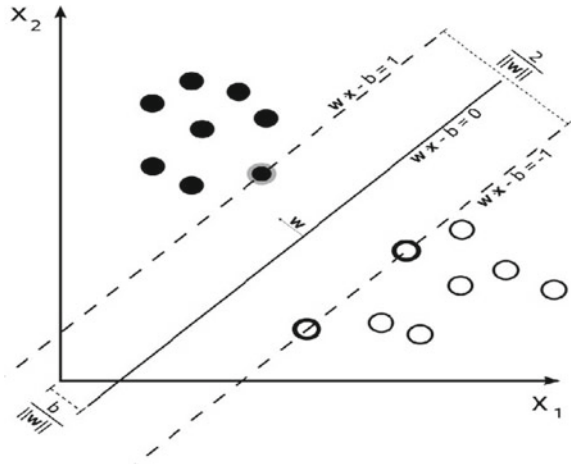
4.2 *K-Nearest Neighbor*

K-nearest neighbor (KNN) is another classifier model that does not require any explicit training for generalization of data points. It is one of the simplest classifiers that discovers the unidentified data points using previously known data points (i.e., nearest neighbor) and classifies them according to the voting system of their neighbors measured by a distance function. This distance function must minimize [10] the distance between the similar group of data points, while maximize the distance between dis-similar data points belonging to different classes. Many distance-based functions have been implemented such as Euclidean distance, Manhattan distance. The selection of the best distance function is one of the major factors that affect the KNN performance.

4.3 *Support Vector Machine*

Support vector machine is a supervised machine learning algorithm that supports both classification and regression analysis. It works well for binary classification where the hyperplane in a multi-dimensional space separates the different class levels. The decision factors for deciding the hyperplane are carried out by various kernel functions. The kernel functions play a significant role in transforming the input data into higher-dimensional form so that the decision boundary can be formed with the help of hyperplanes that classifies the data into distinct groups (Fig. 3).

Fig. 3 Maximum-margin hyperplane [6]



4.4 Proposed Model

Our proposed model is based on weighted averaging ensemble classifier technique. The ensemble classifier combines different classifiers such as artificial neural network (ANN), *K*-nearest neighbor (KNN), and support vector machine (SVM). The main objective of this proposed model is to improve the classification accuracy of the model using weighted averaging ensemble technique, where the higher weightage is assigned to the model whose performance is the best among other classifiers. Initially, the weight values (between 0 and 1) for the individual soft computing classifiers are assigned based on their accuracy value. The detailed working mechanism of the proposed model is explained using followed steps.

First step: Data normalization.

Second step: Splitting of dataset into training and testing data in the ratio 70:30.

Third step: Training phrase of three soft computing classifiers (ANN, KNN, SVM).

Fourth step: Evaluation of model on test data.

Fifth step: Accuracy obtained for each of the classifiers using formula.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Sixth step: Once accuracy is known for each of the individual classifiers, then accordingly, weightage would be assigned.

Seventh step: Highest weigh value is assigned to the one having best accuracy, and second highest weightage is given to the model having accuracy lower as compared to the best one. Likewise, weights are assigned to each one of the individual classifiers.

Eighth step: Now, since ensemble model is the combination of all three classifiers, so the weighted average technique would provide with best result in terms of accuracy.

The schematic representation of the proposed model is shown Fig. 4.

