Vol.1 pp. 12-18 Journal of Operational Management & Marketing Strategies (OMMS)

Feature Selection Aggregation Versus Classifiers Aggregation for Several Data Dimensionalities

Afef Ben Brahim
LARODEC, ISGT
University of Tunis, Tunisia
Email: afef.benbrahim@yahoo.fr

Waad Bouaguel
LARODEC, ISGT
University of Tunis, Tunisia
Email: bouaguelwaad@mailpost.tn

Mohamed Limam
LARODEC, ISGT
University of Tunis, Tunisia
Dhofar University, Oman
Email: mohamed.limam@isg.rnu.tn

***Abstract*—The principal goal of machine learning in classification problems is to achieve the best classification performance for a given application. This purpose is typically accomplished by using a significant set of features that conserves the significance and improves the model generalization. Many feature selection methods are available and this results on a trouble of selecting the appropriate feature selection method for a given classification problem. Since finding the best method is usually difficult in real application, we propose a feature selection framework that fuses the results obtained by different selection methods. We investigate the effect of ensemble feature selection on classification accuracy and we compare the ensemble aggregation in two different levels, classifier based aggregation and selector based aggregation. Experiments on four data sets of different dimensionalities show that ensemble feature selection outperforms individual methods in terms of classification accuracy and that data set dimensionality can guide the choice of the aggregation level of the ensemble feature selection.**

I. Introduction

Choosing the set of features to retain remains the most important factor for any effective classification. Typically, features which are noisy, redundant or irrelevant to the classification task produce inhibited performance for any classifier. Therefore selecting a proper set of features is critical for a successful classification. Thus, it is necessary to identify the valuable set of features and eliminate the undesired ones which may produce worse performance.

The curse of dimensionality is also a big motivation to look for a reduced set of features. In fact with a high number of features the computational time of classification algorithm increase significantly, without any significant change in the performance. With the increase in noise and dimensionality, feature selection becomes an essential step.

Feature selection algorithms are designed to improve the classification performance of a single or a multiple classifiers system, by removing redundant or noisy features from the data. Typically, a feature selection technique looks for a suitable subset of features from the original features set, in order to improve the accuracy of a particular application. Feature selection methods can be divided into three category: Filter , Wrapper, and hybrid methods [1], [2]. Filter methods evaluate features individually and eliminate irrelevant ones before a classification algorithm is trained. Wrapper methods form a second group of feature selection methods, in which the prediction accuracy of a classifier directly measures the value of a feature set. While the filter method is unbiased and fast, the wrapper method gives better results for a particular classifier. Although effective, the exponential number of possible subsets places computational limits. Hybrid method is a fusion of both filter and wrapper methods [3].

Using feature selection algorithms individually may not automatically lead to better performance, because a single feature selection algorithm focuses on one particular region of the feature space. However, different feature selection algorithms will choose different feature subsets, resulting in a classifier that will be trained on a subset that represents the whole set.

The fusion of different features selectors is a step to generate a new feature set from the individual selected set of features. There are two possible levels of aggregation to obtain an ensemble of feature selection methods. The first one is the feature classifiers aggregation level, the second is the selectors aggregation level. Feature ensemble based classifiers combination consists on the parallel combination of decisions from multiple classifiers. Each classifier is trained using variations of the feature representation space, obtained by means of feature selection. The final classification output is obtained by the aggregation of the results of all classifiers in the ensemble.

The second level of aggregation is the selector aggregation level which is based on the combination of feature sets obtained by the application of different selectors. In a first step, a number of different feature selectors are used, and in a final phase the output of these separate selectors is aggregated and returned as the final ensemble result. A single classifier could then be applied on the resulting feature set.

The reminder of the paper is organized as follows. In Section II, we discuss ensemble feature selection and we summarize available techniques based on both classifiers and feature selectors aggregation. Section III presents experimental results on four data sets. We give a discussion of our study in Section IV and we finally conclude this paper in Section V.

II. Ensemble Feature Selection

A. Classifier Based Aggregation

Feature ensemble method based classifiers combination consists on the parallel combination of decisions from multiple classifiers. Each classifier is trained using variations of the

ISSN: 2961-662X

feature representation space, obtained by means of feature selection. With this approach, relevant discriminative information contained in features neglected in a single run of a feature selection method, may be recovered by the application of multiple feature set runs or algorithms, and contribute to the decision through the classifier combination process.

