



Prediction of Skin Disease with Three Different Feature Selection Techniques Using Stacking Ensemble Method

Anurag Kumar Verma¹  • Saurabh Pal² 

Received: 13 October 2019 / Accepted: 5 December 2019 / Published online: 16 December 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Skin disease is the most common problem between people. Due to pollution and deployment of ozone layer, harmful UV rays of sun burn the skin and develop various types of skin diseases. Nowadays, machine learning and deep learning algorithms are generally used for diagnosis for various kinds of diseases. In this study, we have applied three feature extraction techniques univariate feature selection, feature importance, and correlation matrix with heat map to find the optimum data subset of erythemato-squamous disease. Four classification techniques Gaussian Naïve Bayesian (NB), decision tree (DT), support vector machine (SVM), and random forest are used for measuring the performance of model. Stacking ensemble technique is then applied to enhance the prediction performance of the model. The proposed method used for measuring the performance of the model. It is finding that the optimal subset of the erythemato-squamous disease is performed well in the case of correlation and heat map feature selection techniques. The mean value, slandered deviation, root mean square error, kappa statistical error, and area under receiver operating characteristics and accuracy are calculated for demonstrating the effectiveness of the proposed model. The feature selection techniques applied with staking ensemble technique gives the better result as compared to individual machine learning techniques. The obtained results show that the performance of proposed model is higher than previous results obtained by researchers.

Keywords Erythemato-squamous disease · SVM · Stacking · RMSE · KSE

Introduction

The skin of human body is the major organ, with a total 20 square feet covered area. The skin protects us from heat and cold and helps regulate body temperature. The disease related to skin

✉ Saurabh Pal
drsaurabhpal@yahoo.co.in

¹ Research Scholar, MCA Department, VBS Purvanchal University, Jaunpur, India

² Department of MCA, VBS Purvanchal University, Jaunpur, UP, India

is called erythemato-squamous disease. Identification and diagnosis are difficult in erythemato-squamous disease, because all the classes contribute to the same clinical properties: scaling and erythema, with minute changes. The different classes of erythemato-squamous disease are C1: psoriasis, C2: seborrheic dermatitis, C3: lichen planus, C4: pityriasis rosea, C5: chronic dermatitis, and C6: pityriasis rubra. Basic treatment and diagnosing these skin diseases is biopsy. In initial stage, a skin disease of one class may also contain the symptoms of other classes of the skin disease, which is the main problem of dermatologists when they perform the diagnosis of skin diseases. In the beginning, new patient was examined with 12 clinical attributes and if the symptom of disease is found, then the examination of 22 histopathological attributes was done for obtaining skin disease parameters. Histopathological attributes were examined by analyzing the parameters with the help of microscope [1].

The expert system developed with the help of machine learning methods for generating decisions in the medical applications is increasing, as data becomes effortlessly obtained from internet. In the past few decades, an extensive growth has been seen in medical field for predicting the disease with the help of dataset obtained from the previous patients by observing exterior symptoms and using laboratory tests without detailed internal examinations. The applications of machine learning techniques in medical decision support are now proving fruitful for those patients which have not enough resources to do medical tests. Machine learning algorithms are now assist doctors for finding the help to diagnose better. Machine learning methods help doctors in compiling complex diagnostic tests, finding information from various sources (images, clinical data, and scientific knowledge) [2].

Various works were done on the erythemato-squamous disease by different authors. A proposed model FELM for automatic detection of skin diseases combined the fuzzy logic approach and the machine learning techniques. The performance obtained by FELM is superior than other methods in case of accuracy prediction and time complexity. The accuracy obtained by FELM model was calculated 93% [3].

Badrinath et al. [4] uses introduced a hybrid Adaboost ensemble method for prediction of erythemato-squamous disease. Machine learning classifiers neural networks, support vector machines (SVMs), and ANFIS are used, then ensemble method AdaBoost is discussed, and 99.3% of classification accuracy is achieved. Therefore, the proposed hybrid models with AdaBoost used to predict erythemato-squamous disease with higher accuracy as compared to others.

Verma et al. [5] use six machine learning classification techniques to obtain the prediction of erythemato-squamous disease, and different ensemble methods Bagging, AdaBoost, and Gradient boosting are used to enhance the accuracy. A feature importance technique to choose relevant features is also used on erythemato-squamous disease optimal data subset and achieves accuracy of 99.68% by gradient boosting applied on radius neighbor classification.

A two-stage mixed feature selection technique is used to diagnose skin diseases. SVMs are used as classification tools, extended sequential forward search, sequential forward floating search, and sequential backward floating search as searching tools. Normalized F scores are used to find the significance of each attribute. The developed model was claimed to achieve higher accuracy as in comparison to previous studies [6].

Übeyli and Doğdu [7] experimented with five classes of erythemato-squamous disease (excluding pityriasis rubra class) and the k-means clustering classifier was developed to find the erythemato-squamous diseases with 33 attributes and five classes. The classification accuracy of the k-means clustering was calculated as 94.22%. The k-means clustering algorithm can be used considering the misclassification rates in finding erythemato-squamous diseases.

A computer-aided model was developed for the analysis of skin disease. In this model, 22 attributes are used extend it with other parameters. This expert system predicts eight different skin diseases classes and can help doctors for predicting the disease confidence. The data for the study was collected from a limited region which will help to predict the demographic dependence of the disease. The best prediction model was obtained with algorithm developed using J48 [8].

Ozcift and Gulten [9] discussed a rotating forest integrated decision tree that encapsulates the best first search strategy. The wrapper uses positive selection to select the best subset on the erythematous squamous disease dataset. Machine learning techniques are used to calculate the discriminative power of selected features and a bagging algorithm is used to assess the diversity of training data. Based on the characteristics of the rotation forest integration algorithm, the accuracy reaches BNET 98.91% and SL 98.64%.

In order to find the optimal feature value of KSVM for diagnosis of erythematous squamous disease, a new method based on catfish binary particle swarm optimization algorithm was proposed. In order to obtain an optimal subset of features, the AR method is applied as feature reduction in both the training and testing phases. In addition, considering the RBF kernel can improve the performance of the classification. Compared with other methods such as AR-MLP and pure SVM, the experimental results show that the proposed method is more accurate and can achieve an accuracy of 99.09% [10].

Using artificial neural networks (ANN), a framework for diagnosing erythematous squamous disease was developed. The developed system is capable to achieve a high success rate using the artificial neural network technique. Accuracy of proposed model was achieved 90% [11].

The new mixed feature selection technique obtained better classification accuracy with 10 attributes out of 34 attributes from erythemato-squamous diseases dataset. The optimal feature subset then trained with different machine learning techniques and accuracy obtained by decision tree (CART) 95.62%, RBF neural networks 97.26%, SMO-poly kernel 98.36%, and RBF kernel 98.08% [12].

