Comparing Informative Sample Selection Strategies in Classification Ensembles

Hamza Osman İlhan and Mehmet Fatih Amasyal

***Abstract*—Usage of more training data with label information gives more success for classification of datasets in machine learning. But in real life, obtaining data with label information is a cost-effective and long-lasting process. Herein, active learning algorithms are emerged. Active learning algorithms aim to maintain current success rate with fewer samples in train set or increase total success of model in training process. Active learning is not only functional for regular learning methods but also can be used in ensemble learning algorithms with specified techniques. In this study, two different active learning algorithms based on class probabilities of the samples are tested on five datasets classification. Ensemble learning methods are used as classification model. Comparative results presented as graphically and numerically.**

***Index Terms*—Active learning, adaboost, bagging, decision tree, ensemble learning, machine learning.**

I. Introduction

Classification process is a way to analyze datasets according to labels which represent the class information about related data in datasets. It is being used in many research fields such as economy, bioscience, forecasting etc. Not only computer science researchers but also economists, doctors are getting help to strengthen their decisions for later researches and summarize previous years' data to analyze better [1]-[3].

In the manner of doing machine learning, plenty of methods are brought out for classification. ANN (artificial neural network), decision trees, Bayes, naïve Bayes are some of them [3], [4]. Classification methods, typically, create models based on training process with label information which is already given in datasets. Models are performed to rest of data to classify. In the end, success of classification is measured and the success rate of models can be defined based on this rate. Success rate of those methods are generally high because the label information is given in train data. Methods working with labeled train set are named as supervised methods in literature [5]. But datasets with enough labels are not easy to find for studies.

Obtain the label information for data is difficult in real life. Furthermore, it is impossible to get all labels in correctly split up [6]. In that case some other machine learning algorithms called as unsupervised methods are improved [5]. *K*-Means, which based on distances of given clusters' centers, is an example of unsupervised methods. Segmentation of the data is made by the distances, instead of using labels. Procedure is referred as clustering in literature. But unsupervised clustering methods' success is not efficient level when compared to the methods which are using labels [7]. By the purpose of increasing success rate of regular machine learning classification methods, Ensemble learning and specified methods are constituted.

Ensemble learning concept is combination of several machine learning algorithms. This combination concludes the process with higher success because many learners make decision of the classification not only one as in regular machine learning. Regular machine learning algorithms such as Decision Tree or Bayes classifier is defined as base learners in ensemble learning. Broadly speaking, more base learners used in model gives more successful results in classification. Moreover, there are also methods which plays effective role on success such as Voting, Bagging, Boosting and some of cascading models. Methods are generally used for specified purposes. Some methods can give good result whereas others can be failed on the same data. In this sense, many sub techniques arise related with data. LPBoost, BrownBoost, AdaBoost, Logitboost are some of sub techniques for boosting method [8].

Even though there are many ensemble learning methods based on successful base learners, analyzing of datasets with unlabeled or few labeled data and achieving high success rates in classification are still important case for researchers. In this meaning, active learning methods are implemented on ensemble methods for rational selection of samples in literature [6].

In this study, active sample selection called as active learning algorithms on ensemble learning methods are presented. Two different active learning algorithms are tested on 5 datasets having different instances and classes at the stage of sample selection. Moreover, regular sample selection method based on randomly selection algorithm is also tested. Tests are performed on AdaBoost and Bagging methods of ensemble learning.

In the scope of this paper, active sample selection algorithms in ensemble learning methods are explained in Section II with related works information. Section III includes used datasets information and process steps of application. Figures and tables about the results are given in Section IV with the explanations. Conclusion and future works will be present in Section V.

II. Active Sample Selection in Ensemble Learning

Number of sample size used in machine learning algorithms is the main factor of success rate in results. Many studies about definition of the optimal sample size has made in literature [9], [10]. These studies show that the more

Manuscript received October 9, 2013; revised December 6, 2013. This work was supported in part by the Turkey, Yildiz Technical University.

The authors are with the Department of Computer Engineering, Yildiz Technical University, İstanbul, 34720 TR (e-mail: hoilhan@yildiz.edu.tr, mfatih@ce.yildiz.edu.tr).

sample trained by algorithms gives more successful results on classification for rest of data. But as cited previous section of this paper, gathering data with label information is not easy in real life solutions. Many datasets don't have all data' label information. The main problem is that gathering label information takes long time and costly process [5]. On the other hand, unlabeled data is easy to obtain and also plenty in literature for researchers. At this stage, clustering methods are used for unlabeled data classification which is known as unsupervised method in literature but success of the separation the data is low than supervised methods' success [5]. Recently, many studies are made on forming new classification models which is based on supervised techniques but use less labeled data in training process because of difficulty in obtaining [10], [11].