While traditional feature selection algorithms have the goal of finding the best feature subset that is relevant to both the learning task and the selected inductive learning algorithm, the task of ensemble feature selection by classifiers combination has the additional goal of finding a set of feature subsets that will promote disagreement among the base classifiers. To have this disagreement, we used different feature selection algorithms to generate the feature subsets. Opitz proposed an ensemble feature selection approach that is based on genetic algorithms in order to generate a set of classifiers that are accurate and diverse in their predictions. Their proposed approach works by finding a set of feature subsets that will promote disagreement among the ensemble classifiers [4].

Tsymbal et al. introduced also a genetic algorithm-based sequential search for ensemble feature selection (GAS-SEFS). Instead of one genetic process, it employs a series of processes, the goal of each of which is to build one base classifier [5].

Figure 1 illustrates the classifiers based aggregation process.

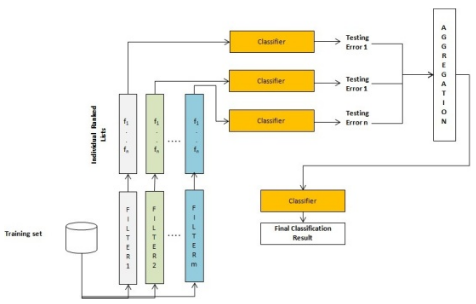

Fig. 1. Classifier Based Aggregation

B. Feature Selector Based Aggregation

The concept of ensemble feature selection based feature selectors aggregation was recently introduced [6]. Ensemble feature selection techniques use an idea similar to ensemble learning for classification [7]. In a first step, a number of different feature selectors are used, and in a final phase the output of these separate selectors is aggregated and returned as the final ensemble result. Similar to the case of supervised learning, ensemble techniques might be used to improve the robustness of feature selection techniques. Different feature selection algorithms may yield feature subsets that can be considered local optima in the space of feature subsets, and ensemble feature selection might give a better approximation to the optimal subset or ranking of features. Also, the representational power of a particular feature selector might constrain its search space such that optimal subsets cannot be reached. Ensemble feature selection could help in alleviating this problem by aggregating the outputs of several feature selectors [6]. This concept was especially applied with high dimensional data with few samples [6], [8], but it can be applied to any data dimensionality as it will be seen in our experiments.

Figure 2 illustrates the feature selectors based aggregation process.

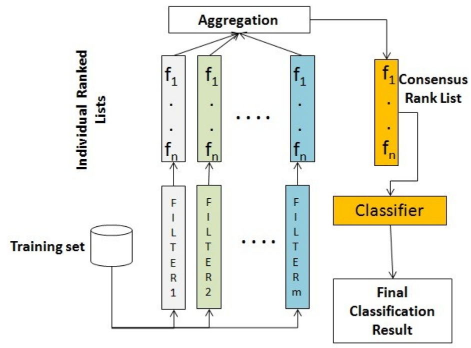

Fig. 2. Feature Selector Based Aggregation

III. Comparative Study

In this section, we compare the use of ensemble methods in two different levels. The first one is the classification level and the second is the feature selection level. Our objective is to study the characteristics and to compare the performance of each setting but especially to search for the level in which the feature selection process is the most effective. We start by applying three different feature selection methods on four data sets resulting on three selected feature subsets for each data set. Then in a first setting, we apply a classification algorithm on the projection of each feature subset on the training data. We then aggregate the classification results of the ensemble. In a second setting, the three selected feature subsets obtained initially are combined in order to obtain a final individual feature subset before proceeding to the classification step. We compare the classification results of the two settings.

A. Datasets

As discussed before, using a large number of features is not always effective. This is especially true when the problem involves unsupervised learning or supervised learning with unbalanced data (many negative observations but minimal positive observations). This paper addresses two issues involving different data dimensionalities. The first issue explores the behavior of ensemble feature selection with its two aggregation levels with data having thousands of dimensions and small sample size. The second issue deals with huge data sets with a massive number of instances and where feature selection is used to extract meaningful rules from the available data.

