Research Article **Bio Informatics**

International Journal of Pharma and Bio Sciences

ISSN 0975-6299

CONSRUCTION OF META CLASSIFIERS FOR APPLE SCAB INFECTIONS

A.KUMARAVEL[1] AND D.UDHAYAKUMARAPANDIAN[2]

[1]*Professor and Dean, Department of Computer Science and Engineering.Bharath University, Selaiyur, Chennai-600073, India.*
[2]*Assistant Professor, Department of Computer Science and Engineering. Bharath University, Selaiyur, Chennai-600073, India.*

ABSTRACT

Impact of infections on fruits like apple is very high if it is not predicted and acted upon. Classification as one of the major data mining methodologies can be applied effectively for this purpose. The objective of this paper is to check the learning algorithms for classification such examples based on selected dataset for apple scab . The main intention in this context is to deal with available data set for high accuracy. For this purpose Ada Boost, Bagging, Logit Boost models are built using an open source mining Weka under supervised learning algorithms. It is necessary to reduce the error before constructing the final models and thus the varying the parameters and number of iterations for training is carried out.

KEYWORDS: **Data mining, Classification, Meta classifiers, Base Classifiers, Apple Scab, Search Methods Bagging Boosting.**

A.KUMARAVEL
Professor and Dean, Department of Computer Science and Engineering.Bharath University, Selaiyur, Chennai-600073, India.

1. INTRODUCTION

In this paper we address the problem of infection prediction based on Primary infection data and such models contribute grower to determine his/her fungicide control options. Some fungicides can eradicate an infection if applied within a specified time of the start of an infection period. Infection is predicted by the Mills table system[7] for apple scab, as modified and adapted by A. L. Jones of Michigan State University. Once green tissue of apples is exposed in the spring, apples are vulnerable to scab infection. The apple scab infection model predicts if a wetting period was sufficient to initiate an infection. The mentioned region had shown zero yields due to these type of infections. Monitoring and controlling such losses physically is expensive. Hence researchers opt for management techniques augmented with predictions

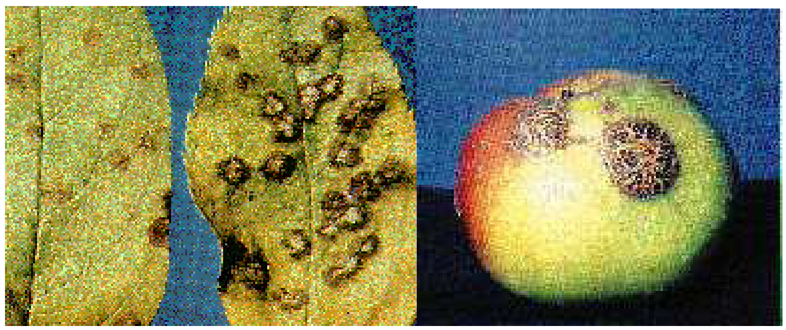

Figure 1
The apple scab fungus infection on leaves and fruits

Selecting this dataset which is clean and simple, and preprocessing for appropriate format we follow rest of the iterative steps as given in figure 2.

