Classification of Skin Disease With Data Mining Based Ensemble Approach

Jitendra Singh Narwariya [1], A.K. Shrivas [2]
[1, 2] Dr. C. V. Raman University, Bilaspur(C.G.)

Abstract- *Data mining play very important role for pattern and classification of data from large amount of data set. In this research work we have proposed ensemble approach for classification of various types of diseases and compare the classification accuracy with proposed ensemble with model build time. In case of individual classifier, Multilayer Perceptron and Naïve Bayes gives same 97.54% of accuracy while proposed ensemble model i. e. (MLP+Naïve Bayes, MLP+SVM and MLP+Naïve Bayes+ SVM) gives same accuracy 97.81%. with 10-fold cross validation. The proposed ensemble model MLP+Naïve Bayes+ SVM is suggested for skin classification with low 5.77 building time.*

Keywords- Ensemble, Data Mining, Dermatology, Classification.

I. INTRODUCTION

Now a days, people are facing many problem related to health and different types of diseases. The doctors and medical science consistently try to diagnosis of health condition and try to better treatment to patients. In this piece of research work we have explored the data mining based classification techniques to identify and classify the health condition. In this research we have collected data set form UCI repository for training and testing purpose and analysis of classifier and proposed ensemble model.There are various authors have worked in the field of classification of skin disease. Dermatology is a study of skin disease that is extremely complex and difficult to diagnose, and ultimately may be a leading cause of skin cancer. Some of the authors have discussed as shown below:Rambhajani M. et al. (2015) [3] have suggested various machine learning techniques such as Data mining, Soft Computing, Hybrid method for classification of various dermatology diseases. Abitha R. et al. (2018) [4] have Naïve bayes classifier with symmetric uncertainty, PSO and hybrid feature selection technique for classification of various types of skin diseases where naïve bayes with hybrid feature selection gives 91.53% of accuracy as best classifier. Saxena A. et al. (2017) [5] have suggested K-NN with Filter-PSO based approach for feature selection. They have applied proposed methods with different data set and compared the performance of classifiers. Patel S. et al.(2017) [6] have suggested Artificial Neural Network for Classification of Dermatology Disease. Rambhojani M. et al. (2015) [7] have suggested Bayes net and Best First Search technique for classification of skin disease.

II. PROPOSED MODEL AND METHODOLOGY

The figure1 shows that classification process of dermatology disease data set where we have collected dermatology data set from UCI repository. We have partition the data set into 10-fold cross validation in which data set is divided is 10-fold where each partition used as testing sample and rest nine partition used training sample. This process applied 10 times to trained and test the classifier. We have compared the accuracy of individual and ensemble classifier and select the best classifier for classification of dermatology disease.

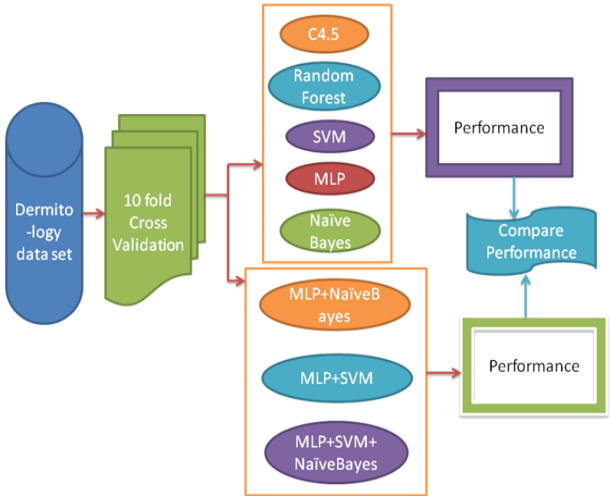

Figure1: Proposed model for classification of dermatology disease

Support Vector Machine (SVM)

Support vector machines [2] design a hyper planes or set of hyper planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. A SVM is a promising new method for classification of both linear and nonlinear data. SVM is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. SVM algorithms divide the n

dimensional space representation of the data into two regions using a hyper plane. This hyper plane always maximizes the margin between the two regions or classes. The margin is defined by the longest distance between the examples of the two classes and is computed based on the distance between the closest instances of both classes to the margin, which are called supporting vectors.

- **Naïve Bayes**

Classification algorithms (Han, J. et al., 2006) [2]have found a simple Bayesian classifier known as the Naive Bayesian classifier to be comparable in performance with decision tree and selected neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases.