The experiments for the first case are conducted on Central Nervous System data set (CNS), which is concerned with the prediction of central nervous system embryonal tumor outcome based on gene expression. This data set includes 60 samples containing 39 medulloblastoma survivors and

ISSN: 2961-662X

21 treatment failures. These samples are described by 7129 genes [9]. We consider also the Leukemia microarry gene expression dataset that consists of 72 samples which are all acute leukemia patients, either acute lymphoblastic leukemia (47 ALL) or acute myelogenous leukemia (25 AML). The total number of genes to be tested is 7129. [10]

For the second case two credit datasets are used, the Australian and the Tunisian credit dataset. The first present an interesting mixture of attributes: 6 continuous, 8 nominal and a target attribute with few missing values. This dataset is composed of 690 instances where 306 are creditworthy and 383 are not. All attribute names and values have been changed to meaningless symbols for confidentiality. The Tunisian dataset covers a sample of 2970 instances of credit consumers where 2523 instances are creditworthy and 446 are not. Each credit applicant is described by a binary target variable and a set of 22 input variables were 11 features are numerical and 11 are categorical. Table I displays the characteristics of the data sets that have been used for evaluation.

TABLE I
DATASETS SUMMARY

Names	Australian	Tunisian	CNS	Leukemia
Total instances	690	2970	60	72
Total features	14	22	7129	7129
Number of classes	2	2	2	2
Missing Values	Yes	Yes	No	No

B. Feature Selection Algorithms

Our feature selection ensemble is composed by three different filter selection algorithms, Relief algorithm [11], Correlation-based feature selection (CFS) [12] and Information gain (IG) [13]. These algorithms are available in Weka 3.7.0 machine learning package [14].

Relief algorithm evaluates each feature by its ability to distinguish the neighboring instances. It randomly samples the instances and checks the instances of the same and different classes that are near to each other.

Correlation-based Feature Selection (CFS) looks for feature subsets based on the degree of redundancy among the features. The objective is to find the feature subsets that are individually highly correlated with the class but have low inter-correlation. The subset evaluators use a numeric measure, such as conditional entropy, to guide the search iteratively and add features that have the highest correlation with the class.

Information gain (IG) measures the number of bits of information obtained for class prediction by knowing the presence or absence of a feature.

The aggregation of these filters in the feature selection level is performed by choosing the selected features shared by the three methods. We refer to the ensemble based feature selection aggregation by ensemble feature aggregation (EFA).

C. Classifiers

We trained our approach using three well-known data mining algorithms, namely Decision trees, Support vector machines and The K-nearest-neighbor. These algorithms are available in Weka 3.7.0 machine learning package [14].

Decision Trees(DT) is a very simple method and can be described as a set of nodes and edges, the root node define the first split of the credit-applicants sample. Each internal node split the set of instances into two subsets. Each node contains individuals of a single class; the operation is repeated until the separation in sub-populations is no more possible.

Support Vector Machines (SVM) is one of the most outstanding machine learning techniques. There many raisons for choosing SVM [15], it requires less prior assumptions about the input data and can perform on small or huge data set by doing a nonlinear mapping from an original input space into a high dimensional feature space.

The K-Nearest-Neighbor (KNN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.

The aggregation in the classifiers level is performed by using five well known combination rules namely, the majority vote, the average probability, the product probability, the minimum probability and the maximum probability combination rule.

D. Performance Metrics

To evaluate the classification performance of each setting and perform comparisons, we used several characteristics of classification performance all derived from the confusion matrix [16]. We define briefly these evaluation metrics.

The precision is the percentage of positive predictions that are correct. The Recall (or sensitivity) is the percentage of positive labeled instances that were predicted as positive. The F-measure can be interpreted as a weighted average of the precision and recall. It reaches its best value at 1 and worst score at 0.

Another the characteristic of a classifier, frequently used is a Receiver Operating Characteristic (ROC) curve. It is a visual characteristic allowing for visualizing classification performance of one or several algorithms. A Receiver Operating Characteristic curve is a plot of the sensitivity (or the TP rate) against one minus its specificity (or the FP rate), as the cut-off criterion for indicating a positive test is varied. This plot depicts relative trade-offs between true positives and false positives. We use the area under the curve (ROC Area) as another performance metric.