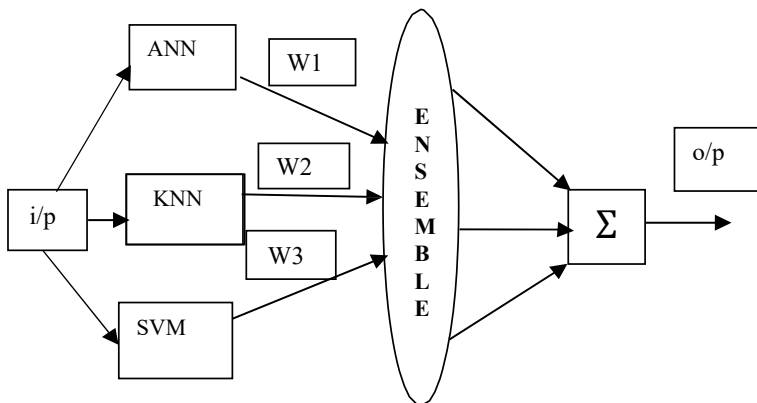


Fig. 4 Proposed model

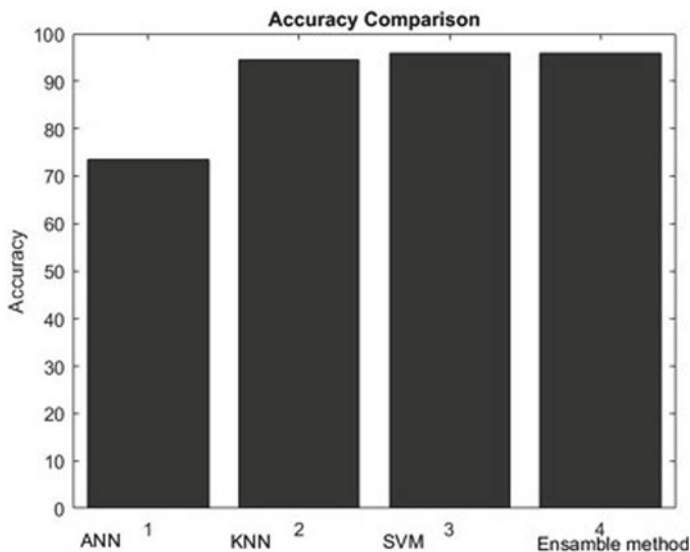


Fig. 5 Comparison of accuracy of existing models and proposed model

5 Experimental Results

The proposed model described in Sect. 4.4 was implemented in MATLAB. The ESD dataset, which included 366 patient records and was retrieved from the UCI repository (erythematous-squamous dataset description, 2016), was used for training and testing. To find the best-suited classifiers, three individual classifiers were tested utilizing data samples. The ensemble's classifiers were trained and tested using tenfold cross-validation, with ninefold for training and onefold for testing. In terms of statistical correctness, the individual models' prediction accuracies are compared to the ensemble classifier. The weighted ensemble outperforms other modeling methods, according to the findings of experiments. The ANN achieved 73.43% accuracy, KNN outperformed ANN by achieving 94.5054%, and SVM achieved 96% accuracy.

The ensemble classifier (i.e., proposed model) performed outstandingly by achieving 96.3% accuracy, thus making the ensemble classifier model, the best model for classification. The results obtained by our proposed model (i.e., ensemble classifier) are compared against all three individual classifiers as shown in Fig. 5.

6 Conclusion

This paper proposed an ensemble classifier model based on artificial neural network (ANN), K -nearest neighbor (KNN), and support vector machine (SVM) with a view to achieve better accuracy in diagnostic process of dermatological diseases.

erythematous-squamous dataset [2] obtained from UCI repository includes all 34 attributes and 366 instances. Experiments were conducted using individual classifiers (ANN, KNN, SVM), and finally our proposed model (i.e., ensemble model) is experimented and compared with baseline models. The accuracy obtained by the ANN is 73.43%, while KNN improved the classification with accuracy 94.5%, and SVM outperformed with accuracy 96%. The overall accuracy obtained by the ensemble model is 96.3%. Thus, the type of dataset taken for the study determined the competence and performance of data mining techniques by using ensemble techniques. The accuracy of the classification can be further improved by using convolutional neural networks (CNN) and transfer learning technique to improve the accuracy of the result.

References

1. Verma AK (2019) Classification of skin disease using ensemble data mining techniques. *Asian Pacific J Cancer Prev APJCP* 20.6
2. Elsayad AM (2018) Analysis and diagnosis of erythematous-squamous diseases using CHAID decision trees In: 2018 15th International Multi-Conference on Systems, Signals and Devices (SSD), IEEE
3. Idoko JB (2018) Fuzzy neural system application to differential diagnosis of erythematous-squamous diseases. *Cyprus J Med Sci* 3.2:90–98
4. Tuba (2017) Support vector machine optimized by elephant herding algorithm for erythematous-squamous diseases detection. *Proc Comput Sci* 122:916–923
5. Danjuma K (2015) Evaluation of predictive data mining algorithms in erythematous-squamous disease diagnosis. *arXiv preprint arXiv: 1501.00607*
6. Parikh (2015) Diagnosing common skin diseases using soft computing techniques. *Int J Bio-Sci Bio-Technol* 7.6:275–286
7. Patel (2017) An empirical study of applying artificial neural network for classification of dermatology disease. *Indian J Sci Technol* 10
8. Kabari LG (2009) Diagnosing skin diseases using an artificial neural network. In: 2009 2nd International conference on adaptive science & technology (ICAST), IEEE
9. Rambhajani M (2015) A survey on implementation of machine learning techniques for dermatology diseases classification. *Int J Adv Eng Technol* 8.2:194
10. Polat K (2006) The effect to diagnostic accuracy of decision tree classifier of fuzzy and k-NN based weighted pre-processing methods to diagnosis of erythematous-squamous diseases. *Digit Signal Process* 16.6:922–930
11. Immagulate I (2015) Categorization of non-melanoma skin lesion diseases using support vector machine and its variants. *Int J Med Imag* 3.2:34–40
12. Güvenira HA, Demiröza G (1998) Learning differential diagnosis of erythematous-squamous diseases using voting feature intervals. *Artif. Intell. Med.* 13:147–165