Almarabeh and Amer [13] discussed application of different machine learning techniques for prediction of various type of disease. They calculated the accuracy for heart disease, breast cancer, lung cancer, and diabetes, and in skin disease, the best accuracy is ANN (97.17%).

Maryam et al. [14] proposed an erythematous squamous disease detection model using mixed feature selection and multi-class support vector machines. The combination of these two methods takes advantage of the filter and wrapper approach, in which a level-based chi-square is used as an evaluation criterion and a genetic algorithm is determined to find the best feature subset. Training and test set for selecting features through multiple types of SVMs. The results show that the proposed model has 18 features and has a high accuracy rate (99.18%).

Different machine learning techniques for skin disease prediction used are correlation and regression tree (CART), SVM, decision tree (DT), random forest (RF), and GBDT to obtain the prediction of skin disease. The best accuracy find is 95.90% from GBDT. A multi-model ensemble method is then applied to combining these five data mining technique to get the highest accuracy of 98.64% [15].

A model based on logistic regression, linear discriminate analysis, k-nearest neighbor, classification and regression tree, Gaussian Bayesian, and support vector machine was developed to classify erythemato-squamous disease. Then, four different ensemble machine learning algorithms boosting AdaBoost and gradient boosting and random forests and extra trees are used to improve the performance of model. The maximum accuracy achieved was 98.64% in gradient boosting [16].

Machine learning algorithms and their abbreviations used in this study are shown in Table 1

Table 1 Machine learning techniques and their abbreviations

Sr. no.	Name of algorithm	Abbreviation used
1	Gaussian Naïve Bayesian classifier	NB
2	Decision tree classifier	DT
3	Support vector machine	SVM
4	Random forest classifier	RF

Methods

In this study, we applied three different feature selection methods: (1) univariate feature selection, (2) feature importance, and (3) correlation matrix with heat map on skin disease dataset obtained by UCI machine learning repository. Skin disease dataset contains of 34 variables; by applying these three feature selection techniques, we obtained 15 most important features (attributes) and obtained new optimum subsets of skin disease dataset. Then, we have applied four machine learning techniques (1) Gaussian Naïve Bayesian classifier, (2) decision tree classifier, (3) support vector machine, and (4) random forest classifier to train the optimum subset of skin disease dataset. The four machine learning algorithms predictions are then improved using stacking ensemble methods.

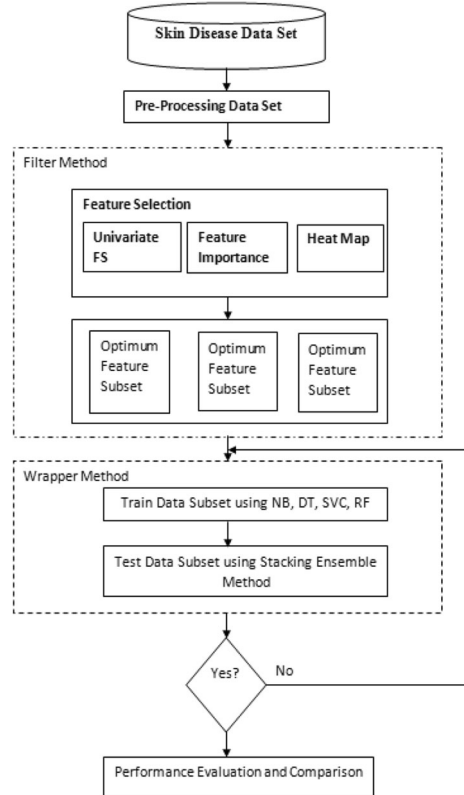
The predictions obtained by three different feature selection methods are compared to choose the best feature selection techniques and best prediction accuracy. The whole proposed methodology used in this research paper is described in Fig. 1.

Dataset Analysis

The database is taken from the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml>). The database was collected for examination of skin diseases and to classify various types of erythematous-squamous diseases. The database contains 34 attributes, 33 of which are linear, and 1 attribute is nominal. The family history (f11) is set to 1 if any diseases are found in the family of patient, and if not found in the family, then the value is 0. The all other remaining attributes (both clinical and histopathological) were assigned a value from the range 0 to 3 (0 = for absence of disease; 1, 2 = comparative intermediate values for disease; 3 = highest value). There are six classes of erythematous-squamous disease, with 366 instances and 34 features as shown in Table 2.

Feature Selection

Machine learning algorithms work rule is defined as, if we send garbage in, we will only let the garbage out. Garbage means error in data or unnecessary number of features. Feature selection is especially more important when the number of features is high. We need not use every attribute for applying any classification algorithm. We can measure the efficiency of algorithm by putting only those features that are really important for prediction. It has been mentioned in many research papers that feature subsets provide better prediction than complete feature datasets of the same algorithm. But if the dataset already contains only important attributes, then feature selection method is not necessary and the use of feature selection technique may not give better results.

Fig. 1 System framework of the proposed system**Table 2** Erythematous disease dataset [1]

Classes (no. of instances)	Clinical	Histopathological attributes
C1: Psoriasis (112)	f1: Erythema	f12: Melanin incontinence
C2: Seborrheic dermatitis (61)	f2: Scaling	f13: Eosinophils in the infiltrate
C3: Lichen planus (72)	f3: Definite borders	f14: PNL infiltrate
C4: Pityriasis rosea (49)	f4: Itching	f15: Fibrosis of the papillary dermis
C5: Chronic dermatitis (52)	f5: Koebner phenomenon	f16: Exocytosis
C6: Pityriasis rubra (20)	f6: Polygonal papules	f17: Acanthosis
	f7: Follicular papules	f18: Hyperkeratosis
	f8: Oral mucosal	f19: Parakeratosis
	f9: Knee and elbow	f20: Clubbing of the rete ridges involvement
	f10: Scalp involvement	f21: Elongation of the rete ridges
	f11: Family history	f22: Thinning of the suprapapillary epidermis
	f34: Age	f23: Spongiform pustule
		f24: Munro microabscess
		f25: Focal hypergranulosis
		f26: Disappearance of the granular layer
		f27: Vacuolization and damage of basal layer
		f28: Spongiosis
		f29: Saw-tooth appearance of rete ridges
		f30: Follicular horn plug
		f31: Perifollicular parakeratosis
		f32: Inflammatory mononuclear infiltrate
		f33: Band-like infiltrate

The main reasons to use feature selections are:

- It reduces the classification algorithm training time.
- It reduces the model complexity and make easy to interpret.
- It enhances the prediction accuracy of classifier algorithm if the accurate subset is selected.
- It decrease over-fitting.

In this paper, we have used three feature selection techniques to select 15 important features for training out reduced subsets of main dataset. These techniques are described below.

