Feature Selection and Ensemble Meta Classifier for Multiclass Imbalance Data Learning

Mohd Shamrie Sainin[1], Rayner Alfred[1], Suraya Alias[1] and Mohamed A.M. Lammasha[2]

[1]Universiti Malaysia Sabah, Malaysia, {shamrie@ums.edu.my, ralfred@ums.edu.my, suealias@ums.edu.my}

[2]Universiti Utara Malaysia, Malaysia

ABSTRACT

The aim of this paper is to investigate the effects of combining feature selection and ensemble classifiers on the prediction performance in addressing the multiclass imbalance data learning . This research uses data obtained from the Malaysian medicinal leaf images shape data and three other large benchmark datasets in which six ensemble methods from Weka machine learning tool were selected to perform the classification task. These ensemble methods include the AdaboostM1, Bagging, Decorate, END, MultiboostAB, and RotationForest. In addition, five base classifiers were used; Naïve Bayes, SMO, J48, Random Forest, and Random Tree in order to examine the performance of the ensemble methods. There are two feature selection approaches implemented which are filter-based (CfsSubsetEval, ConsistencySubsetEval and FilteredSubsetEval) and wrapper-based (WrapperSubsetEval). The results obtained from the experiments show that although the performance accuracy is not much improved, however, with less number of attributes, the classifiers are able to achieve similar accuracy or slightly improved with less processing time. In knowledge management, the findings provide important insight of which algorithm is suitable for decision making when dealing with high dimensional and large data.

Keywords: Ensemble, feature selection, multiclass, imbalance, random forest, filteredsubseteval.

I INTRODUCTION

In multiclass data, the performance of a classifier degrades when imbalance data exists. Imbalanced datasets occur when one of the classes has significantly less examples compared to the other classes. An ensemble approach, which combines several single classifiers, is one of the methods used to solve imbalanced multiclass classification tasks. However, it remains a challenge to how the performance accuracy of the ensemble methods is influenced by the selected base classifiers coupled with different feature selection techniques used. The proposed ensemble classifier is combining the advantages data-level approach (feature selection) and algorithm level approach (ensemble classifiers). This may enable the enhancement of ensemble classifier performance in classifying multiclass imbalance data. Thus, the proposed framework of the ensemble classifier model consists of two parts which are combined together. These two parts include various feature selection techniques and various ensemble classifiers, including the various types of base classifiers used for the proposed ensemble classifier design. The aim of this paper is to investigate the effects of combining the feature selection and ensemble classifiers on the prediction performance in addressing the multiclass imbalance data learning. Thus, the objectives of this research include 1) Proposing and outlining the framework that combines several feature selection techniques and ensemble classifiers, which are available in Weka machine learning tool; 2) Evaluating the prediction performance of the proposed combination framework.

II RELATED WORK

Data is imbalance if there exists unequal distribution between its classes. Researcher in (Ding, 2011) stated that if imbalance ratio in a general classification problem is no less than 19:1 with the size of minority class is only 5% of the entire size of the data, and then it is called as a highly imbalanced classification problem.

The class imbalance has been recently discovered as an important issue in the machine learning and data mining. This problem takes place when there is no even distribution of the training data among classes, whereby, when a class is noticeably larger than the other, then, the data set becomes imbalance. The majority class usually tends to overshadow the standard classifier by overlooking the minority class examples, which is providing unwanted and also unsatisfactory classification performance. This value is based on the fact that a few conventional classifiers assume a balanced distribution of data and misclassification of cost between the classes. Therefore, the traditional algorithms need to be improved for a better handling of the imbalance data (López, Fernández, García, Palade, & Herrera, 2013).

Methods for imbalance problem can be categorized in two groups based on their approaches, namely data-level and algorithm-level. There are two methods that are associated with data-level method which are row-based (e.g. sampling) and column-based (feature selection).

Feature selection (also known as attribute subset selection or attribute reduction) is another important research issue in data mining and machine learning, and can be viewed as part of data pre-processing techniques. The technique works by selecting the subsets of the available features for application of a learning algorithm, with the aim to increase the performance of a classifier. The study reported in this paper considers feature selection to address the multiclass imbalance problems, mainly to find the subset of relevant features and to improve or to achieve similar prediction accuracy with reduced model build time.