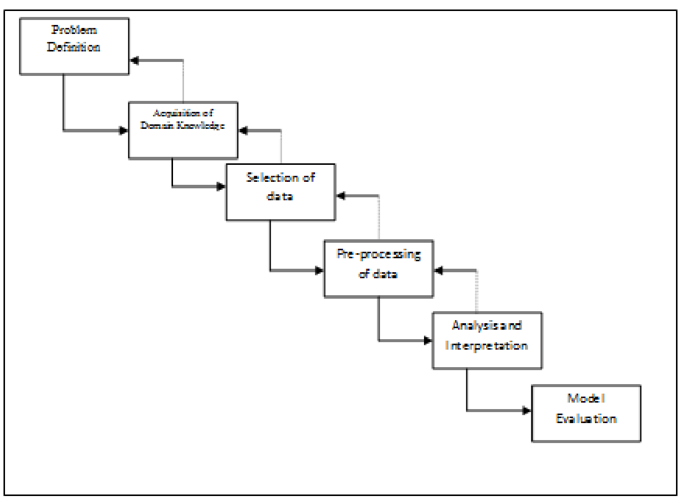

Figure 2
Iterative steps in Data Mining

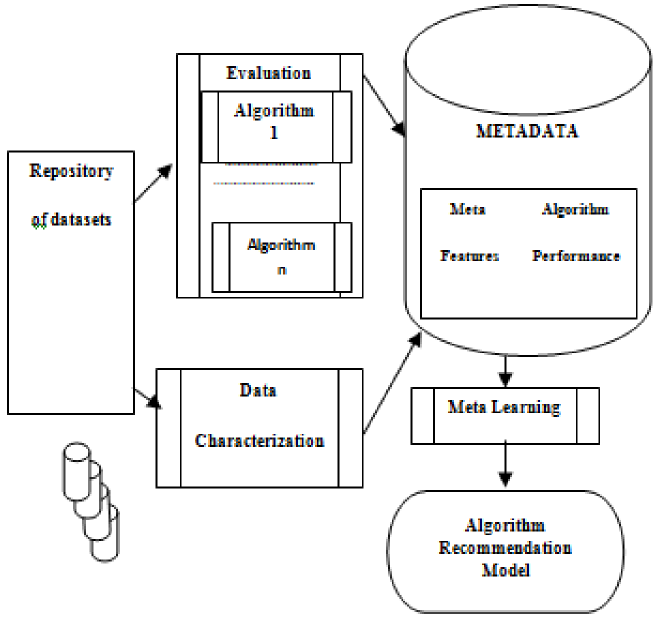

Figure 3
System for constructing meta- classifier

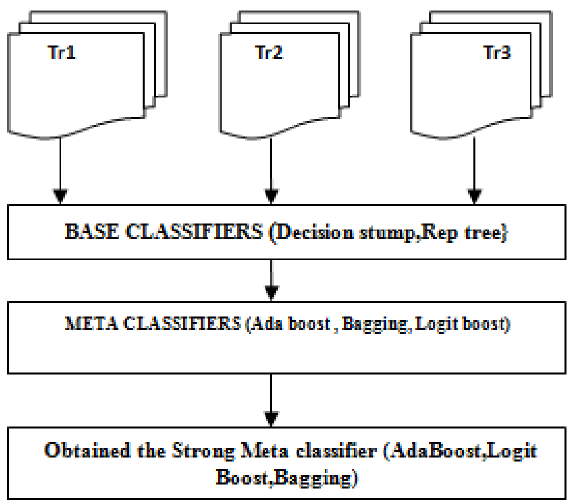

Figure 4
Flow to obtain a Strong Meta classifier

2. Meta Classifiers

Meta Classifier has showed spectacular success in reducing classification error from learned classifiers. These techniques develop a classifier in the form of a committee of classifiers. The committee members are applied to a classification task and their individual outputs combined to create a single classification. Meta learning approaches like AdaBoost, Logit boost, Bagging, Parameter Selection have received extensive attention. They are the recent methods for improving the predictive power of classifier learning systems.

Meta Classifier name	Category	Functions
Meta	AdaboostM1	Boost using the AdaBoostM1 method
	Bagging	Bag a classifier; works for regression too
	Logit Boost	Boost using performing additive logistic regression
Base	Decision Stump	Usually used in conjunction with a boosting algorithm.
	J48	For generating a pruned or unpruned C4.5 decision tree
	Bayes	Numeric estimator precision values are chosen based on analysis of the training data

2. EXPERIMENTAL ANALYSIS

In this section, we test the implementation efficiency of various methods and compare with whole dataset and the selected attributes. Weka tool is used to construct classification models.