Active learning is a method mainly used to maintain high success rates with less labeled data [9], [11]. Learning processes are initially started with few labeled data by machine learning methods. In the scope of this paper, Ensemble learning methods are used. Formed model based on learning process is performed over rest of samples which named as data pool in this study. Samples within the data pool have probabilities values of belonging to classes after testing iteration. Calculation process can vary up to selected active learning algorithm. Many algorithms are defined in literature such as query by committee [12], uncertainty sampling [13], margin sampling [14], entropy [14] etc. The most effective labeled data decided by using active learning algorithm is referred as informative data. New train set is formed with many informative data for next training process in the scope of active learning. This process lasts until the top limit defined by user for train data.

The simplest and most common algorithm is uncertainty sampling algorithm in literature [13]. Active learner selects new samples which don't belong to any class information from the data pool in uncertainty sampling algorithm. This algorithm is useful for datasets having two classes because most informative label information is used and other label information and their possibilities of belonging to classes are ignored. Margin sampling method is emerged to fix this problem [14]. Most informative two possibilities of data' label are used in margin sampling method. Method selects samples which have minimum margin differences as mostly informative samples. But margin sampling method also ignores the rest of the distribution of data' label information.

In this study, two different modified active learning algorithms are presented. Modified algorithms are composed from regular uncertainty sampling and margin sampling algorithms. Both algorithms start with 1% randomly selected data from all dataset as train data. Two different models which are based on two different ensemble learning methods with decision tree combination are trained by randomly selected initial train data. AdaBoost and Bagging methods are used to observe response of active sample selection algorithms on different ensemble models. Created classification models are performed on rest of data called as data pool. Classification success rate depends on informative sample selection and probabilities of each samples belonging to classes are obtained.

Active sample selection algorithms which are called as active learning start at second iteration. Calculated probabilities of samples in each iteration are used to make decision on creating new train set in both active sample selection algorithms. As it is known from general probability equation, total of the all probabilities' values for specified sample should be equal to 1 as in (1).

$$\sum_{r=1}^{N} P(x_k)^r = 1 \tag{1}$$

N is donated as total class number and $P(x_k)^r$ represents probabilities of k^{th} sample x for class r.

Higher value in probabilities of sample belonging to specified class means possession of sample mostly related this class. In both active learning algorithms, this information will be used for new sample selection step.

A. Active Learning Based on Strict Separation

New train set is created based on probabilities' values of samples which are calculated in previous iteration. Values of sample classification probabilities are checked for not equal to 1 in spite of uncertainty sampling [13]. In this study, this algorithm is named as strictly separation algorithm. Samples whose probability equal to 1 for specific class indicates that the sample is already classified in other words stable samples. Otherwise, samples having many different probabilities for classes called as unstable samples (2). Strictly separation algorithm works based on the unstable samples. Randomly selected 1% of all data based on being unstable structure is added to form new train set in each iteration.

$$\begin{matrix} r = 1,2 \ldots . N \\ k = 1,2 \ldots . T \end{matrix} \quad \begin{cases} P(x_k)^r = 1 & \textit{Stable Sample} \\ P(x_k)^r < 1 & \textit{Unstable Sample} \end{cases} \tag{2}$$

where r is class number and N is total number of classes. k represents the sample number and T is defined as total samples in data. $P(x_k)^r$ is the probability value of k^{th} sample x for class r.

B. Active Learning Based on Margin Distances

Starting process is the same as mentioned in strictly separation algorithm. Difference is at the processing step using the probability values. Margin distance algorithm selects new data based on distance measurements between probabilities belong to classes for the particular sample. Probabilities' values are calculated previous iteration similar to strictly separation algorithm. Distances between sample's probabilities according to the classes are used to define unstable or stable samples as in (3). The higher differences between assigned class probabilities make the sample more stable than others because of the general probability rule (1). Calculation for margin difference information is made between maximum and minimum values of sample probabilities. But in regular margin sampling method, this calculation is made on the first two values of probabilities [14]. As it's known from regular probabilistic rule, if the value is 1, it means sample is strictly assigned to one class which is sample is already stable. That information is used for previous algorithm. In margin distances algorithm, new train set is formed by samples whose probabilities' difference belonging to specific class are minimum. As it is another algorithm called as margin distances based active sample

selection is also tested in this study. 1% of all data selected by margin distance algorithm is added to train set with a descending sequence algorithm from max to min in each iteration by using (4).