Univariate Feature Selection

Statistical tests can be used to select those attributes that have the significant relationship with the target attribute. This univariate feature selection method uses the chi-squared (χ^2) test for non-negative attributes to select 15 best features from the skin disease dataset.

Chi-square test is applied for categorical attributes in a dataset. Chi-square values are calculated between each attribute and the target attribute, and we can choose required number of attributes with best Chi-square values. Chi-square score is given by the following:

$$\chi^2 = \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}}$$

Here:

Observed frequency = Number of observations of a class

Expected frequency = Number of expected observations of a class if no relationship exists between the attribute and the target attribute.

The important 15 features obtained by univariate feature selection method and their chi-square values are shown in Table 3.

Table 3 Univariate feature important table

Features	Chi-square values
f33: Band-like infiltrate	759.365
f31: Perifollicular parakeratosis	690.579
f15: Fibrosis of the papillary dermis	688.000
f27: Vacuolization and damage of basal layer	671.800
f6: polygonal papules	669.667
f29: Saw-tooth appearance of rete ridges	667.742
f12: Melanin incontinence	604.333
f25: Focal hypergranulosis	577.996
f8: Oral mucosal	563.500
f30: Follicular horn plug	552.411
f7: Follicular papules	537.371
f22: Thinning of the suprapapillary epidermis	513.184
f20: Clubbing of the rete ridges involvement	513.124
f34: Age	429.796
f21: Elongation of the rete ridges	401.305

Feature Importance

The importance of the feature is calculated by reducing the impurity of the node and weighting it to obtain the probability that the impurity will reach the node. The probability of a node is calculated by dividing the number of observations arriving on the node by total number of observations. The high value of a feature shows the importance of that feature in feature importance method. Then, the importance of each attribute on the decision tree is evaluated as follows:

$$\hat{f}_i = \frac{\sum_j n_{ij}}{\sum_k n_{ik}} \quad \text{where } j = \text{node } j \text{ split on feature } i \text{ and } k \in \text{all nodes}$$

where \hat{f}_i = the importance of feature i

n_{ij} = importance of node j

These values are then normalized by dividing the sum of all feature importance values between 0 and 1:

$$\text{norm}\hat{f}_i = \frac{\hat{f}_i}{\sum_j \hat{f}_j} \quad \text{where } j \in \text{all features}$$

At the “random forest” level, the feature importance of the last element is the average of all trees. Calculate the sum of the importance values of the features on each tree and divide by the total number of trees:

$$\text{RF } \hat{f}_i = \frac{\sum_j \text{norms } \hat{f}_{ij}}{T} \quad \text{where } j \in \text{all trees}$$

where RF = the importance of feature i calculated from all trees in the Random Forest model

$\sum_j \text{norms } \hat{f}_{ij}$ = the normalized feature importance for i in tree j

T = total number of trees

The top most important 15 features obtained by feature importance method and their node probability using random forest values are shown in Fig. 2.

Correlation Matrix with Heat Map

Correlation shows how the attributes are associated to each other or the target attribute. Correlation can be positive (increase in value of an attribute increases the value of the target

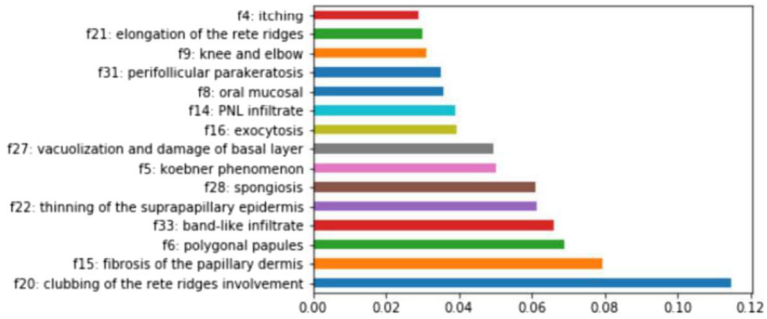


Fig. 2 Important attributes using feature importance

attribute) or negative (increase in value of a attribute decreases the value of the target attribute). It is easy to recognize which attributes are most related to the target attribute using heat map. The values of correlation coefficient lie between -1 and 1 .

- The value near to 0 shows weak correlation (0 means no correlation)
- The value near to 1 shows strong positive correlation
- The value near to -1 shows strong negative correlation

The heat map is shown in the Fig. 3.

We will only choose those attributes which have value of correlation coefficient more than 0.4 (absolute value only) and the best 15 features are listed in Table 4.

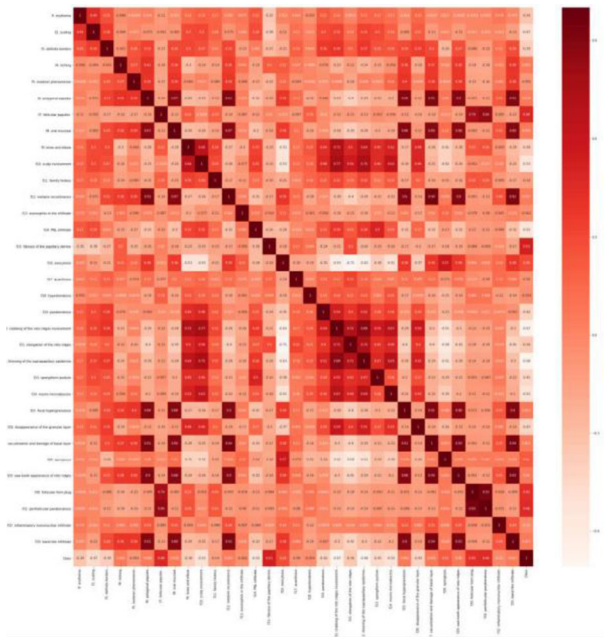


Fig. 3 Heat map

Machine Learning Classifiers

Ensemble methods are used to combine multiple classifiers predicted values into a single value to generate a strong model. Mainly, there are two types of ensemble methods—combining multiple models of same type and combining multiple models of different type. In this study, we have used staking ensemble technique so we have used combining multiple model of different type and therefore four different types of models are used namely.

Gaussian Naïve Bayesian Classifier

Gaussian Naïve Bayesian algorithms are used for classification. The probability of the attributes is assumed to be Gaussian:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

The values of σ_y and μ_y are calculated by maximum likelihood [17].

Decision Tree Classifier

Decision tree classifier generates classification or regression models in the shape of a structured tree. It divides a dataset into various smaller data subsets, and this forms associated decision tree. The end result is a tree with decision nodes which has two or more branches and leaf nodes which represents a classification or decision. The highest decision node in the tree represents to the best predictor called the root node. Decision trees can use categorical data and digital data [18].