E. Results Analysis

We consider information retrieval measures of data sets when individual filter methods are applied, using the learning algorithms by 10-fold cross validation. Then we apply the ensemble feature selection, first based on the classifier aggregation then on the selectors aggregation. We measure the

ISSN: 2961-662X

performance of those methods using the performance metrics described precedently. Tables II - V show the obtained results.

In most cases, the ensemble methods produced better performance than individual filter methods. In particular, the average of probability aggregation method performed even better than other methods especially for the Australian and the Tunisian datasets with DT and KNN learning algorithms. Also, the product of probability aggregation worked as good as the average of probability with DT and KNN learning algorithm once applied to the Central Nervous System and Leukemia datasets.

We conducted the same experiments with those datasets and we applied ensemble feature selection based on the feature set aggregation and evaluated each method by the same process. As expected, ensemble methods again give great results. In fact, for very small dataset sizes as the Central Nervous System and Leukemia datasets, ensemble methods outperform individual feature selection methods. However, the learning performance with the Australian and the Tunisian datasets did not improve so much. This might be due to the massive number of instances bias.

TABLE II
RESULTS SUMMARY FOR THE AUSTRALIAN DATASET

	Precision	Recall	F-Measure	ROC Area
Decision Tree				
Cfs	0.906	0.833	0.868	0.882
Relief	0.878	0.849	0.863	0.889
InfoGain	0.93	0.799	0.86	0.832
Majority V	0.883	0.843	0.862	0.852
Average P	0.887	0.841	0.863	0.901
Product P	0.887	0.843	0.865	0.897
Max P	0.901	0.83	0.864	0.897
Min P	0.901	0.83	0.864	0.893
EFA	0.899	0.833	0.864	0.881
SVM				
Cfs	0.936	0.799	0.862	0.865
Relief	0.919	0.802	0.856	0.857
InfoGain	0.927	0.799	0.858	0.86
Majority V	0.921	0.794	0.853	0.855
Average P	0.921	0.794	0.853	0.863
Product P	0.935	0.799	0.862	0.861
Max P	0.914	0.802	0.854	0.864
Min P	0.935	0.799	0.862	0.861
EFA	0.925	0.802	0.859	0.86
KNN				
Cfs	0.89	0.843	0.866	0.897
Relief	0.866	0.864	0.865	0.886
InfoGain	0.853	0.896	0.874	0.892
Majority V	0.857	0.875	0.866	0.846
Average P	0.866	0.875	0.87	0.903
Product P	0.864	0.877	0.87	0.897
Max P	0.865	0.869	0.867	0.91
Min P	0.865	0.869	0.867	0.901
EFA	0.868	0.856	0.862	0.884

TABLE III
RESULTS SUMMARY FOR THE TUNISIAN DATASET

	Precision	Recall	F-Measure	ROC Area
Decision Tree				
Cfs	0.865	0.981	0.92	0.553
Relief	0.85	1	0.919	0.497
InfoGain	0.862	0.983	0.919	0.547
Majority V	0.865	0.985	0.921	0.557
Average P	0.85	1	0.919	0.653
Product P	0.85	0.999	0.918	0.647
Max P	0.85	0.999	0.918	0.653
Min P	0.85	0.999	0.918	0.646
EFA	0.85	1	0.919	0.497
SVM				
Cfs	0.851	0.994	0.917	0.505
Relief	0.85	1	0.919	0.5
InfoGain	0.868	0.907	0.887	0.563
Majority V	0.851	0.994	0.917	0.505
Average P	0.851	0.994	0.917	0.505
Product P	0.851	1	0.919	0.505
Max P	0.85	1	0.919	0.505
Min P	0.851	1	0.919	0.505
EFA	0.85	1	0.919	0.5
KNN				
Cfs	0.864	0.959	0.909	0.675
Relief	0.862	0.932	0.895	0.602
InfoGain	0.86	0.94	0.898	0.607
Majority V	0.86	0.967	0.91	0.539
Average P	0.86	0.957	0.906	0.67
Product P	0.86	0.937	0.897	0.658
Max P	0.86	0.937	0.897	0.67
Min P	0.86	0.937	0.897	0.643
EFA	0.861	0.931	0.895	0.596

Based on the discussions above, we conclude that if the dataset size is very small and the number of features exceed the number of instances the best way to introduce aggregation is in the pre-processing step before the learning process. In case the dataset is big and the number of feature does not exceed the number of instances, aggregation is more benefic once used on the learning algorithms trained over the reduced data by the individual filters.