$$\begin{aligned} max(P(x)) - min(P(x)) &= 1 \quad \textit{Stable Sample} \\ max(P(x)) - min(P(x)) &< 1 \quad \textit{Unstable Sample} \end{aligned} \tag{3}$$

Minimum differences of the samples margin point is calculated by (4) where r represents the class, k is the sample number and $P(x_k)^r$ is the probability of k^{th} sample x for class r.

$$\begin{aligned} r &= 1,2 \ldots N \\ k &= 1,2 \ldots T \end{aligned} \qquad min(\max(P(x_k)^r) - \min(P(x_k)^r)) \tag{4}$$

Both algorithms' main steps can be seen on Fig. 1. E represents train algorithm which is many Decision Tree combination with AdaBoost and Bagging methods. Labeled data set is abbreviated as L. U is donated as data pool (rest of all data).

- Generate first train set with 1% of all data randomly selected

Repeat k times the following steps :

- Generate classification ensemble

C = Ensemble Learning (E, L) (E = Adaboost or Bagging Method)

- $\forall x_j \in U$ calculate probabilities for each sample $P(x_k)^r$
- Most informative 1% of data selected in S as samples based on active learning algorithm
- Obtain labels of S, from U
- Delete S from U and add to L

Output:

Ensemble Learning (E, U) (Adaboost or Bagging)

Fig. 1. Algorithms process steps.

III. Application

Algorithms are programmed in Matlab programming interface. Two different active sample selection algorithms are tested on ensemble learning methods (bagging and AdaBoost) formed by 100 decision tree combination. Application and properties of datasets used in tests will be explained in the following sub sections.

TABLE I: Datasets Information Used in Application

Name	Number of Instances	Number of Attributes	Number of Class
d159	7182	33	2
mushroom	8124	112	2
letter	20000	16	26
Kr vs Kp	3196	40	2
ringnorm	7400	20	2

A. Data Information

Five datasets downloaded from UC Irvine Machine Learning Repository [15] are used in application. Active sample selection algorithms are implemented on all datasets and the results are presented on both figures and tables. Algorithms start with 1% of all data as train data and rest of data assumed as data pool where tests applied on. Each iteration, train set increased with 1% of all data up to 70% whereas 1% decreases in test set. Increment in train set is made based on active sample selection and random sample selection algorithms respectively. 70% of all data in datasets is arranged as upper limit to terminate iterations. Number of samples in datasets is important to indicate effects. Large datasets are more useful to show the benefits of using active sample selection algorithms on learning process.

B. Process Steps

Active sample selection algorithms are implemented on both ensemble methods which are bagging and AdaBoost. 100 Decision Trees are combined in ensemble model as base learners.

As a starting point for learners, 1% of all data is arranged as first train data by random selection. First classification result can be vary according to random sample selection but in the second iteration, active selection is applied on methods and effects observed. Besides active sample selection algorithms, also random sample selection is provided in tests to compare the effects on graph.

IV. Test Results

Active learning based on strictly separation algorithm is abbreviated as Algorithm 1 in figures and tables. Additionally, Algorithm 2 represents Margin distances based active learning results. Random Sample Selection method is defined as regular sample selection in application. All tests started with 1% randomly selected data as train data. Each iteration is repeated after 1% increment in train data.

Regular section in the tables also represents random sample selection for the ensemble learning methods. Every 5% part of the testing records is listed in tables. Figures are the graphical demonstration generated from all recorded data in tables.

Classification success in the tables and figures represents the success of the separation for informative samples in data pool. Classification success values become higher when using informative samples in learning process.

Fig. 2 shows the effects of active sample selection on ensemble learning process for d159 dataset. Adaboost method is used in ensemble model. Sample selection based on algorithm 2 makes remarkable difference in results unlike the algorithm 1. Algorithm 1 isn't enough to decide precisely for samples relation to classes, therefore, success rates of the algorithm 1 is near to regular method.

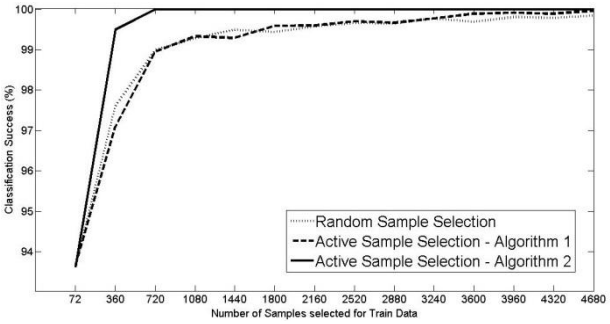

Fig. 2. Classification success of d159 dataset in adaboost method.