Table 4 Correlation and heat map feature important table

Features	Correlation values (absolute)
f22: Thinning of the suprapapillary epidermis	0.684864
f20: Clubbing of the rete ridges involvement	0.668781
f14: PNL infiltrate	0.550195
f10: Scalp involvement	0.533208
f15: Fibrosis of the papillary dermis	0.526976
f24: Munro microabscess	0.520995
f7: Follicular papules	0.477813
f2: Scaling	0.468688
f31: Perifollicular parakeratosis	0.457025
f23: Spongiform pustule	0.449407
f30: Follicular horn plug	0.431972
f26: Disappearance of the granular layer	0.427348
f19: Parakeratosis	0.420742
f27: Vacuolization and damage of basal layer	0.416742
f6: Polygonal papules	0.410342

Support Vector Machine

Support vector machines (SVMs) are supervised learning algorithms used for classification, regression, and outlier detection that analyze data used for classification and regression analysis. Support vector machines (SVMs) are discriminant classifiers that are correctly defined by separate hyperplanes. The SVM uses labeled training data (supervised learning), and the algorithm outputs the best hyperplane, which classifies the new records [18].

Random Forest Classifier

A random forest classifier is a supervised learning technique and can be used for classification and regression analysis. This algorithm is most simple and flexible to use. A forest is collection of various trees. If high number of trees is present, then the forest is more robust. Random forests randomly select data to create decision trees and give prediction from each tree and choose the best solution by use of voting technique. It also provides an attractive excellent display of the feature importance [19].

Stacking Ensemble Technique

Stacking is an ensemble method for joining multiple classification models of different types through a meta-classifier. The individual four classification models NB, DT, SVM, and RF are trained using complete training set; then, the meta-classifier is fitted based on the outputs “meta-features” of the individual classification models in the ensemble techniques. The meta-classifier can either be trained on the predicted class labels or probabilities from the ensemble techniques. The stacking method is shown in Fig. 4.

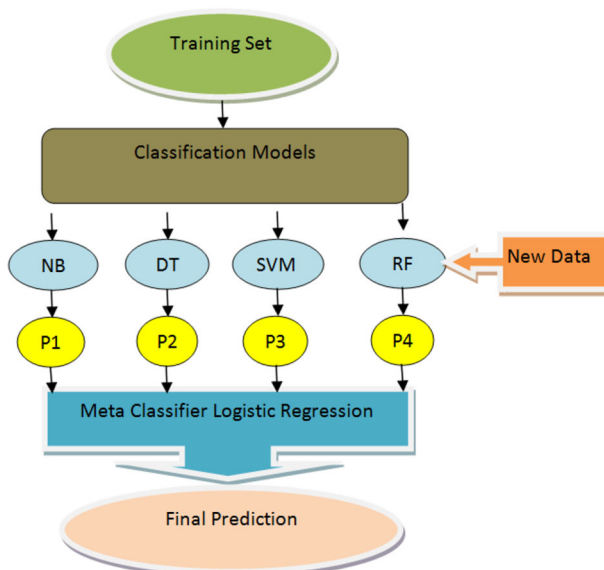


Fig. 4 Stacking ensemble technique

Results

We performed three experiments for finding the predictions for skin disease datasets using three different feature selection techniques. In first experiment, we choose the skin disease data subset taken from UCI Machine repository by choosing only 15 most important attributes using univariate feature selection methods, which consists 15 attributes and one target attribute “Class.” Second experiment has done with feature importance method, and third experiment has done using correlation matrix with heat map.

The important features obtained by three feature selection techniques used in three experiments are listed in Table 5 to illustrate the features selected.

We visualize the bar chart for distribution values of attributes as shown in Fig. 5. Figure shows the distribution of values for each attribute. Each feature bar chart shows the

Table 5 Important features selected by three different feature selection techniques

Features	Univariate feature selection	Feature importance	Correlation matrix with heat map
f1: Erythema			
f2: Scaling			✓
f3: Definite borders			
f4: Itching		✓	
f5: Koebner phenomenon		✓	
f6: Polygonal papules	✓	✓	✓
f7: Follicular papules	✓		✓
f8: Oral mucosal	✓	✓	
f9: Knee and elbow		✓	
f10: Scalp involvement			✓
f11: Family history			
f12: Melanin incontinence	✓		
f13: Eosinophils in the infiltrate			
f14: PNL infiltrate		✓	✓
f15: Fibrosis of the papillary dermis	✓	✓	✓
f16: Exocytosis		✓	
f17: Acanthosis			
f18: Hyperkeratosis			
f19: Parakeratosis			✓
f20: Clubbing of the rete ridges involvement	✓	✓	✓
f21: Elongation of the rete ridges	✓	✓	
f22: Thinning of the suprapapillary epidermis	✓	✓	✓
f23: Spongiform pustule			✓
f24: Munro microabscess			✓
f25: Focal hypergranulosis	✓		
f26: Disappearance of the granular layer			✓
f27: Vacuolization and damage of basal layer	✓	✓	✓
f28: Spongiosis		✓	
f29: Saw-tooth appearance of rete ridges	✓		
f30: Follicular horn plug	✓		✓
f31: Perifollicular parakeratosis	✓	✓	✓
f32: Inflammatory mononuclear infiltrate			
f33: Band-like infiltrate	✓	✓	
f34: Age	✓		

value of 366 instances, now they are distributed between the 4 values using univariate feature selection, feature importance, and correlation matrix with heat map techniques.

Python code is developed to calculate the prediction and to obtain different metrics and to evaluate the performance of different classifiers used on skin diseases data subset obtained by univariate feature selection technique, feature importance, and correlation matrix with heat map techniques. To measure the performance of different classifiers, the calculated values of mean, standard deviations, and accuracy are shown in Table 6.

Table 6 shows that the highest accuracy achieved is 89.18% using Gaussian Naïve Bayesian classification and the lowest accuracy obtained is 77.02% in Random Forest Classification in case of univariate feature selection technique. In feature importance technique, we achieve highest accuracy in case of SVM classifier and lowest accuracy in Naïve Bayesian classifier. The highest accuracy is obtained in random forest classifier 95.94% and minimum in decision tree classifier in case of correlation matrix with heat map.

The overall highest accuracy achieved is 97.29% by SVM using feature importance technique. This shows that the feature importance produced the best calculation for choosing the important features for skin disease data subset.

The comparison of different accuracy obtained by four classifiers, and three feature extraction techniques are shown by the box and whisker plot Fig. 6.

The accuracy and performance of any classifier and ensemble methods are related with some metrics. These metrics are calculated for measuring the performance of different features used in the feature selection techniques. We have calculated three different metrics root mean square error (RMSE), kappa statistic error (KSE) and area under receiver operating characteristics (AUC).