Feature selection approaches have been applied in various studies. Among the method presented in the studies are comparisons of feature selection methods, e.g. Information Gain, Gain Ratio, etc. (Mohsin, Hamdan, & Bakar, 2014), and wrapper-based genetic algorithm (Barati, Abdullah, Mahmod, Mustapha, & Udzir, 2013).

The domain of Machine learning and Data mining, researches are greatly faced with the problem of multiclass imbalance. Although class imbalance problem has been extensively investigated. However, the issue of high dimensionality in data remain unsolved since high dimensionality is a common feature of class imbalance problem. In a study centered on Malaysian medicinal leaf identification (Sainin, Ahmad, & Alfred, 2016), leaf shape features generate enormous possibilities (high dimensional data) for leaf species. In this kind of identification, though, high accuracy may be recorded by a classifier in identifying the dominant leaf features (majority class) but there is greater tendency for the same classifier to record low performance in identifying the non-dominant features.

Weka's feature selection evaluators and search methods were investigated for their effect on the multiclass imbalance classification performance.

A. CfsSubsetEval, ConsistencySubsetEval and FilteredSubsetEval

These evaluators are in the type of filter-based feature selection. CfsSubsetEval or Correlation-based feature selection method (CFS) is concern with the hypothesis which contain features that are highly correlated with the class, but has no correlation with each other (Hall, 2000). It then will compute the correlation between attributes by first, applying the discretization and followed by the symmetrical uncertainty measure. In the study, CFS is proven to be comparable to wrapper feature selection method, but better on small datasets and overall running time.

ConsistencySubsetEval (CSE) is based on probabilistic approach to feature selection that is claimed to be simple and fast feature selection algorithm, thus guaranteed to find the optimal given the suitable resources (Liu & Setiono, 1996). The probabilistic approach called Las Vegas Algorithm (LVF) makes probabilistic choices as guide for the search of feature subset. In the experiment of this filter-based feature selection, it produces minimum features for the tested datasets with promising error rates. An analytical comparison on filter based feature selection has been conducted on CFS and CSE using decision tree classifier for accuracy measurement (Onik, Haq, Alam, & Mamun, 2015), where CFS provides less feature subset most of the time but CSE with BestFirstSearch strategy has higher performance.

FilteredSubsetEval (FSE) is simply a filter-based feature selection which available in Weka that running an arbitrary subset evaluator on the training data and produce the best feature subset (Cuaya et al., 2011).

B. WrapperSubsetEval

WrapperSubsetEval (Kohavi, 1995) in Weka is a feature selection method that using an induction algorithm as a blackbox (evaluator) for feature subset, where accuracy estimation technique is applied to measure how good is the features. In the study, the method is shown to improve significantly for some datasets with two induction algorithms namely decision tree and Naïve Bayes.

III METHODOLOGY

The methodology consists of Phase 1 (data acquisition), Phase 2 (feature selection), Phase 3 (training and testing of the combination of feature selection and ensemble classifiers), and Phase 4 (comparison). WEKA tool software program version 3.8 is adopted to implement the experiment study in this paper.

A. Phase 1: Data Acquisition

In data collection phase, a Malaysian medicinal leaf images were collected to construct the preliminary dataset in this domain. The dataset for the experiment is obtained from villages situated in the Perlis state where, 65 leaf samples are randomly selected from specified leaf species for the experimental data. The leaf sample size is selected in this preliminary study due to enormous time required to process the images without specific automated image processing. Table 1 is the list of leaf species selected in this research and Table 2 illustrates the description of the data. In addition, three benchmark datasets that comprise of imbalance multiclass (high imbalance ratio) as listed in Table 3.

Table 1. Sample Leaf species datasets.

Class	Leaf	Name	Train	Test
1		Cemumar (CM)	11	4
2		Kapal Terbang (KP)	12	4
3		Kemumur Itik (KI)	11	4
4		Lakom (LK)	5	4
5		Mengkudu (MK)	6	4
		Total	45	20

Table 2. Medicinal Leaf Dataset Information.

Description	Value #
#Examples	65
#Attributes	624
#Training	45
#Testing	20
#Majority	12
#Minority	5

Table 3. Benchmark Dataset With High Imbalance Ratio And Large Data.