Effect of the algorithms on bagging method on the same dataset is demonstrated in Fig. 3. Bagging method in ensemble model gives more successful results in gathering more informative samples of d159 dataset for algorithm 2. Algorithm 2 increased success rates of bagging method more

than adaboost method.

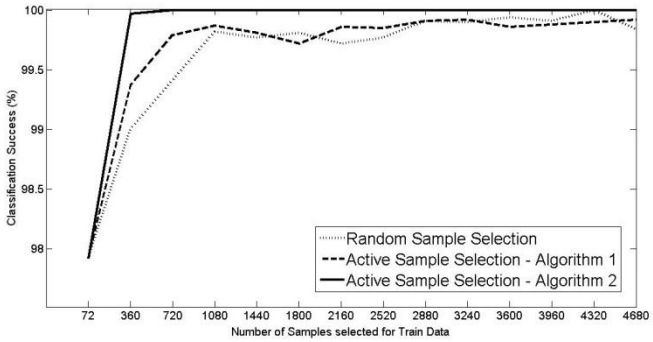

Fig. 3. Classification success of d159 dataset in bagging method.

The General purpose of using active sample selection is that obtain high success rates with minimum number of samples in train. As it can be seen in figures, Algorithm 2 provided it. Algorithm 1 also increased the success compared to regular method especially in bagging method. However, algorithm 2 is more useful for better success. Table II shows numerical results of the figures. The numbers in bold represents the success point resulted with algorithms better than random selection method.

TABLE II: NUMERICAL RESULTS OF D159 DATASET CLASSIFICATION

	Bagging			AdaBoost		
Num.	**Algorithm 1**	**Algorithm 2**	**Regular**	**Algorithm 1**	**Algorithm 2**	**Regular**
72	97,92	97,92	97,92	90,77	90,77	90,77
360	**99,79**	**99,97**	99,00	97,09	**99,50**	97,41
720	**99,87**	**100**	99,41	**98,95**	**100**	98,81
1080	99,81	**100**	99,82	99,33	**100**	99,44
1440	99,72	**100**	99,77	99,29	**100**	99,55
1800	**99,86**	**100**	99,81	**99,59**	**100**	99,49
2160	**99,85**	**100**	99,72	99,60	**100**	99,80
2520	**99,91**	**100**	99,77	99,70	**100**	99,81
2880	**99,92**	**100**	99,91	99,67	**100**	99,77
3240	99,86	**100**	99,90	99,77	**100**	99,85
3600	99,88	**100**	99,94	99,89	**100**	99,89
3960	99,90	**100**	99,91	99,91	**100**	99,91
4320	99,92	**100**	100	**99,89**	**100**	99,86
4680	99,37	**100**	99,84	99,96	**100**	99,96

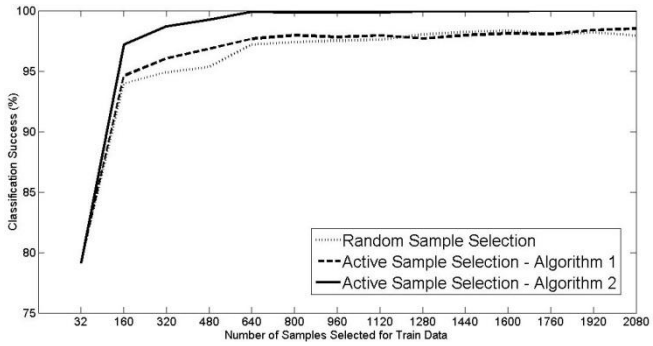

Fig. 4. Classification success of Kr vs Kp dataset in adaboost method.

Table II proves that using active sample selection based on margin distances (algorithm 2) provides better results on separation of unstable samples. It provides more success with the less labeled data. Classification success rates vary according to used ensemble learning methods. Active sample selection based on strict separation (algorithm 1) doesn't have effective role. Also in some stages, random selection has more successful results. The main reason for that is about dataset properties. D159 is in complex structure. Generally, $P(x_k)^r$ values for the samples are not equal to 1. Thus, algorithm 1 results with very close probabilities for samples. Fig. 4 belongs to King-rook vs King-Pawn (Kr vs Kp) dataset trained by ensemble learning with adaboost method. Algorithm 2 has also significant effect on success. It maintains more success with less train data. Besides algorithm 2, algorithm 1 also gives efficient results. Classification success with algorithm 1 reached the same success with regular selection in earlier step which is main purpose of the active sample selection algorithms.

