Comparison of Ensemble-Based Multiple Instance Learning Approaches

Pelin Yıldırım Taşer
Department of Software Engineering
Manisa Celal Bayar University
Manisa, Turkey
pelin@cs.deu.edu.tr

Kökten Ulaş Birant
Department of Computer Engineering
Dokuz Eylul University
Izmir, Turkey
ulas@cs.deu.edu.tr

Derya Birant
Department of Computer Engineering
Dokuz Eylul University
Izmir, Turkey
derya@cs.deu.edu.tr

***Abstract*— Multiple instance learning (MIL) is concerned with learning from training set of bags including multiple feature vectors. This paradigm has various algorithms as a solution for multiple instance problem. Recently, ensemble learning has become one of the most preferred machine learning technique because its high classification ability. The main goal of ensemble learning is combining multiple learning models and obtaining a decision from all outputs of these models. Considering this motivation, the study presented in this paper proposes an ensemble-based multiple instance learning approach which merges standard algorithms (MIWrapper and SimpleMI) with ensemble learning methods (Bagging and AdaBoost) to improve classification ability. The proposed approach includes ensemble of combination of MIWrapper and SimpleMI learners with Naive Bayes, Support Vector Machines (SVM), Neural Networks (Multilayer Perceptron (MLP)), and Decision Tree (C4.5) as base classifiers. In the experimental studies, the proposed ensemble-based approach was compared with individual MIWrapper and SimpleMI algorithms in terms of accuracy. The obtained results indicate that the ensemble-based approach shows higher classification ability than the conventional solutions.**

Keywords—multiple instance learning, ensemble learning, AdaBoost, bagging, machine learning

I. Introduction

Multiple instance learning (MIL) is a variation of supervised learning where each training instances are assigned to sets, called *bag*, and a single class label is specified for all instances in a bag. The goal of MIL is to classify bags with related unlabeled instances. According to this paradigm, if a bag contains at least one positive input, it is labeled as positive. On the contrary, a bag is labeled as negative, if all instances in this bag are negative. This study focuses on combining multiple instance paradigm with ensemble learning techniques.

In the literature, there are several ways of solution for the multiple instance learning problem. One of them is converting the data including bags to a standard representation. The most popular algorithms of this technique are MIWrapper and SimpleMI. Considering this motivation, MIWrapper and SimpleMI algorithms were selected with Naive Bayes, Support Vector Machines (SVM), Neural Networks (Multilayer Perceptron (MLP)), and Decision Tree (C4.5) as base learners for Bagging and Boosting ensemble methods.

In the last decide, ensemble learning has become mostly preferred machine learning technique due to high classification ability it provides. *Ensemble learning* is a machine learning paradigm which combines a set of individual learners to obtain single final class label. The aim of ensemble approach is improving classification accuracy of the system by merging different decisions of the models. Ensemble learning methods are divided into four main types: Bagging, Boosting, Stacking, and Voting. In this study, bagging and boosting methods were selected for the construction of ensemble structure.

The novelty and main contributions of this paper are as follows: (i) it provides a brief survey of multiple instance learning and ensemble learning which are combined to improve classification performance, (ii) it proposes an ensemble-based multiple instance learning approach which consists of combination of AdaBoost and Bagging versions of MIWrapper and SimpleMI learners with Naive Bayes, Support Vector Machines (SVM), Neural Networks (Multilayer Perceptron (MLP)), and Decision Tree (C4.5) as base classifiers, and (iii) it presents experimental studies conducted on five different real-world multiple instance datasets to demonstrate that the proposed ensemble-based approach shows better classification results than individual multiple instance algorithms in terms of accuracy rates.

The remainder of this paper is structured as follows: In the following section, related literature and previous works on the subject are given. In Section 3, background information about multiple instance learning and are explained. In addition, this section also describes the proposed ensemble-based multiple instance learning approach in detail. Section 4 explains benchmark multiple instance datasets used in the experimental studies and gives information about the application of the proposed model on the datasets. In this section, the obtained results are presented with discussions. Finally, some concluding remarks and future directions are given in Section 5.