Data	#S	#A	#C	Min	Max	Ratio	Previous result
Landsat	6345	36	6	56	1072	19	89.3%[1]
PageBlocks	5473	10	5	28	4913	175	97.3% [2]
Shuttle	58000	9	6	10	45586	4559	96.3%[3]

Note: #S: Number of samples, #A: Number of attributes, #C: Number of classes

[1] (Ghosh, Biswas, Sarkar, & Sarkar, 2014), [2] (Eschrich, Chawla, & Hall, 2002), [3] (Cohen, Cozman, Sebe, Cirelo, & Huang, 2004)

B. Phase 2: Feature Selection

Feature selection combined with ensemble classifier is implemented in this phase. Weka feature selection evaluators and search methods are investigated for their effect on multiclass imbalance classification performance. The experiments are carried out using three filter-based and one wrapper-based feature selection methods. Three filter-based methods (CFS, CSE, and FSE), each using search methods (BestFirst (BF), GeneticSearch (GS), GreedyStepwise (GSW) and LinearForwardSelection (LF)) and full training set approach. Wrapper-based feature selection (WrapperSubsetEval) in Weka is implemented using Naïve Bayes as the induction algorithm.

IV RESULT AND DISCUSSION

The first discussion in this section discusses the evaluation of the ensemble classifier with base classifier using original data and then followed by combining Weka feature selection over the data and ensemble classifier.

A. Performance of Ensemble Method using Original Data (All Features)

The classification performance of the classifiers when using all features with the ensemble classifier is listed in Table 4. The experiment uses seven ensemble methods and classifiers (Naïve Bayes (NB), SMO, Decision Tree (J48), Random Forest (RF), and Random Tree (RT) found in Weka using their best settings. The performance measures that were observed in each ensemble are F-measure and ROC, which is normally used in measuring the true positive rate as well as the accuracy of positive prediction among the classes (in multiclass). The reported results in Table 4 are the selected best classification performance in this paper. Each ensemble method used a single base classifier which produced up to 10 classifiers (as an ensemble) and produces the classification accuracy on one dataset.

Table 4. Ensemble Methods Classification Performance (In Percentage %).

Ensemble Classifier	Single Base Classifier					Avg.
	NB	***SMO***	***J48***	***RF***	***RT***	
AdaBoostM1	50.00	60.00	**70.00**	70.00	65.00	63.00
Bagging	50.00	40.00	55.00	60.00	65.00	54.00
Decorate	50.00	60.00	60.00	60.00	60.00	58.00
END	45.00	55.00	65.00	60.00	60.00	57.00
MultiBoostAB	55.00	60.00	**70.00**	70.00	65.00	64.00
RotationForest	60.00	55.00	55.00	65.00	60.00	59.00

According to the results, ensemble methods using AdaboostM1, MultiBoostAB and Stacking almost produce similar performance, which is 70% when using J48 or RF as base classifiers. The best base classifier in this experiment is the Random Forest with an average performance in all ensemble methods at 75%. Generally, MultiBoostAB, and AdaboostM1 performed better than the other ensembles tested in this experiment.

B. Feature Selection and Ensemble Methods Performance

In order to implement the ensemble method with feature selection, 20 experiments were conducted each with feature selection using evaluator and search method as presented in Table 5.

Based on Table 5, CFS and CSE produce less than 10 attributes (out of 624) when using BF, GSW and LF search methods. It can be seen that LF search method gives the minimum number of features consistently in the experiments, where FSE+SMOTE output the minimum features (3 features). Further tests were carried out to compare the performance of selected features using two single classifiers as shown in Table 6 and Table 7.

Table 5. Number Of Feature Selected Using The Evaluator And Search Method.

Feature selection evaluator	Search Method			
	BF	*GS*	*GSW*	*LF*
CFS	9	229	9	6
CSE	8	266	6	5
FSE + Resample	20	176	20	8
FSE + SMOTE	14	261	14	3
Wrapper + NB	8	291	3	11

Table 6. Performance Of The Feature Selection Methods Using NB.