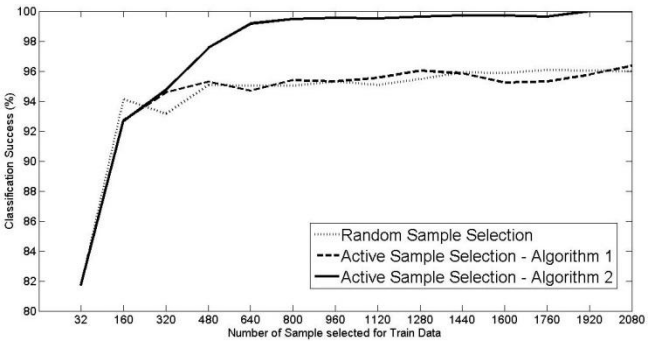

Fig. 5. Classification success of Kr vs Kp dataset in bagging method.

Each machine learning algorithms and also methods is specified for the problems as it is cited in introduction section. Therefore, bagging method is also not convenient for Kr_Ks dataset because of the dataset properties. Regular sample selection method success rate for bagging method isn't efficient but algorithm 2 still gives significant success as it can be observed in Fig. 5. Algorithm 1 resulted with almost the same rates of regular sample selection method.

Active sample selection based on strict separation algorithm is also not useful for all datasets because of different diversity and number of class as uncertainty sampling. Table III also shows that algorithm 2 is more successful than algorithm 1 for Kr vs Kp dataset.

TABLE III: NUMERICAL RESULTS OF KR VS KP DATASET CLASSIFICATION

	Bagging			Adaboost		
Num.	**Algorithm 1**	**Algorithm 2**	**Random**	**Algorithm 1**	**Algorithm 2**	**Random**
32	81,72	81,72	81,72	79,14	79,14	79,14
160	92,70	92,67	94,09	**94,61**	**97,19**	93,95
320	**94,57**	**94,75**	93,16	**96,04**	**98,69**	94,90
480	**95,29**	**97,54**	95,08	**96,84**	**99,24**	95,33
640	94,69	**99,17**	95,03	**97,67**	**99,88**	97,19
800	**95,40**	**99,46**	95,02	**97,97**	**99,86**	97,38
960	**95,32**	**99,54**	95,31	**97,82**	**99,85**	97,52
1120	**95,54**	**99,51**	95,06	**97,94**	**99,84**	97,59
1280	**96,03**	**99,64**	95,46	97,69	**99,94**	98,02
1440	95,85	**99,72**	95,92	97,96	**99,94**	98,23
1600	95,23	**99,70**	95,88	98,11	**99,93**	98,33
1760	95,31	**99,64**	96,06	**98,07**	**100**	98,06
1920	95,77	**100**	96,04	**98,38**	**100**	98,19
2080	**96,05**	**100**	95,75	**98,51**	**100**	97,93

The best ensemble learning method to classify for Letter dataset is adaboost method. Tests are made with only adaboost method for that purpose. Letter dataset is the largest dataset in this study which takes time for process. Fig. 6 indicates the significant success of algorithm 2 in the long run.

Algorithm 1 also increased the success of model in short term. Effects of the algorithms can be observed in Table IV, but it is not noticeable for algorithm 1 in long term. However, Algorithm 2 has more effect on results.

Mushroom dataset is commonly used in studies many times so far. Many classification methods and algorithms are implemented on it. Success rates of the classification for

mushroom dataset are already efficient level in literature. In this study, ensemble learning methods also give decent results. Moreover, success rates are enhanced with the active sample selection algorithms as it can be seen in Fig. 7 and Fig. 8.

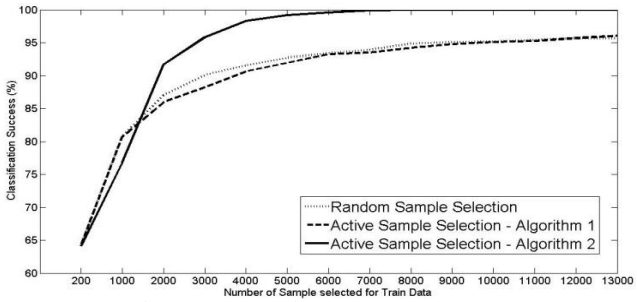

Fig. 6. Classification success of letter dataset in adaboost method .