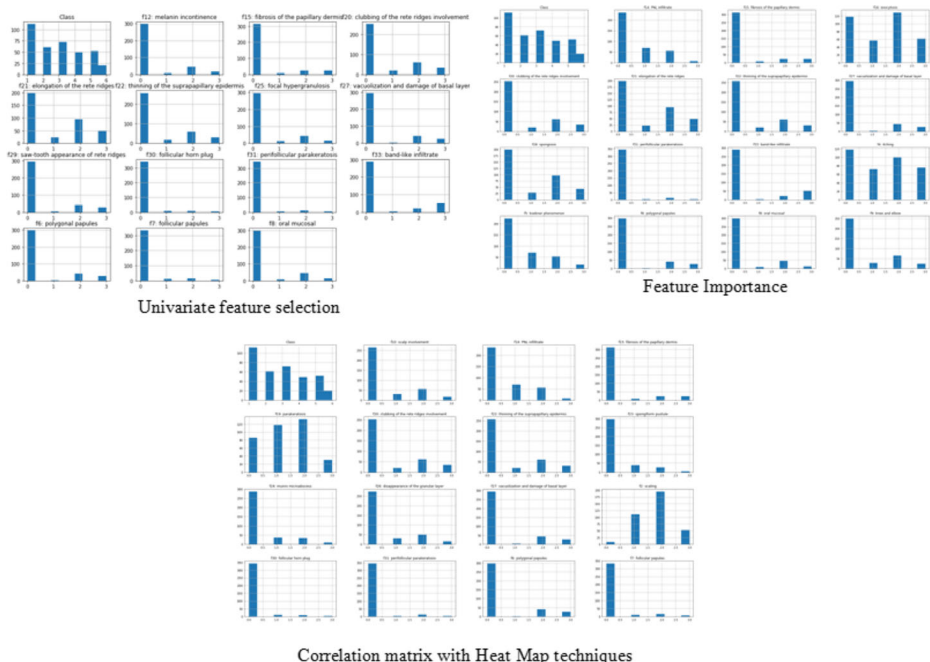


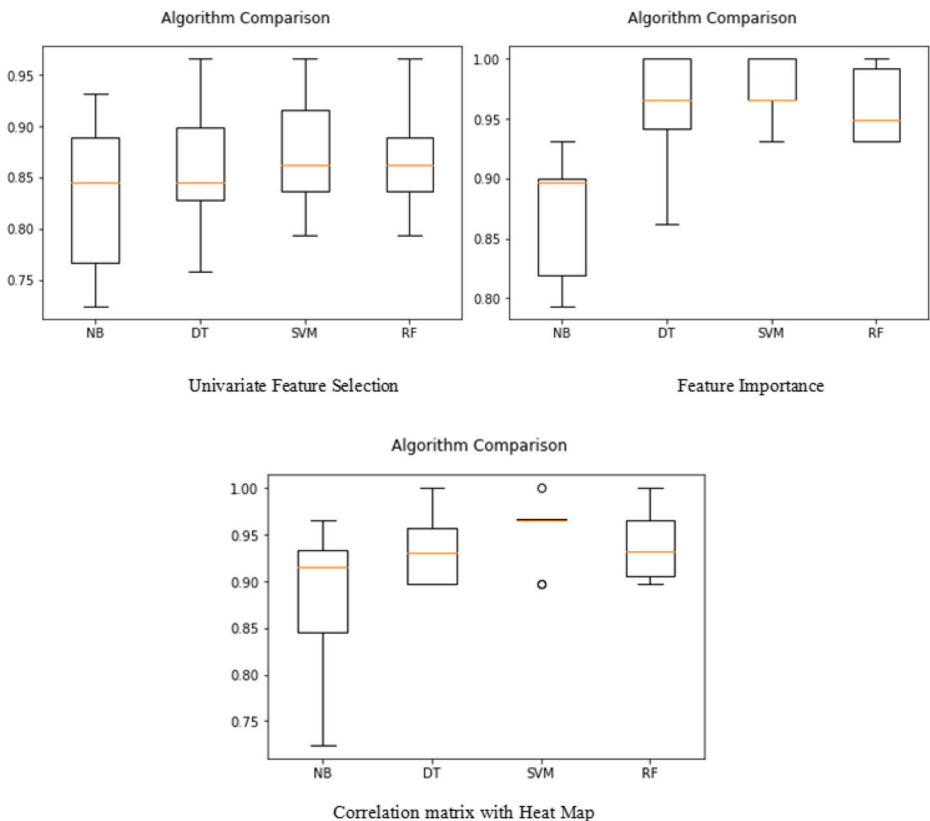
Fig. 5 Visualization of skin disease data subset obtained using three feature selection techniques

Table 6 Mean value, standard deviation, and accuracy

Algorithms	Univariate feature selection			Feature importance			Correlation matrix with heat map		
	Mean value	Standard deviation	Accuracy (%)	Mean value	Standard deviation	Accuracy (%)	Mean value	Standard deviation	Accuracy (%)
NB	0.8283	0.0712	89.18	0.8731	0.0538	91.89	0.8833	0.0806	93.25
DT	0.8591	0.0611	77.03	0.9624	0.0419	91.90	0.9348	0.0388	91.89
SVM	0.8728	0.0517	79.72	0.9727	0.0254	97.29	0.9554	0.0311	93.24
RF	0.8662	0.0524	77.02	0.9590	0.0296	94.59	0.9382	0.0338	95.94

Root Mean Square Error RMSE plays important role in the performance of classifiers. It is defined as the values predicted by a classifier and the values actually observed. The values of RSME for training and testing datasets are similar if we have developed the good classifier; in other case, if the RMSE values are much higher in testing of data than training data, the classifier developed is not good. The RMSE values is calculated using the formula

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$


Fig. 6 Accuracy of different classifier algorithms

Kappa Statistic Error The KSE is a metric which compares between calculated accuracy and expected accuracy. The kappa statistic error is calculated for both a single classifier and for ensemble classifiers. The kappa statistic error can be calculated with the help of following formula:

$$\text{KSE} = \frac{(\text{Observed Accuracy} - \text{Expected Accuracy})}{(1 - \text{Expected Accuracy})}$$

The value of KSE lies between -1 and 1 . If the calculate KSE value is almost 1 , then the performance of classifier is more accurate rather than by observation.

Area Under Receiver Operating Characteristics To understand the concept of confusion matrix, there are 4 types of results:

- True negative—Observation is negative and we predicted that the class is negative.
- False negative—Observation is positive but we predict that the class is negative.
- False positive—Observation is negative and we predicted that the class is positive.
- True positive—Observation is positive and we predicted that the class is positive.

From these 4 types of outcome, we calculate

$$\text{True Positive Rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{True Negative Rate (TNR)} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Now, AUC is calculated from the formula

$$\text{AUC} = \frac{1}{2} (\text{TPR} + \text{TNR})$$

Calculated RMSE values for different classification methods using stacking ensemble techniques using three different feature selection techniques are shown in Table 7.