Methods	Classifier: NB & Search				
	BF	*GS*	*GSW*	*LF*	*Avg.*
CFS	50	40	50	55	48.75
CSE	45	65	45	20	43.75
FSE+Resample	**70**	40	**70**	55	**58.75**
FSE+SMOTE	50	50	50	40	47.5
Wrapper+NB	45	50	35	45	32.5

Table 7. Performance Of The Feature Selection Methods Using Random Forest.

Methods	Classifier: Random Forest & Search				
	BF	*GS*	*GSW*	*LF*	Avg.
CFS	65	60	65	40	57.5
CSE	50	65	35	45	48.75
FSE+Resample	50	65	50	70	**58.75**
FSE+SMOTE	50	50	50	60	52.5
Wrapper+NB	55	70	30	**75**	57.5

Taking the methods with highest classification rate (using FSE+Resample) from Tables 6-7, the detailed performance (F-measure) on the class labels for each method is shown in Table 8. The F-measure values indicate that although the accuracies of some methods are similar, however the effects of feature selection of the imbalance data are varied. It is proven that when the dataset has imbalance problem, high accuracy is actually poor choice for model evaluation as it just relies on majority class. The 1st row in the table illustrates this problem, where the majority class gets high F-measure while the minority class (Lakom) gives low F-measure value.

Two of the methods show that the F-measure values are almost balanced. In this case, the FSE+Resample with BF and GSW have better performance distribution except for class Kapal Terbang, where this class is supposed to be the majority class. FSE+Resample with LF search method produce a higher F-measure on minority class, but in turn, gets lower value in majority class, thus, the weighted average accuracy based on F-measure is lower (0.69) than the percentage accuracy (70%). Unfortunately, wrapper-based feature selection namely Wrapper+NB with GS has the similar F-measure values when all features are used. An interesting observation on Wrapper+NB with LF that this method provides the best accuracy, however the F-measure values are not seen promising compared to FSE+Resample with BF or GSW, where the majority and minority class gives low F-measure values. Increased performance in class 'Kemumur Itik' is shown by the higher average F-measure for wrapper+NB.

Table 8. Performance (F-Measure) Of The Best Feature Selection Methods On Each Class (*Best Value) Using Single Classifier (RF and NB).

	C1	C2	C3	C4	C5	Av.
All Features + Classifier: RF	0.33	0.80	0.73*	0.40	1.00	0.65
FSE+Resample +Search: BF +Classifier: NB	0.89*	0.44	0.50	0.67	1.00	0.70
FSE+Resample +Search: GSW +Classifier: NB	0.89*	0.44	0.50	0.67	1.00	0.70
FSE+Resample +Search: LF +Classifier: RF	0.67	0.33	0.55	0.89*	1.00	0.69
Wrapper+NB +Search: GS +Classifier: NB	0.33	0.89*	0.67	0.40	1.00	0.66
Average	0.89*	0.40	0.73*	0.57	1.00	0.72

C1 = Cemumar, C2 = Kapal Terbang, C3 = Kemumur Itik, C4 = Lakom,C5 = Mengkudu

Performance using all features indicate that it is better than most of feature selection methods when using single classifier. According to the results, filter-based feature selection methods almost performed similar given by the three classifiers. It can be seen that FSE+Resample performs better in every tested classifiers, where GS is the best search method when J48 and Random Forest is used as the classifiers. However, the Naive Bayes made the BF to perform better in average. Interestingly, despite that FSE+Resample perform better in average, only NB evaluates the feature selection methods with 70% accuracy using BF or GSW as search technique.

In relation to small feature subset selection, LF is surprisingly coming in second place for the average performance. In fact, although not the most minimum number of features, LF combined with Wrapper+NB provides the highest classification accuracy of 75%. This shows that the wrapper-based feature selection method has successfully selected the best feature subset (11 features) and evaluated by Random Forest (using default settings).

Comparing the classification performance between all features and selected features (which is very small), the feature selection effect is notable where it can represent the dataset significantly using small number of features. The performance is similar or even better as shown by the FSE+Resample and Wrapper +NB. Thus, it is proven that feature selection can improve the classification and in the also reduce the running time than using all features.

C. 5-Cross Validation Result

In another perspective of the previous results using training and testing data, Table 9 shows the performance results starting from original data and different Weka feature selection algorithms based on 5-cv.

Table 9. Ensemble Classifier With FSE+Resample And Bestfirst Search Method Using 5-CV.