TABLE IV: NUMERICAL RESULTS OF LETTER DATASET CLASSIFICATION

	AdaBoost		
Num.	Algorithm 1	Algorithm 2	Random
200	64,52	64,52	64,52
1000	**80,70**	76,79	80,06
2000	85,96	**91,71**	86,54
3000	88,21	**95,82**	89,56
4000	90,66	**98,32**	91,04
5000	**91,95**	**99,18**	91,84
6000	**93,27**	**99,60**	92,87
7000	93,50	**99,92**	93,62
8000	94,22	**99,96**	94,30
9000	94,80	**99,98**	94,86
10000	**95,10**	**99,98**	95,06
11000	95,26	**100**	95,55
12000	95,77	**100**	96,04
13000	**96,05**	**100**	95,75

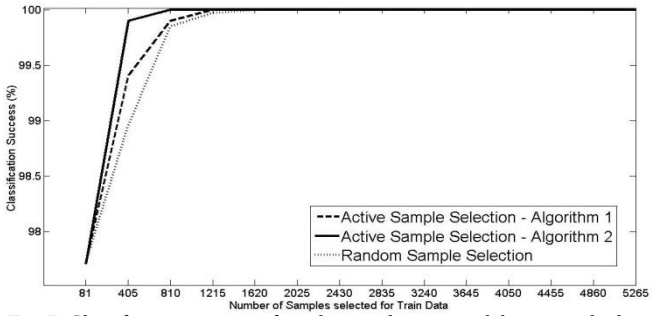

Fig. 7. Classification success of mushroom dataset in adaboost method.

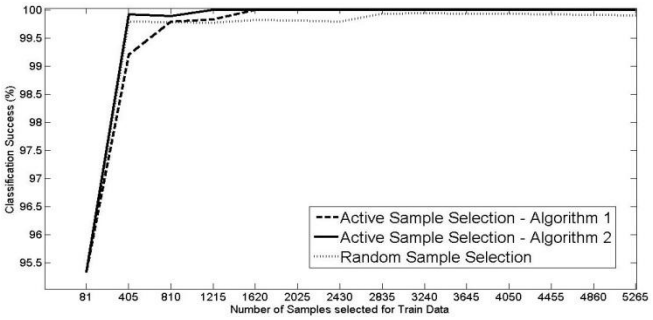

Fig. 8. Classification success of mushroom dataset in bagging method.

Algorithm 1 is also convenient for mushroom dataset because it can be separated easily whereas the regular method has also enough success. Moreover, algorithm 2 results more successful as other tests because of based on not only probabilities between classes but also margin distances.

Success of the algorithm 1 can be seen on Table V more easily. It provides better success over regular method in bagging method.

TABLE V: NUMERICAL RESULTS OF MUSHROOM DATASET CLASSIFICATION

	Bagging			AdaBoost		
Num.	**Algorithm 1**	**Algorithm 2**	**Random**	**Algorithm 1**	**Algorithm 2**	**Random**
81	95,33	95,33	95,33	97,71	97,71	97,71
405	99,20	**99,92**	99,69	99,40	**99,90**	98,95
810	**99,79**	**99,89**	99,55	**99,90**	**100**	99,85
1215	**99,83**	**100**	99,77	**100**	**100**	99,97
1620	**100**	**100**	99,64	**100**	**100**	100
2025	**100**	**100**	99,64	**100**	**100**	100
2430	**100**	**100**	99,90	**100**	**100**	100
2835	**100**	**100**	99,87	**100**	**100**	100
3240	**100**	**100**	99,86	**100**	**100**	100
3645	**100**	**100**	99,89	**100**	**100**	100
4050	**100**	**100**	99,88	**100**	**100**	100
4455	**100**	**100**	99,86	**100**	**100**	100
4860	**100**	**100**	99,85	**100**	**100**	100
5265	**100**	**100**	99,83	**100**	**100**	100

Last tests are made on Ringnorm Dataset. Fig. 9 shows that algorithm 2 plays effective role on success of the model. Success rates are raised up to 100% whereas regular methods around 95%. Algorithm 1 is also efficient on classification in adaboost method as it can be observed in Fig. 9. Both algorithms provide more success rate with the less sample in training.