The accuracy of a classification method is evaluated by confusion matrix using the following equation:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Table 7 RMSE, KSE, and AUC values for ensemble method

Algorithms	Univariate feature selection			Feature importance			Correlation matrix with heat map		
	RMSE	KSE	AUC	RMSE	KSE	AUC	RMSE	KSE	AUC
NB	0.0685	0.9841	0.983	0.0666	0.9864	0.985	0.0656	0.9897	0.992
DT	0.0798	0.9785	0.976	0.0659	0.9795	0.986	0.0668	0.9856	0.985
SVM	0.0696	0.9985	0.980	0.0621	0.9989	0.999	0.0645	0.9979	0.986
RF	0.0672	0.9798	0.972	0.0654	0.9689	0.993	0.0667	0.9865	0.997

In another term is can be represented as

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

The accuracy of stacking ensemble methods using four classifier algorithms is discussed in Table 8.

Here, we combine four classification algorithms by stacking classifier as a model. The results obtained after applying these three feature selection techniques to obtained performance of skin disease data subset methods and different values for confusion matrix, precision, recall, f1 score, and support.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall(Sensitivity) = \frac{TP}{TP + FN}$$

$$F1-Score = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right)$$

The support is defined as the number of observations of the true response obtained in that class.

The accuracy of staking ensemble techniques and different values is shown in Table 9.

Discussion

Feature selection acts very vital role in classification methods as unnecessary attributes used in the dataset are removed and the performance of classification models are improved. Using feature selection the dataset is reduced into a subset which also contains the enough information for classification as original dataset. In this research paper, we have performed three experiments univariate feature selection, feature importance, and correlation matrix with heat map to select 15 important features related to erythematous squamous dataset with 366 instances and 34 features which deals with six classes of skin disease. The three feature selection techniques gives different attributes as important features; they are not same because they are calculated using three different techniques chi-square, decision tree, and correlation

Table 8 Output of evaluating algorithms on the ensemble dataset

Ensemble stacking methods	Accuracy (%)			
	NB	DT	SVM	RF
Univariate feature selection	98.35	96.23	94.57	97.65
Feature importance	98.21	97.56	96.58	98.72
Correlation matrix with heat map	95.62	98.57	99.86	95.61
Time to build the model (in s)	0.02	0.02	0.04	0.03

Table 9 Accuracy, confusion matrix, and other values obtained by stacking ensemble method

Feature selection techniques	Accuracy	Confusion matrix	Values				
				Precision	Recall	f1 Score	Support
Univariate feature selection	98.35%	[23 1 0 0 0 0]	1	1.00	0.96	0.98	24
		[0 9 0 0 0 1]	2	0.38	0.90	0.53	10
		[0 0 11 0 0 0]	3	1.00	1.00	1.00	11
		[0 14 0 0 0 0]	4	0.00	0.00	0.00	14
		[0 0 0 0 11 0]	5	1.00	1.00	1.00	11
		[0 0 0 0 0 4]	6	0.80	1.00	0.89	4
		Avg/total		0.72	0.78	0.73	74
Feature importance	98.72%	[24 0 0 0 0 0]	1	1.00	1.00	1.00	24
		[0 8 0 1 0 1]	2	0.67	0.80	0.73	10
		[0 0 11 0 0 0]	3	1.00	1.00	1.00	11
		[0 4 0 10 0 0]	4	0.91	0.71	0.80	14
		[0 0 0 0 11 0]	5	1.00	1.00	1.00	11
		[0 0 0 0 0 4]	6	0.80	1.00	0.89	4
		Avg/total		0.93	0.92	0.92	74
Correlation matrix with heat map	99.86%	[23 1 0 0 0 0]	1	1.00	0.96	0.98	24
		[0 9 0 1 0 0]	2	0.75	0.90	0.82	10
		[0 0 11 0 0 0]	3	1.00	1.00	1.00	11
		[0 2 0 12 0 0]	4	0.92	0.86	0.89	14
		[0 0 0 0 11 0]	5	1.00	1.00	1.00	11
		[0 0 0 0 0 4]	6	1.00	1.00	1.00	4
		Avg/total		0.95	0.95	0.95	74

coefficient. The features selected by three feature selection techniques are presented in Table 6. It is clear that in univariate feature selection and correlation matrix with heat map techniques, 4 clinical and 11 histopathological features are selected. In feature importance techniques, 5 clinical and 10 histopathological features are selected.

After selecting the skin disease data subset using feature selection techniques, we have applied four classification algorithms NB, DT, SVM, and RF. We have chosen both linear and non-linear classification algorithms, so that the performance can be measure from both types of algorithms. These classification algorithms are evaluated on the basis of mean vale, slandered deviation, accuracy, root mean square error, kappa statistic error, and area under receiver operating characteristics. Standard deviation is a number used to inform how measurements for a feature are widen out from the average (mean) or expected value. The lower value of standard deviation represents that most of the numbers are closer to the mean value. A higher value for standard deviation means that numbers are more widen out. The values for mean, standard deviation, and accuracy are presented in Table 3. Table 3 shows that highest accuracy achieved is 89.18% for Gaussian Naïve Bayesian algorithm, and the lowest accuracy obtained is 77.02% in random forest classification in case of univariate feature selection technique. In feature importance technique, we achieve highest accuracy in case of SVM classifier and lowest accuracy in Naïve Bayesian classifier. The highest accuracy obtained in random forest classifier 95.94% and minimum in decision tree classifier in case of correlation matrix with heat map. The error measures, RMSE and KSE, and to evaluate the effectiveness of selected features area under receiver operating characteristics (AUC) value are evaluated in Table 4. If the values of RMSE are near to 0, the accuracy of used algorithm is more accurate. From this principle, it is easily find that the SVM classifier has best RMSE error for two out of four classifiers. If KSE value find for a classifier tends to 1, then performance of that classifier is

Table 10 Few results obtained by previous researcher which conduct experiment on skin disease