II. Related Work

Multiple instance learning has been widely used in many fields, including education [1, 2], text mining [3], health [4 - 6], image categorization [7], bioinformatics [8, 9], and computer networks [10]. Quellec et al. [5] developed a multiple instance learning framework for classifying diabetic retinopathy screening in 2D retinal images as relevant or irrelevant using content-based image retrieval. In contrast to these studies, the research presented in this paper proposes an ensemble-based multiple instance learning approach that combines a set of individual learners using MIWrapper and SimpleMI algorithms.

In the literature, there are several studies which implements MIWrapper and SimpleMI algorithms to solve multiple instance learning problems [2, 11 - 13]. Tarrago et al. [11] proposed a novel multi-instance learning wrapper method using the Rocchio classifier for the web index recommendation (WIR) problem. The proposed approach was

978-1-7281-1862-8/19/$31.00 ©2019 IEEE

conducted on nine WIR benchmark data sets to demonstrate its performance.

Nowadays, ensemble learning techniques which has been one of the active research fields in machine learning have been commenced to be combined with multiple instance learners [14-16]. For example, Sastrawaha and Horata [14] developed an ensemble extreme learning machine for multi-instance learning (E-ELM-MIL) which combines several single hidden layer neural networks and predicts the final label using majority voting mechanism. In the other study, researchers proposed an ensemble approach for multi-view multi-instance learning [15]. In this study, multiple view learners are combined and a class label is determined according to consensus among the weighted class predictions.

Differently from existing studies, the study presented in this paper focuses on the application of ensemble-based multiple instance learning approach on various real-world and well-known multi-instance datasets. It proposes an ensemble multiple instance learner which consists of combination of MIWrapper and SimpleMI learners with traditional learners: Naive Bayes, Support Vector Machines (SVM), Neural Networks (Multilayer Perceptron (MLP)), and Decision Tree (C4.5) as base classifiers to enhance classification accuracies.

III. Material and Methods

A. Multiple Instance Learning

In traditional supervised learning techniques, training sets consists of feature vectors including certain class labels. However, in *multiple instance learning*, training sets consists of bags including various feature vectors and each bag is assigned a class label [17]. As opposed to the standard single-instance learning, we have no information about the feature vectors' class labels of bags in multiple instance learning. A multiple instance (MI) learning algorithm learns an MI dataset which consists of bags of instances, rather than single instances. As shown in Figure 1, only the bags are labeled, and each can contain many instances. The number of instances in each bag can vary independently of other bags, and the same instance can belong to different bags.

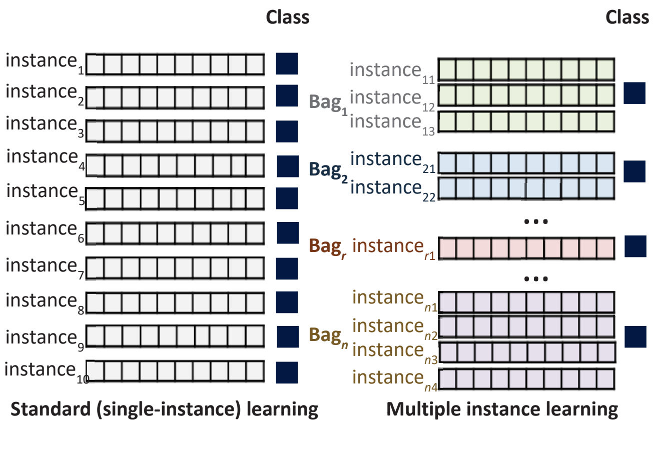

Fig. 1. Standard (single-instance) learning and multiple instance learning.

The description of multiple instance learning is as follows. Let χ as the instance space. Given a dataset D consists of a set of bags, denoted by $D = \{(B_1,L_1), (B_2,L_2), ..., (B_n,L_n)\}$, where $B_i = \{x_{i1}, x_{i2},...,x_{im}\} \subset \chi$ is called a *bag*, and n is the number of training bags. Here $x_{ij} \in B_i$ is the j^{th} instance in bag i and m is the number of intances in B_i. Let $L_i \in L$ is the class label of B_i, i.e. for binary classification $L=\{0,1\}$. The aim is to built a multiple instance learner $h(B)$ for correctly labeling unseen bags. The multiple instance learner works at the bag level and takes a bag as its input and generates a decision for the bag. Actually, a standard (single-instance) learner is a special case where each bag contains only one instance $B_i=\{x_{i1}\}$.