IV. DISCUSSION

In this paper, we investigate the use of ensemble methods for different sample size data classification. The use of ensemble methods is studied in two different levels. The first is the classification level and the second is the feature selection level. Our objective is to study the characteristics and to compare the performance of each setting but especially to search for the level in which the feature selection process is the most effective. First of all, we apply three different feature selection methods on four datasets resulting on three selected feature subsets for each dataset. Then in a first setting, we apply a classification algorithm on the projection of each feature

ISSN: 2961-662X

TABLE IV
RESULTS SUMMARY FOR THE CENTRAL NERVOUS SYSTEM DATASET

Decision Tree				
	Precision	Recall	F-Measure	ROC Area
Cfs	0.676	0.641	0.658	0.512
Relief	0.600	0.538	0.568	0.399
InfoGain	0.674	0.744	0.707	0.535
Majority V	0.69	0.744	0.716	0.562
Average P	0.69	0.744	0.716	0.426
Product P	0.48	0.75	0.585	0.595
Max P	0.5	0.333	0.400	0.411
Min P	0.5	0.813	0.619	0.595
EFA	0.775	0.795	0.785	0.690
SVM				
Cfs	0.700	0.718	0.709	0.573
Relief	0.632	0.615	0.623	0.474
InfoGain	0.737	0.718	0.727	0.621
Majority V	0.737	0.718	0.727	0.621
Average P	0.737	0.718	0.727	0.58
Product P	0.737	0.718	0.727	0.58
Max P	0.704	0.487	0.576	0.542
Min P	0.704	0.760	0.731	0.553
EFA	0.9	0.923	0.911	0.866
KNN				
Cfs	0.677	0.538	0.600	0.531
Relief	0.659	0.692	0.675	0.513
InfoGain	0.727	0.615	0.667	0.593
Majority V	0.688	0.564	0.62	0.544
Average P	0.688	0.564	0.62	0.563
Product P	0.688	0.564	0.62	0.571
Max P	0.739	0.436	0.548	0.574
Min P	0.739	0.436	0.548	0.574
EFA	0.878	0.923	0.9	0.842

TABLE V
RESULTS SUMMARY FOR THE LEUKEMIA DATASET

Decision Tree				
	Precision	Recall	F-Measure	ROC Area
Cfs	0.933	0.894	0.913	0.865
Relief	0.933	0.894	0.913	0.865
InfoGain	0.913	0.894	0.903	0.871
Majority V	0.933	0.894	0.913	0.873
Average P	0.933	0.894	0.913	0.873
Product P	0.933	0.913	0.923	0.883
Max P	0.915	0.915	0.915	0.873
Min P	0.933	0.913	0.923	0.883
EFA	0.911	0.872	0.891	0.843
SVM				
Cfs	0.958	0.979	0.968	0.949
Relief	0.979	0.979	0.979	0.969
InfoGain	0.938	0.957	0.947	0.919
Majority V	0.979	0.979	0.979	0.969
Average P	0.979	0.979	0.979	0.968
Product P	0.978	0.978	0.978	0.959
Max P	0.92	0.979	0.948	0.966
Min P	0.978	0.978	0.978	0.959
EFA	0.958	0.979	0.968	0.949
KNN				
Cfs	0.938	0.957	0.947	0.911
Relief	0.957	0.957	0.957	0.936
InfoGain	0.956	0.915	0.935	0.92
Majority V	0.957	0.957	0.957	0.939
Average P	0.957	0.957	0.957	0.958
Product P	0.957	0.957	0.957	0.958
Max P	0.92	0.979	0.948	0.956
Min P	0.92	0.979	0.948	0.956
EFA	0.978	0.936	0.957	0.947

subset on the training data. We then aggregate the classification results of the ensemble. In a second setting, the three selected feature subsets obtained initially are combined in order to obtain a final individual feature subset before proceeding to the classification step. The comparison of the two settings performances conducts to the following conclusions.