2.1 DATASET

The datasets for these experiments are from [7]. The original data format has been slightly modified and extended in order to get relational format.

2.1.1 DATASET INFORMATION

The database of Apple scab this dataset describes a set for apples affected by two attributes like time and temperature in the range as shown in the table 1. The output is categorized into light, medium and high. The output class is denoting the possible category of infection affected. Number of Instances in this database is 138.

	Data Type	Range		Mean	Standard deviation
		Min	Max		
Temperature	Numeric	33	78	55.5	
Time	Numeric	7.9	96	25.369	19.89
Infection Class	Nominal	No .of classes:3			

Table 1

2.2 METHODOLOGY

The first step of our analysis was to reduce the high data dimensionality. For this purpose we used Weka tool for attribute selection based on various search methods[16] made in the attribute space as shown in table 1. We used factors which are selected after preprocessing as new predictors.

2.2.1 METHOD DESCRIPTION

Here we use three meta data classifications with different iterations in Ada Boost is decision stump and getting 57.2464 accuracy in my represented iterations so skipped this and Bagging is Rep tree classifier with maximum of 79.7101commonly for three iterations ,in Logit Boost is Decision stump classifier with maximum of 76.8116.

Ada Boost

This is meant for boosting a nominal class classifier method. Only nominal class problems can be tackled. Often dramatically improves performance, but sometimes over fits.

Bagging

This is meant for bagging a classifier to reduce variance. It can do classification and regression depending on the base learner.

Logit Boost

This is meant for performing additive logistic regression. This model enables to classify the dataset using regression method as base classifier and can be applied for more than binary class problemsThe following tables show the performance of the above methods in Weka which is a java implementation

Meta Classifier	Base Classifier	Parameter (No of iterations)	Accuracy	Maximum Accuracy
Bagging	Decision stump	10	56.5217	56.5217
		20	56.5217	
		30	56.5217	
	J48	10	85.5072	85.5072
		20	83.3333	
		30	84.7826	
	Bayes	10	39.8551	39.8551
		20	39.1304	
		30	39.1304	
	Random forest	10	84.7826	84.7826
		20	84.058	
		30	84.058	

Table 2
Bagging Classifier Iterations to get Maximum Accuracy

Meta Classifier	Base Classifier	Parameter (No of iterations)	Accuracy	Maximum Accuracy
Ada Boost	Decision stump	10	57.2464	57.2464
		20	57.2464	
		30	57.2464	
	J48	10	83.3333	84.058
		20	82.6087	
		30	84.058	
	Bayes	10	40.5797	40.5797
		20	40.5797	
		30	40.5797	
	Random forest	10	85.5072	85.5072
		20	85.5072	
		30	85.5072	

Table 3
Ada Boost Classifier Iterations to get Maximum Accuracy

Meta Classifier	Base Classifier	Parameter (No of iterations)	Accuracy	Maximum Accuracy
Multi BoostAB	Decision stump	10	57.2464	57.2464
		20	57.2464	
		30	57.2464	
	J48	10	81.8841	81.8841
		20	82.6087	
		30	83.3333	
	Bayes	10	40.5797	40.5797
		20	40.5797	
		30	40.5797	
	Random forest	10	84.7826	85.5072
		20	85.5072	
		30	85.5072	

Table 4
Multi Boost Classifier Iterations to get Maximum Accuracy

Meta Classifier	Base Classifier	Parameter (No of iterations)	Accuracy	Maximum Accuracy
RandomForest	Decision stump	10	57.9712	57.9712
		20	57.9712	
		30	57.9712	
	J48	10	78.2609	78.2609
		20	77.5362	
		30	76.8111	
	Bayes	10	37.6812	40.5797
		20	38.4052	
		30	39.1304	
	Random forest	10	81.8841	84.058
		20	84.0581	
		30	84.0581	

Table 5
RandomForest Classifier Iterations to get Maximum Accuracy

3. RESULTS

The Best meta classifier seen from above tables happens to be Bagging and the parameter values with the J48 as base classifier . This is confirmed after iterating the training at least three times for its convergence as shown in the tables 2-4.