There are numerous MIL learners in the literature: Citation-kNN (k-Nearest Neighbor), MIDD (Diverse Density for MIL), MDD (Modified DD), MISVM (Support Vector Machines for MIL), MIWrapper, SimpleMI, etc. In this study, MIWrapper and SimpleMI algorithms were applied on five real-world MIL datasets individually and then they were combined with AdaBoost (Adaptive Boosting) and Bagging ensemble methods in the proposed approach.

- **MIWrapper:** MIWrapper is a multiple instance learner, introduced by Eibe Frank and Xin Xu [18]. It constructs a training set using inheritance rule as follows:

$$LR(X) = \frac{1}{|X|}\sum_{x \in X} lr(x) \tag{1}$$

where *LR(x)* bag-level learner and *lr(x)* defined as instance-level learner. Then, each instance in the bag is assigned the weight value of *1/m.* It means that the instances in the bags have the same weight values, so they have an equal impact on class labels.

- **SimpleMI:** It is a simple and fast method which implements traditional single-instance learning algorithms on multiple instance learning applications [19]. In this approach, only one instance is built to symbolize the whole bag. This summarizing process is generally performed by evaluating arithmetic mean M of instances in each bag as follows:

$$M(X) = \frac{1}{|X|}\sum_{x \in X}(x) \tag{2}$$

B. Ensemble-Based Multiple Instance Learning

Ensemble learning is a machine learning approach which merges multiple learners and classify an instance by a voting mechanism over all learners. The most commonly preferred voting mechanisms are major class labels for categorical target attributes, and average, weighted average, median, minimum, and maximum operations for numerical target attributes.

The goal of ensemble learning technique is obtaining a strong classifier from a set of individual learners to improve classification performance. Many ensemble learning studies stated that ensemble-based approaches show more accurate classification results than the individual learners.

The ensemble learning techniques are classified into four types: Bagging, Boosting, Stacking, and Voting. In our work, Bagging and Boosting (AdaBoost algorithm) methods with MIWrapper and SimpleMI learners were implemented on five different datasets.

- **Bagging:** Bagging (bootstrap aggregating) is commonly implemented ensemble method that chooses instances randomly from the dataset and builds training subsets using these instances. In this structure, individual learners are trained with these subsets separately and so multiple learners are comprised. Finally, the predicted labels from each learner are aggregated and a final output is obtained by a voting mechanism.

- **Boosting:** The aim of boosting method is obtaining strong learners from weak ones. Firstly, the same weight value is assigned to each sample in the training set. Then, misclassified instances' weight values are increased and conversely the weight of correctly classified samples are decreased.

  The most commonly applied boosting algorithm is AdaBoost which implements boosting process on each instances in the dataset. So, multiple learners are generated and the obtained outputs from them are aggregated using a weighted voting method. In this study, AdaBoost algorithm was applied as boosting method.

In the proposed ensemble approach, Naive Bayes, Support Vector Machines (SVM), Neural Networks (Multilayer Perceptron (MLP)), and Decision Tree (C4.5) classifiers are used as base learners for bagging and AdaBoost methods separately.

The general structure of the proposed approach is presented in Figure 2. Firstly, multiple training subsets are generated according to applied ensemble method. While samples in the dataset are selected randomly in bagging method, samples are selected considering their weight values. Then, a learner that produced as a combination of ensemble methods and MIL algorithms with single-instance classifiers (Bagging.MIWrapper.NaiveBayes, AdaBoost.MIWrapper.MLP, AdaBoost.SimpleMI.C4.5, etc.) is applied on the training subsets separately and so multiple ensemble-based multiple instance learners are obtained. Each learners produces an output label. The majority vote of these labels is chosen as final predicted class label.