Ensemble Classifier	Original data	Feature Selection
Random Forest	73.85	73.85
AdaboostM1+RF	70.77	70.77
Bagging+RF	73.85	73.85
Decorate+RF	75.38	75.38
END+RF	75.38	73.85
MultiBoostAB+RF	70.77	70.77
RotationForest+RF	75.38	**78.46**
Stacking+RF	64.62	73.85
Average	72.50	73.85

Further investigation is done to evaluate Weka feature selection combined with ensemble classifier. In the experiments, the combination of all data (training and testing) and 5-cv was applied to the data using ensemble classifier with FSE+Resample (BF). This technique is selected due to the best results compared to other feature selection methods in the experiments. According to the result, FSE+Resample can provide the support to ensemble classifier with similar or slightly improved classification accuracy using 20 features (compared to 624 in original data). Highest classification accuracy is achieved by RotationForest with RF as a base classifier (78.83%).

D. Benchmark data

In this section, the experiments are conducted to compare the predictive accuracy of the selected methods on the benchmark datasets. The selected large datasets comprise of imbalance multiclass (high imbalance ratio) as depicted in Table 3. First, the performances of ensemble classifiers were investigated and recorded on the original datasets. Table 10 presents the results of the experiments.

Table 10. Classification Performance Of Ensemble Classifier On Original Benchmark Dataset.

Ensemble Classifier	Landsat	Shuttle	PageBlocks
Random Forest	95.31 Time: 2.01	99.99 Time: 22.16	97.35 Time: 1.81
AdaboostM1+RF	**95.64** Time: 2.21	99.99 Time: 22.55	97.11 Time: 8.62
Bagging+RF	95.36 Time: 18.55	99.99 Time: 174.70	97.11 Time: 13.15
Decorate+RF	95.20 Time: 259.07	-	97.53 Time: 148.60
MultiBoostAB+RF	**95.64** Time: 2.25	99.99 Time: 21.78	96.97 Time: 12.69
RotationForest+RF	95.62 Time: 34.18	99.98 Time: 311.04	97.28 Time: 24.22

According to the results, almost all of the ensemble classifiers performs similar with improved accuracy compared to the previous results in the respective research. AdaBoostM1+RF and MultiBoost+RF produce the high accuracy and fairly fast in Landsat and Shuttle data but drop for PageBlock dataset. Decorate combined with RF is a good ensemble design, however, its processing time is too high which is not good for large dataset problem. Decorate combined with RF is a good ensemble design, but the processing time is too high which is not good for large dataset problem. Specifically, in Shuttle dataset where the classifier could not finish the experiment due to a very long processing time and high memory load error, thus marked as '-' as noted in Table 10, implemented on Intel Core-i5 (2.50GHz, 8GB RAM, 64 bit) computer.

Table 11. Classification Performance Of Ensemble Classifier With Feature Selection On Benchmark Dataset.

Ensemble Classifier	Landsat	Shuttle	PageBlocks
Random Forest	95.79 Time:1.51	95.45 Time:6.74	97.00 Time:0.77
AdaboostM1+RF	**95.81** Time:1.73	95.45 Time:46.81	96.78 Time:6.51
Bagging+RF	95.46 Time:14.79	95.44 Time:66.36	97.08 Time:6.93
Decorate+RF	**95.81** 232.65	-	**97.61** Time:201.72
MultiBoostAB+RF	**95.81** Time:1.81	95.43 Time:57.37	96.91 Time:8.51
RotationForest+RF	95.55 Time:28.19	95.43 Time:114.97	96.27 Time:12.15
Features	22/36	2/9	3/10

Feature selection and ensemble classifier combination is further experimented. The results are presented in Table 11. The results show that the trend follows the results in Table 10, where all ensemble classifiers perform almost similar on the benchmark data. AdaboostM1+RF, MultiBoostAB+RF and Decorate+RF improved classification rate for Landsat using less number of attributes (22 of 36). However, the performance of all of the ensemble classifiers is seen to drop in Shuttle dataset. This is due to FSE+RESAMPLE method that is able to capture the right attributes. In PageBlocks dataset, the performance of the ensemble classifiers except Decorate+RF is lower compared to the previous results. Interestingly, the model build time on all classifiers were reduced when an attribute selection is performed.