Author	Year	Method	Classification accuracy
Polat and Güneş [20]	2009	C4.5 and one-against-all	96.71%
Ubeyli [21]	2009	CNN	97.77%
Chang and Chen [22]	2009	decision tree neural network	80.33%
			92.62%
Ubeyli and Dogdu [7]	2010	K-mean clustering	94.22%
Lekka andMikhailov [23]	2010	Evolving fuzzy classification	97.55%
Xie and Wang [24]	2011	IFSFS and SVM	98.61%
Cataloluk and Kesler [25]	2012	Basic-KNN	94.4%
		Weighted KNN	96.36%
Olatunji and Arif [26]	2013	ANN	78.09%
		ELM	98.36%
Sharma and Hota [27]	2013	ANN	97.98%
		SVM	96.97%
		Ensemble	98.99%
Ravichandran et al. [3]	2014	FELM	93.00%
Olatunji and Arif [28]	2014	SVM	97.26%
		ELM	98.36%
Amarathunga et al. [29]	2015	AdaBoost	85% for eczema
		BayesNet	95% for impetigo
		J48, MLP (Naïve Bayes)	85% for melanoma.
Parikh et al. [30]	2015	ANN	97.17%
		SVM	94.04%
Badrinath et al. [4]	2016	FELM	98.21%
		ELM	95.87%
		Hybrid Adaboost	99.26%
Maghooli et al. [31]	2016	CART	93.69%
Zhou et al. [32]	2017	ANN	92.4%
Idoko et al. [33]	2018	CART	94.84%
		ARFCMC	75.96%
		ANFIS	.95.50%
		AEC	97.32%
		FNN	98.37%
Zhang et al. [34]	2018	ANN	96.8%
Verma et al. [5]	2019	Bagging	98.56%
		Adaboost	99.25%
		Gradient boosting	99.86%
Chaurasia and Pal [16]	2019	LR	97.94%
		LDA	96.22%
		KNN	85.57%
		CART	93.50%
		NB	89.02%
		SVM	92.10%
Verma et al. [15]	2019	CART	94.17%
		SVM	96.93%
		DT	93.82%
		RF	97.27%
		GBDT	96.25%
		Ensemble method	98.64%

more accurate. From this principle, the reliability of SVM classification is better to others. To evaluate the effectiveness of used classifiers, AUC value is calculated for each classifier. AUC values are given in Table 4.

In last, we have combined all four classifiers using stacking ensemble techniques to improve the performance of models on three feature selection data subset. Table 5 shows

the accuracy obtained by stacking ensemble method. Table 6 shows the accuracy, confusion matrix, and different values for six classes of target variable like precision, recall, f1 score, and support. Precision is the percentage of results which are relevant to target variable, whereas recall (sensitivity) is the percentage of total relevant results correctly classified by the algorithm. F1 Score is the weighted average of precision and recall. The f1 score gives the harmonic mean of precision and recall. Therefore, this score takes both false positives and false negatives into consideration. The support is the number of samples of the true response that lie in that class. The highest accuracy 99.86% obtained by staking ensemble method is in the case of correlation matrix with heat map feature selection techniques. This shows that the correlation coefficient proves the better result in feature selection technique.

To demonstrate the achievement of our proposed model, the results evaluated by present study were compared to other results obtained from the previous study. To compare the accuracy of this model, with others, a huge number of studies of skin disease using the same dataset but using another classifications methods and feature selection methods have been done. According to these studies, the same dataset of skin disease and test datasets are used. To demonstrate this, the efficiency obtained by this model is compared to other studies and we achieve the highest accuracy as compared to other studies as shown in Table 10.

Conclusion

Machine learning plays an important role in healthcare. The knowledge obtained by previous stored information from healthcare industry/organizations is used by various machine learning algorithms and is used to build up a machine which makes an effective decision to advance and develop healthcare industry/organizations. This paper uses different machine learning techniques for improving skin disease prediction. Since skin disease dataset consists of 35 attributes for evaluation of disease, but all the 35 attributes are not necessary to check the disease, so we use three feature selection techniques univariate feature selection, feature importance, and correlation matrix with heat map to obtain the best featured data subset. Four machine learning techniques NB, DT, SVM, and RF are used on reduced data subset to classify the prediction of skin disease. The best accuracy find among these different techniques is 97.29% by SVM. We have also calculated the various performance metrics like mean value, standard deviation, root mean square error (RMSE), kappa statistic error (KSE), and area under receiver operating characteristics (AUC). A stacking ensemble method is then applied to combining these four machine learning techniques; we get the highest accuracy of 99.86% in case of correlation matrix with heat map feature selection techniques. We get the highest accuracy in the literature available on skin disease dataset, and we recommend that the correlation coefficient obtained using heat map is best feature selection techniques.

Author's Contributions All authors contributed both the concepts and contents of this study. AKV provided the manuscript under supervised by SP. All authors also performed discussion intensively for contents improvement. All authors read and approved the final manuscript. Data Availability The datasets used in this paper are publicly online available as described in "Dataset Analysis" section.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no competing interests.