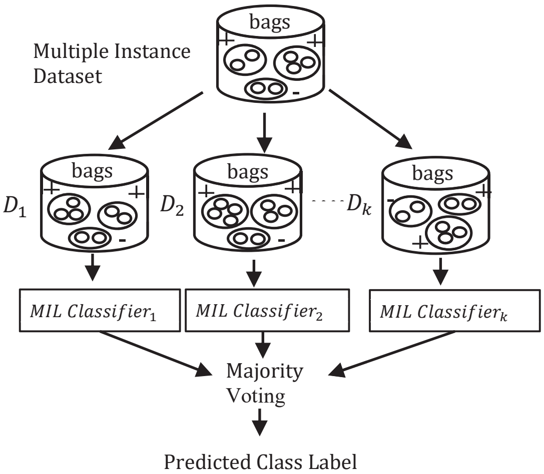

Fig. 2. The general structure of proposed ensmeble-based MIL approach.

IV. Experimental Studies

In the experimental studies, the proposed ensemble-based was applied on five different well-known and real-world multiple instance datasets to demonstrate its classification success over the present multiple instance learning algorithms such as MIWrapper and SimpleMI. The application was developed by using Weka open source data mining library [20]. The standard MIL algorithms (MIWrapper and SimpleMI) and the proposed ensemble-based MIL approach (bagging and AdaBoost) using four different classification algorithms (Naive Bayes, SVM, MLP, and C4.5) as base learners were applied on the MIL datasets separately and they were compared in terms of accuracy rates. The experimental results in this study are presented with the help of tables and graphs.

A. *Dataset Description*

In this experimental study, six different datasets which are available for public use were selected to present the proposed ensemble-based MIL approach's classification performance. The utilized datasets were obtained from the data archive in Sourceforge.net's Weka directory [21]. These datasets were obtained for prediction of molecule activities. The detailed descriptions about the investigated MIL datasets are given in Table 1.

TABLE I. Dataset Characteristics

Dataset	# of Bags			Average Bag Size	# of Instances
	Positive	Negative	Total		
Musk1	47	45	92	5.17	476
Musk2	39	63	102	64.69	6598
Mutagenesis-atoms	125	63	188	8.61	1618
Mutagenesis-bonds	125	63	188	21.25	3995
Mutagenesis-chains	125	63	188	28.45	5349

B. *Experimental Results*

In this study, two different MIL algorithms such as MIWrapper and SimpleMI were preferred for the solution for MIL and applied on five different benchmark datasets in the experiments separately. In the first experiment, MIWrapper and SimpleMI algorithms were applied on these datasets using traditional classification algorithms: Naive Bayes, SVM, MLP, and C4.5 as base learners with default parameters. Then, in the second experiment, AdaBoost and Bagging versions of these single solutions were implemented on the same datasets by modifying iteration number as 100. Thus, individual, AdaBoost-Based and Bagging-Based approaches were compared with each other in terms of classification accuracy. In order to evaluate the accuracy rates of them, *n*-fold cross validation technique was implemented by selecting *n* as 10.

The computed accuracy rates of the implemented approaches on five different benchmark datasets in the experiments are presented in Table 2 and Table 3. In these tables, the ensemble-based algorithms' accuracy rates which have higher values than related single approaches are marked with the character *. Furthermore, the algorithm which gives the highest accuracy rate among the single, AdaBoost and Bagging versions of it is represented by making it bold. The obtained results show that the proposed ensemble-based MIL approach (AdaBoost and Bagging based versions of MIWrapper and SimpleMI algorithms) generally gives more accurate classification results than the individual MIWrapper and SimpleMI algorithms. For example, AdaBoost.MIWrapper.SVM (84.31%) and Bagging.MIWrapper.SVM (83.33%) algorithms show higher classification accuracy than single MIWrapper.SVM (82.35%) algorithm on Musk 2 dataset. When the average classification accuracies for all used datasets, the proposed method in this paper shows generally better classification performance than the traditional ones.