Finally, in most cases from the experiments, RF is performing better as classifier or as a base classifier for another ensemble method. This has been discussed in (Breiman, 2001) where RF is actually consists of several decision trees which constructed using a random process. RF is also known as an ensemble methods of random decision trees (Vens, 2013), combining the predictions of the individual trees. In previous research such as (Kamanksha & Sanjay, 2018) proved that RF is the better classifier in their analysis. Thus, the results in this study shows that RF is a good classifier or combined as a base classifier for other ensemble methods.

V CONCLUSION

Data from medicinal leaf images (shape data) and large multiclass imbalance benchmark data were trained and tested using a combination of feature selection methods coupled with the ensemble classifiers. While the combinations tested in this study were performed almost similarly, there is no single combination that best fit to the problem. However, if there is no modification to the data, ensemble classifier with Random Forest can perform better than any single classifier. Furthermore, when Random Forest is combined as a meta classifier such as AdaBoostM1 or MultiBoostAB, and feature selection method using FSE+RESAMPLE, the performance can be further improved.

REFERENCES

Barati, M., Abdullah, A., Mahmod, R., Mustapha, N., & Udzir, N. I. (2013). Features Selection for IDS in Encrypted Traffic using Genetic Algorithm. In *Proceedings of the 4th International Conference on Computing and Informatics* (pp. 279–285).

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Cohen, I., Cozman, F. G., Sebe, N., Cirelo, M. C., & Huang, T. S. (2004). Semisupervised Learning of Classifiers: Theory, Algorithms, and Their Application to Human-Computer Interaction. *IEEE Trans. Pattern Anal. Mach. Intell.*, *26*, 1553–1567.

Cuaya, G., Muñoz-Meléndez, A., Morales, E. F., Cuaya, G., Muñoz-Meléndez, A., & Morales, E. F. (2011). A Minority Class Feature Selection Method. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, *7042*, 417–424.

Ding, Z. (2011). *Diversified Ensemble Classifiers for Highly Imbalanced Data Learning and Their Application in Bioinformatics*. *Computer Science Department, Georgia State University*. Georgia State University. Retrieved from http://digitalarchive.gsu.edu/cs_diss/60

Eschrich, S., Chawla, N. V., & Hall, L. O. (2002). Generalization methods in bioinformatics. In *2nd International Conference on Data Mining in Bioinformatics (BIOKDD'02)* (pp. 25–32).

Ghosh, S., Biswas, S., Sarkar, D., & Sarkar, P. (2014). A Tutorial on Different Classification Tech-niques for Remotely Sensed Imagery Datasets. *Smart Computing Review*, *4*(1), 34–43.

Hall, M. A. (2000). Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 359–366). Morgan Kaufmann Publishers Inc.

Kamanksha, D. P., & Sanjay, A. (2018). A Critical Analysis of Twitter Data for Movie Reviews Through `Random Forest' Approach. In S. C. Satapathy & A. Joshi (Eds.), *Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 2* (pp. 454–460). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-63645-0_52

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *In Proc. of the 14th Joint International Conference Artificial Intelligence (IJCAI'95)*, *2*, 1137–1143.

Liu, H., & Setiono, R. (1996). A probabilistic approach to feature selection - A filter solution. In *13th International Conference on Machine Learning* (pp. 319–327).

López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, *250*, 113–141. https://doi.org/10.1016/j.ins.2013.07.007

Mohsin, M. F. M., Hamdan, A. R., & Bakar, A. A. (2014). An Evaluation of Feature Selection Technique for Dendrite Cell Algorithm. In *Proceeding of International Conference on IT Convergence and Security (ICITCS)* (pp. 1–5). IEEE.

Onik, A. R., Haq, N. F., Alam, L., & Mamun, T. I. (2015). An Analytical Comparison on Filter Feature Extraction Method in Data Mining using J48 Classifier. *International Journal of Computer Applications*, *124*(13), 1–8.

Sainin, M. S., Ahmad, F., & Alfred, R. (2016). Improving the Identification and Classification of Malaysian Medicinal Leaf Images Using Ensemble Method. In *Proceeding of the First International Conference on ICT for Transformation 2016* (pp. 1–6).

Vens, C. (2013). Random Forest. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 1812–1813). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_612