On most cases, the ensemble results, even obtained by one ensemble setting or the other outperform those obtained by the application of a single feature selection algorithm followed by a single classifier.

For datasets with small dimensionality, the best performance results are obtained by classifiers aggregation and never with feature selectors aggregation.

For high dimensional datasets, the best performance results are achieved even by classifiers or selectors aggregation, with special high values when feature selectors aggregation is outperforming.

A possible explanation of the performance of feature selection aggregation on high dimensional data sets and not on small size datasets is that on the latter individual feature subsets obtained by different feature selection methods may be very similar as the initial number of features is small. However, in the case of high dimensional datasets, obtained feature subsets from the ensemble feature selection process may be very different as the feature space is very large. Thus the features combination effect on classification performance will be much more apparent in case of high dimensional datasets.

Sample size is therefore determinant of the choice of one or the other setting when classification accuracy is the performance criterion taken into account. Stability is another important criterion for evaluating feature selection results and in terms of this performance metric we expect that feature selection algorithm ensembles will be preferred as they focus on improving classification results by strengthening feature selection results. It is not the case for classifier ensembles which focus on strengthening classification results without a special care to feature selection phase.

V. CONCLUSION AND FUTURE DIRECTIONS

We investigated the use of ensemble methods in classification level and in feature selection level. Our experiments on four datasets of different dimensionality showed that sample

ISSN: 2961-662X

size influence classification results obtained by one or the other setting and thus determine the level on which aggregation must be performed, classification or feature selection level. We plan to do the proposed research study on a higher number of datasets from both small and high dimensionality in order to validate our conclusions about sample size influence on the choice of the aggregation level. We expect also to evaluate stability as another performance criterion as it becomes important to have stable feature selection results with the increasing data dimensionality due to high technologies.

REFERENCES

[1] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, Apr. 2005.

[2] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artif. Intell.*, vol. 151, no. 1-2, pp. 155–176, 2003.

[3] E. Tuv, A. Borisov, G. Runger, K. Torkkola, I. Guyon, and A. R. Saffari, "Feature selection with ensembles, artificial variables, and redundancy elimination," *JMLR*, 2009.

[4] D. W. Opitz, "Feature selection for ensembles," in *In Proceedings of 16th National Conference on Artificial Intelligence (AAAI.* Press, 1999, pp. 379–384.

[5] A. Tsymbal, M. Pechenizkiy, and P. Cunningham, "Sequential genetic search for ensemble feature selection," in *Proceedings of the 19th international joint conference on Artificial intelligence*, ser. IJCAI'05. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005, pp. 877–882.

[6] Y. Saeys, T. Abeel, and Y. Peer, "Robust feature selection using ensemble feature selection techniques," in *Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases - Part II*, ser. ECML PKDD '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 313–325.

[7] T. G. Dietterich, "Ensemble methods in machine learning," in *Proceedings of the First International Workshop on Multiple Classifier Systems*. London, UK, UK: Springer-Verlag, 2000, pp. 1–15.

[8] B. Schowe and K. Morik, "Fast-ensembles of minimum redundancy feature selection," in *Ensembles in Machine Learning Applications : Studies in Computational Intelligence*, vol. 373.

[9] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. H. Kim, L. C. Goumnerova, P. M. Black, C. Lau, J. C. Allen, D. Zagzag, J. M. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S. Lander, and T. R. Golub, "Prediction of central nervous system embryonal tumour outcome based on gene expression," *Nature*, vol. 415, no. 6870, pp. 436–442, 2002.

[10] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, and C. D. Bloomfield, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, pp. 531–537, 1999.

[11] K. Kira and L. Rendell, "A practical approach to feature selection," in *International Conference on Machine Learning*, D. Sleeman and P. Edwards, Eds., 1992, pp. 368–377.

[12] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proceedings of the Seventeenth International Conference on Machine Learning*. Morgan Kaufmann, 2000, pp. 359–366.

[13] J. R. Quinlan, *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., 1993.

[14] R. R. Bouckaert, E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, *Weka manual (3.7.1)*, Jun. 2009.

[15] J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.

[16] O. Okun, *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations*, 2011.

ISSN: 2961-662X