TABLE II. THE COMPARISON OF MIWRAPPER BASED ALGORITHMS WITH THEIR ENSEMBLE VERSIONS IN TERMS OF ACCURACY

	Datasets ⇨ / Algorithms ⇩	Accuracy (%)					
		Musk1	**Musk2**	**Mutagenesis-atoms**	**Mutagenesis-bonds**	**Mutagenesis-chains**	**Average**
Single	MIWrapper.NaiveBayes	78.26	77.45	66.49	63.83	58.51	68.91
	MIWrapper.SVM	84.78	82.35	66.49	66.49	67.02	73.43
	MIWrapper.MLP	82.61	60.78	67.55	71.81	77.66	72.08
	MIWrapper.C4.5	81.52	78.43	77.13	81.38	85.69	80.83
AdaBoost-Based	AdaBoost.MIWrapper.NaiveBayes	83.7*	84.32*	66.49	67.55*	61.17*	**72.65**
	AdaBoost.MIWrapper.SVM	82.61	84.31*	67.55*	69.68*	80.32*	**76.89**
	AdaBoost.MIWrapper.MLP	85.87*	79.41*	73.4*	82.98*	85.64*	**81.46**
	AdaBoost.MIWrapper.C4.5	86.96*	85.29*	86.7*	81.91*	82.45	**84.66**
Bagging-Based	Bagging.MIWrapper.NaiveBayes	77.17	73.53	66.49	65.96*	51.6	66.95
	Bagging.MIWrapper.SVM	88.04*	83.33*	66.49	66.49	66.49	**74.17**
	Bagging.MIWrapper.MLP	85.87*	68.63*	73.4*	75*	66.49	**73.88**
	Bagging.MIWrapper.C4.5	90.22*	85.29*	79.79*	80.32	86.7*	**84.46**

TABLE III. THE COMPARISON OF SIMPLEMI BASED ALGORITHMS WITH THEIR ENSEMBLE VERSIONS IN TERMS OF ACCURACY

	Datasets ⇨ / Algorithms ⇩	Accuracy (%)					
		Musk1	**Musk2**	**Mutagenesis-atoms**	**Mutagenesis-bonds**	**Mutagenesis-chains**	**Average**
Single	SimpleMI.NaiveBayes	85.87	80.39	68.09	73.94	78.72	77.4
	SimpleMI.SVM	85.87	78.43	68.09	73.94	79.79	77.22
	SimpleMI.MLP	81.52	82.35	78.19	85.64	83.51	82.24
	SimpleMI.C4.5	83.7	79.41	79.26	80.32	79.79	80.5
AdaBoost-Based	AdaBoost.SimpleMI.NaiveBayes	84.78	91.18*	69.15*	80.32*	81.91*	**81.47**
	AdaBoost.SimpleMI.SVM	85.87	81.37*	74.47*	78.19*	80.32*	**80.04**
	AdaBoost.SimpleMI.MLP	85.87*	83.33*	79.79*	84.57	82.98	**83.31**
	AdaBoost.SimpleMI.C4.5	81.52	83.33*	80.32*	86.7*	82.45*	**82.86**
Bagging-Based	Bagging.SimpleMI.NaiveBayes	83.7	81.37*	69.15*	73.4	80.32*	**77.59**
	Bagging.SimpleMI.SVM	88.04*	79.41*	68.62*	73.4	79.26	**77.75**
	Bagging.SimpleMI.MLP	88.04*	80.39	79.79*	84.57	84.57*	**83.47**
	Bagging.SimpleMI.C4.5	84.78*	84.31*	80.85*	88.3*	82.98*	**84.24**

The graphs presented in Figure 3 and 4 show the average rank values of the individual MIWrapper (Figure 3) and SimpleMI (Figure 4) algorithms with Naive Bayes, SVM, MLP, and C4.5 base learners and AdaBoost and Bagging based versions of them on the five different MIL datasets. In this comparative method, accuracy rates of the applied approaches are rated by firstly assigning 1 to the approach which provides the highest classification accuracy and then the rank values are increased until giving r value to the most unsuccessful approach. In the case of tie, average ranking values of the approaches are evaluated by taking mean values of the ranks of them. According to this ranking method, the applied approach which has the lowest ranking value is the most successful classifier. Thus, the results indicate that AdaBoost and Bagging based approaches give more accurate classification results than the individual versions for both MIWrapper (Figure 3) and SimpleMI (Figure 4) algorithms.