References

- Güvenir, H. A., Demiröz, G., & Ilter, N. (1998). Learning differential diagnosis of erythematous-squamous diseases using voting feature intervals. *Artificial Intelligence in Medicine*, 13(3), 147–165. [https://doi.org/10.1016/S0933-3657\(98\)00028-1](https://doi.org/10.1016/S0933-3657(98)00028-1).
- Barati, E., Saraee, M., Mohammadi, A., Adibi, N., & Ahamadzadeh, M. R. (2011). A survey on utilization of data mining approaches for dermatological (skin) diseases prediction. *Journal of Selected Areas in Health Informatics*, 2(3), 1–11.
- Ravichandran, K. S., Narayanamurthy, B., Ganapathy, G., Ravalli, S., & Sindhura, J. (2014). An efficient approach to an automatic detection of erythematous-squamous diseases. *Neural Computing and Applications*, 25(1), 105–114. <https://doi.org/10.1007/s00521-013-1452-5>.
- Badrinath, N., Gopinath, G., Ravichandran, K. S., & Soundhar, R. G. (2016). Estimation of automatic detection of erythematous-squamous diseases through AdaBoost and its hybrid classifiers. *Artificial Intelligence Review*, 45(4), 471–488. <https://doi.org/10.1007/s10462-015-9436-8>.
- Verma, A. K., Pal, S., & Kumar, S. (2019). Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study. *Applied Biochemistry and Biotechnology*, 1–19. <https://doi.org/10.1007/s12010-019-03093-z>.
- Xie, J., Lei, J., Xie, W., Shi, Y., & Liu, X. (2013). Two-stage hybrid feature selection algorithms for diagnosing erythematous-squamous diseases. *Health Information Science and Systems*, 1(1), 1–14. <https://doi.org/10.1186/2047-2501-1-10>.
- Übeyli, E. D., & Doğdu, E. (2010). Automatic detection of erythematous-squamous diseases using k-means clustering. *Journal of Medical Systems*, 34(2), 179–184. <https://doi.org/10.1007/s10916-008-9229-6>.
- Manjusha, K. K., Sankaranarayanan, K., & Seena, P. (2015). Data mining in dermatological diagnosis: a method for severity prediction. *International Journal of Computer Applications*, 117(11), 11–14. <https://doi.org/10.5120/20597-3102>.
- Ozçift, A., & Gulten, A. (2012). A robust multi-class feature selection strategy based on rotation forest ensemble algorithm for diagnosis of erythematous-squamous diseases. *Journal of Medical Systems*, 36(2), 941–949. <https://doi.org/10.1007/s10916-010-9558-0>.
- Giveki, D. (2012). Detection of erythematous-squamous diseases using AR-CatfishBPSO-KSVM. *Signal & Image Processing : An International Journal*, 2(4), 57–72. <https://doi.org/10.5121/sipij.2011.2406>.
- Kabari, L. G., & Bakpo, F. S. (2009). Diagnosing skin diseases using an artificial neural network. *ICAST 2009 - 2nd International Conference on Adaptive Science and Technology*, 187–191. <https://doi.org/10.1109/ICASTECH.2009.5409725>.
- Aruna, S., Nandakishore, V. L., & Rajagopalan, P. S. (2012). A hybrid feature selection method based on IGSBFS and naive Bayes for the diagnosis of erythematous-squamous diseases. *International Journal of Computer Applications*, 41(7), 13–18. <https://doi.org/10.5120/5552-7623>.
- Almarabeh, H., & Ehab, F. E. (2017). A study of data mining techniques accuracy for healthcare. *International Journal of Computer Applications*, 168(3), 12–17. <https://doi.org/10.5120/ijca2017914338>.
- Maryam, Setiawan, N. A., & Wahyunggoro, O. (2017). A hybrid feature selection method using multiclass SVM for diagnosis of erythematous-squamous disease. *AIP Conference Proceedings*, 1867(August). <https://doi.org/10.1063/1.4994451>.
- Verma, A. K., Pal, S., & Kumar, S. (2019). Classification of skin disease using ensemble data mining techniques. *Asian Pacific Journal of Cancer Prevention : APJCP*, 20(6), 1887–1894. <https://doi.org/10.31557/APJCP.2019.20.6.1887>.
- Chaurasia, V., & Pal, S. (2019). Skin diseases prediction: binary classification machine learning and multi model ensemble techniques. *Research Journal of Pharmacy and Technology*, 12(August), 3829–3832. <https://doi.org/10.5958/0974-360X.2019.00656.5>.
- Wang, N., Xu, H. L., Zhao, X., Wen, X., Wang, F. T., Wang, S. Y., Fu, L. L., Liu, B., & Bao, J. K. (2012). Network-based identification of novel connections among apoptotic signaling pathways in cancer. *Applied Biochemistry and Biotechnology*, 167(3), 621–631. <https://doi.org/10.1007/s12010-012-9704-x>.
- Banerjee, A. K., Ravi, V., Murty, U. S. N., Sengupta, N., & Karuna, B. (2013). Application of intelligent techniques for classification of bacteria using protein sequence-derived features. *Applied Biochemistry and Biotechnology*, 170(6), 1263–1281. <https://doi.org/10.1007/s12010-013-0268-1>.
- Behbahani, M., Nosrati, M., & Moradi, M. (2019). Using Chou's general pseudo amino acid composition to classify laccases from bacterial and fungal sources via Chou's five-step rule. *Applied Biochemistry and Biotechnology*, 1–14. <https://doi.org/10.1007/s12010-019-03141-8>.
- Polat, K., & Güneş, S. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications*, 36(2), 1587–1592. <https://doi.org/10.1016/J.ESWA.2007.11.051>.

21. Übeyli, E. D. (2009). Combined neural networks for diagnosis of erythematous-squamous diseases. *Expert Systems with Applications*, 36(3), 5107–5112. <https://doi.org/10.1016/J.ESWA.2008.06.002>.
22. Chang, C. L., & Chen, C. H. (2009). Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Systems with Applications*, 36(2), 4035–4041. <https://doi.org/10.1016/J.ESWA.2008.03.007>.
23. Lekkas, S., & Mikhailov, L. (2010). Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases. *Artificial Intelligence in Medicine*, 50(2), 117–126. <https://doi.org/10.1016/J.ARTMED.2010.05.007>.
24. Xie, J., & Wang, C. (2011). Using support vector machines with a novel hybrid feature selection method for diagnosis of erythematous-squamous diseases. *Expert Systems with Applications*, 38(5), 5809–5815. <https://doi.org/10.1016/J.ESWA.2010.10.050>.
25. Çataloluk, H., & Kesler, M. (2012). A diagnostic software tool for skin diseases with basic and weighted K-NN. *INISTA 2012 - International Symposium on INnovations in Intelligent SysTems and Applications*, 0–3. <https://doi.org/10.1109/INISTA.2012.6246999>.
26. Olatunji, S. O., & Arif, H. (2013). Identification of erythematous-squamous skin diseases using extreme learning machine and artificial neural network. *ICTACT Journal on Soft Computing*, 4(1), 627–632. <https://doi.org/10.21917/ijsc.2013.0090>.
27. Sharma, D., & Hota, H. (2013). Data mining techniques for prediction of different categories of dermatology diseases. *Journal of Management Information and Decision Sciences*, 16(2), 103.
28. Olatunji, S., & Arif, H. (2014). Identification of erythematous-squamous skin diseases using support vector machines and extreme learning machines: a comparative study towards effective diagnosis. *Transactions on Machine Learning and Artificial Intelligence*, 2(6). <https://doi.org/10.14738/tmlai.26.812>.
29. Amarathunga, A. A. L. C., Ellawala, E. P. W. C., Abeysekara, G. N., & Amalraj, C. R. J. (2015). Expert system for diagnosis of skin diseases. *International Journal of Scientific & Technology Research*, 4(1), 174–178.
30. Parikh, K. S., Shah, T. P., Kota, R. K., & Vora, R. (2015). Diagnosing common skin diseases using soft computing techniques. *International Journal of Bio-Science and Bio-Technology*, 7(6), 275–286. <https://doi.org/10.14257/ijbsbt.2015.7.6.28>.
31. Maghooli, K., Langarizadeh, M., Shahmoradi, L., Habibi-Koolae, M., Jebraili, M., & Bouraghi, H. (2016). Differential diagnosis of erythematous-squamous diseases using classification and regression tree. *Acta informatica medica : AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Društva za medicinsku informatiku BiH*, 24(5), 338–342. <https://doi.org/10.5455/aim.2016.24.338-342>.
32. Zhou, H., Xie, F., Jiang, Z., Liu, J., Wang, S., & Zhu, C. (2017). Multi-classification of skin diseases for dermoscopy images using deep learning. In *2017 IEEE International Conference on Imaging Systems and Techniques (IST)* (pp. 1–5). IEEE. <https://doi.org/10.1109/IST.2017.8261543>.
33. Idoko, J. B., Arslan, M., & Abiyev, R. (2018). Fuzzy neural system application to differential diagnosis of erythematous-squamous diseases. *Cyprus Journal of Medical Sciences*, 90–97. <https://doi.org/10.5152/cjms.2018.576>.
34. Zhang, X., Wang, S., Liu, J., & Tao, C. (2018). Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. *BMC Medical Informatics and Decision Making*, 18(S2), 59. <https://doi.org/10.1186/s12911-018-0631-9>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.