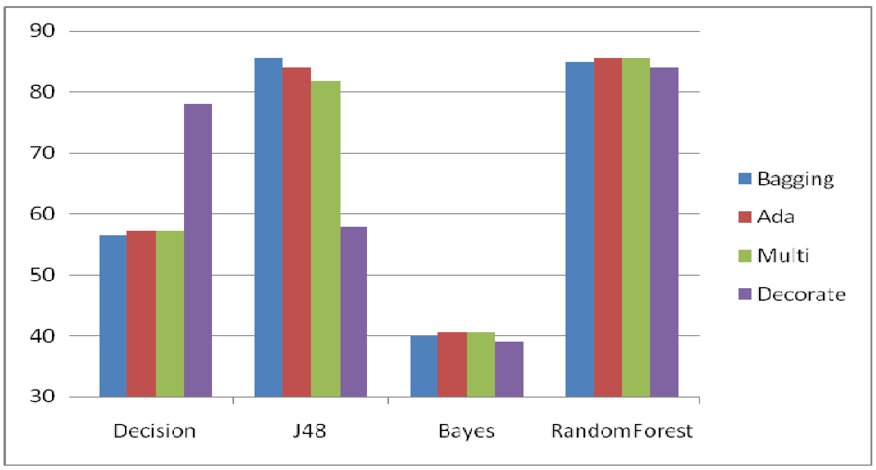

Figure 5
Comparison of meta classifier algorithms for accuracy

4. CONCLUSION

The above results improve the previously obtained accuracies and this study will help to formulate better schemes for preventing infections and enhancing the yields. However the size of the dataset is not large, future research can accommodate with either large dataset or aggregating small data set into bigger size.

ACKNOWLEDGEMENT

The authors would like to thank the management of Bharath University for the support and encouragement for this research work.

REFERENCES

1. https://www.waset.org/journals/waset/v68/v68-21.pdf world academy of science, engineering and technology, 2012.(Accessed on 29th july 2013)
2. Companion slides for the text by Dr. .H.Dunham, *Data Mining, Introductory and Advanced Topics*, Prentice Hall, 2002
3. Sourceboutwekahttp://www.cs.waikato.ac.nz/ml/weka/
4. A.Kumaravel, Pradeepa.R, Efficient molecule reduction for drug design by intelligent search methods Int J Pharm Bio Sci 2013 Apr; 4(2): (B) 1023 - 1029
5. A.Gelman, Y. S. Su, M.Yajima, J. Hill, M. Pittau, J. Kerman, and T. Zheng, “arm: Data Analysis Using Regression and Multilevel/Hierarchical Models,” R package version 1.5-02.://CRAN.Rproject.org/package=arm,2012.
6. L. Breiman, “ RandomForests,”in*Machine Learning*, vol. 45, pp. 5-32, 2001.
7. http://ipm.ncsu.edu/apple/chptr5.html(Accessed on 30th july 2013)
8. Dietterich, T. G., Jain, A., Lathrop, R., Lozano-Perez, T. (1994). A comparison of dynamic reposing and tangent distance for drug activity prediction.Advances in Neural Information Processing Systems, 6. San Mateo, CA: Morgan Kaufmann. 216--223.
9. A.Stensvand, T. Amundsen, L. Semb, D.M. Gadoury, and R.C. Seem. 1997. Ascospore release and infection of apple leaves by conidia and ascospores of Venturia inaequalis at low temperatures. Phytopathology 87:1046-1053.
10. Biggs, A. R. "Apple Scab" in Jones, A. L., and Aldwinckle, H. S. (editors). 1990. Compendium of apple and pear diseases. APS Press, St. Paul, MN.
11. http://weka.sourceforge.net/doc/weka/(Downloaded on 15th July 2013.