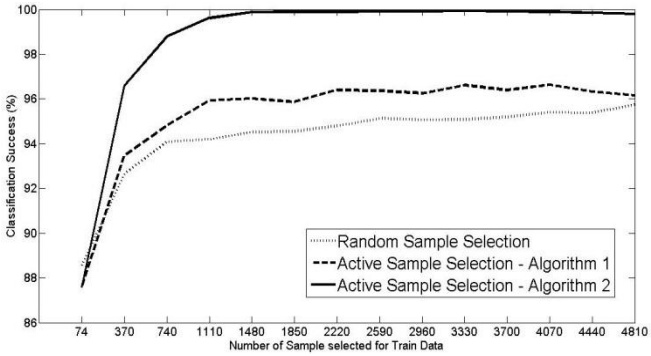

Fig. 9. Classification success of ringnorm dataset in adaboost method.

On the other hand, algorithm 1 is not efficient in bagging method. Algorithm 2 is still has an impact over success as it can be observed in Fig. 10. Main reason is related with the process which is underlying of the algorithms. As noted in section 2, algorithm 1 is based on only probabilities of the samples belonging to classes. However, algorithm 2 also deals with the margin differences between classes.

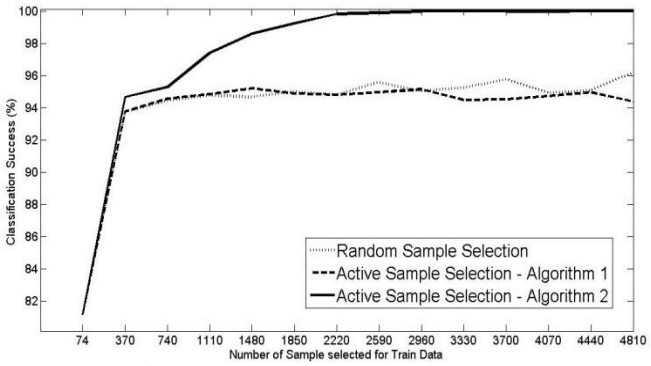

Fig. 10. Classification success of ringnorm dataset in bagging method.

Table VI provides better observation with numerical values. Margin distance active learning algorithm has more sensible effects on success rates because of including not only one probability of the sample but also distribution of the sample probabilities between classes.

TABLE VI: NUMERICAL RESULTS OF RINGNORM DATASET CLASSIFICATION

	Bagging			Adaboost		
Num.	**Algorithm 1**	**Algorithm 2**	**Random**	**Algorithm 1**	**Algorithm 2**	**Random**
74	81,14	81,14	81,14	87,58	87,58	87,58
370	**93,75**	**94,65**	93,70	**93,47**	**96,61**	92,10
740	**94,54**	**95,27**	94,43	94,82	**98,79**	95,11
1110	**94,84**	**97,39**	94,77	**95,93**	**99,62**	94,71
1480	**95,19**	**98,59**	94,66	**96,01**	**99,88**	95,51
1850	94,88	**99,23**	95,00	95,86	**99,90**	95,97
2220	**94,80**	**99,82**	94,78	**96,39**	**99,89**	95,78
2590	94,95	**99,89**	95,57	**96,37**	**99,91**	96,17
2960	**95,13**	**99,98**	95,02	**96,25**	**99,92**	96,10
3330	94,47	**99,98**	95,23	**96,62**	**99,93**	96,11
3700	94,51	**99,98**	95,75	**96,39**	**99,91**	95,48
4070	94,72	**99,97**	94,91	**96,63**	**99,90**	95,49
4440	94,95	**100**	95,06	**96,32**	**99,86**	95,60
4810	94,37	**100**	96,18	96,16	**99,80**	95,38

Starting success rates of some dataset classification can vary according to randomly selection as referred in application section. Starting point is a vital issue in classification algorithms. Intent of this study is to show active learning effects. Starting train set arranged as minimum size (1% of all data) in order to minimize the effect of random selection at initial step. Thus, minimal count of sample with random selection in the beginning can be ignored within this study. During the all test, with the increment of the sample size, success rate also increases. This success can be provided with few samples by active learning methods as it shows in Tables.

Classification rate of first step is inefficient when comparing to stages used active learning sample selection algorithms. In that meaning, this paper also shows that initializing step plays effective role on success besides each iterations because of random selection. Many studies has made on defining starting point of the classification algorithms in literature [11].

V. CONCLUSION

In this study, active learning algorithms' effects are presented. Labeled data is important to make classification with higher success rate in contrast to it is hard to label all data. Studies over this problem focus on rational sample selection theories. Informative samples selection named as active learning sample selection algorithms contributes more success on result compare to randomly sample selection. Presented paper includes two different active learning algorithms on two different ensemble methods concurrently. Five datasets from UC Irvane Machine Learning Repository are used to observe effects. Results prove that samples can be classified with few label data because the samples with rational selection plays effective role. Selecting more informative data for classification provides to achieve same or more success rate at earlier steps. This process is called as active learning.