- **Multilayer Perceptron(MLP)**

MLP (Pujari, A. K., 2001) [1] is a development from the simple perceptron in which extra hidden layers (layers additional to the input and output layers, not connected externally) are added. More than on hidden layer can be used. The network topology is constrained to be feed forward, i.e., loop-free. Generally, connections are allowed from the input layer to the first (and possible only) hidden layer, from the first hidden layer to the second and so on, until the last hidden layer to the output layer.

- **Ensemble Model**

Ensemble technique [2] denotes a system which employs a combination of methods and techniques from machine learning or intelligent techniques. An ensemble model is a combination of two or more models to avoid the drawbacks of individual models and to achieve high accuracy.

III. EXPERIMENT AND RESULT

Dermatology is a study of skin disease that is very difficult to diagnosis and classify by doctors and finally it may be cause of skin cancer. In this experiment, we have used WEKA data mining tool for analysis of dermatology data set with 10-fold cross validation. We have used different data mining based classifiers for classification of different types of skin. Firstly we have analyzed the different types of skin diseases with individuals and ensemble classifiers.In this research work we have used C4.5, Random Forest, Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Naïve Bayes classifiers. We have developed the ensemble to improve the accuracy compare to other individuals classifiers. We have achieved 97.81% of accuracy in case of ensemble classifiers (MLP+Naïve Bayes, MLP+SVM and MLP+Naïve Bayes+ SVM). We have suggested the ensemble classifier gives better accuracy compared to others classification techniques. Table 1shows that accuracy individual classifier and ensemble classifier (voting scheme) with 10 fold cross validation. Table 2 shows that other performance measures like True positive rate(TPR), False positive rate (FPR), precision, F-measure and ROC-areas.

Table 1: Accuracy individual classifier and ensemble classifier (voting scheme) with 10 fold cross validation:

Classifier	Correctly classified	Incorrectly classified	Time taken to build model (In second)
C4.5	95.90	4.09	0.01
Random Forest	96.99	3.00	0.14
MLP	**97.54**	2.45	6.57
SVM	97.26	2.73	0.09
Naïve Bayes	**97.54**	2.45	0.01
MLP+Naïve Bayes	**97.81**	2.18	6.85
MLP+SVM	**97.81**	2.18	6.85
MLP+Naïve Bayes+ SVM	**97.81**	2.18	5.77

Tabble2:Detailed Accuracy based on class:

Class	TPR	FPR	Precision	F-measure	ROC Area
Psoriasis	0.902	0.007	0.965	0.932	0.984
Seborgic Dermatitis	0.991	0	1	0.996	1
Lichen Planus	1	0	1	1	1
PityriasisRosea	1	0	1	1	1
Chronic Dermatitis	0.980	0.019	0.932	0.922	0.951
PityriasisRubra Pilaris	1	0	1	1	1
Weighted Avg	0.978	0.004	0.979	0.974	0.991

IV. CONCLUSION

The main objective of this research work is developed robust classifier which can classify various types of skin diseases with high accuracy. In this research work we have used different data mining based classification techniques for classification of skin disease. We have suggested new classifier i.e. ensemble of SVM, MLP and Naïve bayes which gives 97.81% of accuracy compare to other individual classifiers. We have recommended the ensemble classifier for classification of skin diseases. In future we will validate the proposed classifier with new data set and also apply the existing and new feature selection technique to improve the performance of classifier.

REFERENCES

[1] Pujari, A. K., "Data Mining Techniques", University press India, ISBN: 9788173716720, 2001.

[2] Han, J. and Kamber, M., "Data Mining Concepts and Techniques", Morgan Kaufmann. San Francisco. 2nd ed., ISBN: 13: 978-1-55860-901-3, 2006.

[3] Rambhajani M., Deepanker W. and Pathak N. , "A survey on implementation of machine learning Techniques for dermatology diseases Classification", International Journal of Advances in Engineering & Technology, vol.8, no. 2, pp.194-202, 2015.

[4] Abitha R. and Vennila M., "An efficacious feature selection based on SU-PSO for healthcare datasets", International Journal of Pure and Applied Mathematics, vol. 119, no.16, pp. 4095-4116,2018

[5] Saxena A. and Shrivas M.M. , "Filter – PSO based approach for Feature Selection", International Journal of Advanced Research in Computer Science, vol. 8,no. 5,pp.2063-2073,2017

[6] Patel S. and Patel A. (2017), "An Empirical Study of Applying Artificial Neural Network for Classification of Dermatology Disease", Indian Journal of Science and Technology, vol. 10, no. 17, pp. 1-7, 2017.

[7] Rambhajani M., Deepanker W. and Pathak N. , "Classification of Dermatology Diseases through Bayes net and Best First Search", International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 5,pp.116-119,2015.