In this study, it is observed as bagging process makes the train set more complex, hence decision for the algorithm 1 become more complex as well. But algorithm 2 solves that problem with the margin distances. In contrast to bagging method, adaboost method gives different weight values to each sample selected by randomly or active algorithms. Weights values also provide simplicity for algorithms. Therefore, algorithm 1 generally gives more success on only adaboost method.

Randomly selected minimal number of samples (1% of all data) is used to form first classification models on both tests in this study. This approach made the results inefficient at initial state. In future works, some clustering methods will be used for initial sample selection to increase success rate more.

REFERENCES

[1] D. Zhang, "Machine learning in value-based software test data generation," in *Proc. 18th IEEE International Conference on Tools with Artificial Intelligence*, *ICTAI '06*, November 2006, pp.732-736.

[2] M. N. Wernick, Y. Y. Yang; J. G. Brankov, G. Yourganov, and S. C. Strother, "Machine learning in medical imaging," *Signal Processing Magazine, IEEE* , vol. 27, no. 4, pp. 25-38, July 2010.

[3] P. D. Yoo, M. H. Kim and T. Jan, "Machine learning techniques and use of event information for stock market prediction: a survey and evaluation," in *Proc. International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, November 2005, vol. 2, pp. 835-841.

[4] L. Y. Wei, Y. Y. Yang, R. M. Nishikawa, and Y. L. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, pp. 371-380, March 2005

[5] P. Chaovalit and L. N. Zhou, "Movie review mining: a comparison between supervised and unsupervised classification approaches," in *Proc. the 38th Annual Hawaii International Conference on System Sciences, HICSS '05*, January 2005, pp. 112c-112c, 03-06.

[6] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin, Madison.

[7] P. Arabie, L. J. Hubert, and G. Soete, "Clustering and classification," *World Scientific*, London, England, 1999.

[8] N. Abe and H. Mamitsuka, "Query learning strategies using boosting and bagging," in *Proc. the Fifteenth International Conference on Machine Learning*, July 24-27, 1998, pp. 1-9 .

[9] D. Cohn, L. Atlas, and R. Ladner, "Improving generalization with active learning," *Machine Learning*, vol. 15, no. 2, pp. 201-221, 1994.

[10] C. Agan and M. F. Amasyali, "Active learning with committees and the selection of starting sets," in *Proc. Signal Processing and Communications Applications Conference (SIU) 21st*, April 2013, pp. 1-4.

[11] P. Melville and R. Mooney, "Diverse ensembles for active learning," *in Proc. the 21st Int. Conference on Machine Learning*, 2004, pp. 584–591.

[12] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proc. COLT '92*, 1992, pp. 287-294.

[13] H. L. Xu, X. D. Wang, Y. Liao, and C. Y. Zheng, "An uncertainty sampling-based active learning approach for support vector machines," in *Proc. International Conference on Artificial Intelligence and Computational Intelligence, AICI '09*, November 2009, vol. 3, pp. 208-213.

[14] D. Tuia, F. Ratle, F. Pacifici, A. Pozdnoukhov, M. Kanevski, F. D. Frate, D. Solimini, and W. J. Emery, "Active learning of very-high resolution optical imagery with svm: entropy vs margin sampling," in proc. *IEEE International Geoscience and Remote Sensing Symposium, 2008, IGARSS 2008*, July 2008, vol. 4, pp. 73-76.

[15] UC Irvine Machine Learning Repository [Online]. Available: http://archive.ics.uci.edu

Hamza O. İlhan received the B.Sc. degree in electronics and computer science from Marmara University, Istanbul, Turkey in 2010. M.Sc. degree is received in computer engineering from Yalova University, Yalova, Turkey in 2012. He is currently a Ph.D. student in Yıldız Technical Universiy (YTU), Istanbul, Turkey. He is appointed to Yıldız Technical University as a research assistant in 2011. His research interests are in the areas of autonomous robots, image and signal processing, machine learning and pattern recognition with applications to biomedical engineering.

M. Fatih Amasyal received the MSc degree from the Yildiz Technical University, Turkey, in 2003, and the Ph.D. degree from the same university in 2008. Dr. Amasyali is currently an assistant professor at the Computer Engineering Department, Yildiz Technical University. His interests include machine learning, natural language processing and autonomous robotics. He has published several scientific papers.