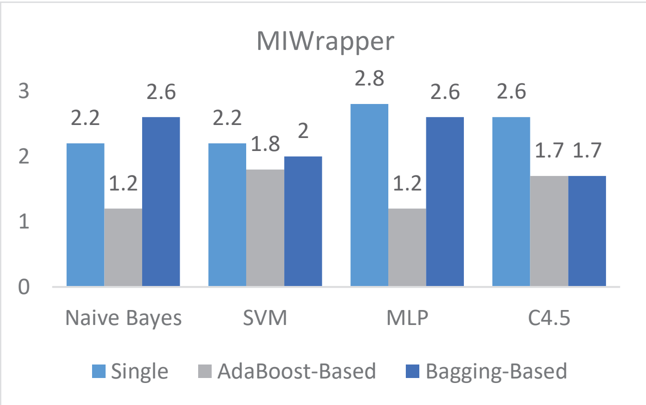

Fig. 3. The average ranks of MIWrapper algorithm with base learners and AdaBoost and Bagging based versions.

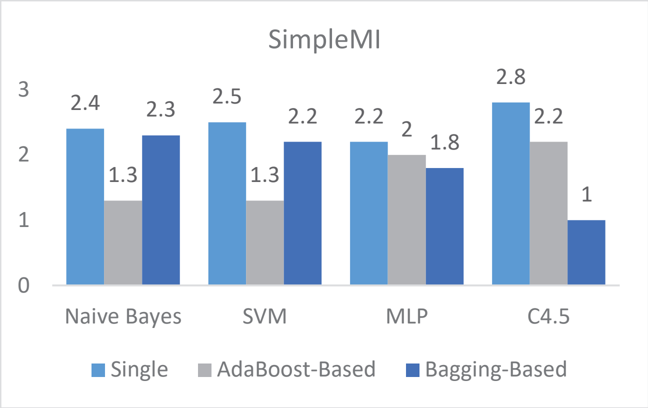

Fig. 4. The average ranks of SimpleMI algorithm with base learners and AdaBoost and Bagging based versions.

V. Conclusion and Future Works

Multiple instance learning has become an area which has received pretty much attention in machine learning community. This paradigm is proposed for an appropriate solution for real-world classification problems. On the contrary of supervised learning technique, in MIL we have no prior information about the class labels of feature vectors in each bag. Meanwhile, ensemble learning approach shows remarkable classification results than traditional individual classifiers in various areas. Considering these reasons, we have presented an ensemble-based MIL approach which consists of combination of AdaBoost and Bagging versions of MIWrapper and SimpleMI learners with Naive Bayes, Support Vector Machines (SVM), Neural Networks (Multilayer Perceptron (MLP)), and Decision Tree (C4.5) as base classifiers. We have applied the proposed approaches on five different real-world multiple instance datasets and empirically evaluated the performance of these approaches by comparing with individual standard classifiers in terms of accuracy. The experimental results stated that the proposed ensemble-based MIL approach provides higher classification performance than the conventional solutions.

As future work, the other ensemble learning methods that were not implemented in this study such as stacking and voting can be applied on MIL datasets. Furthermore, ensemble-based multiple instance clustering study can be developed. This ensemble structure can be constructed by using different training sets, different features, different clustering algorithms or different parameters.

References

[1] A. Zafra, C. Romero, and S. Ventura, “Multiple instance learning for classifying students in learning management systems,” Expert Sys. App., vol. 38, no. 12, pp. 15020-15031, 2011.

[2] S. A. Kumar and M. N. Vijayalakshmi, “Efficiency of Multi-instance Learning in Educational Data Mining,” in Knowledge Computing and its Applications, S. M. Anouncia and U. Wiil, Eds. Singapore: Springer, 2018, pp 47-64.

[3] M. Surdeanu, J. Tibshirani, R. Nallapati, and C. D. Manning, “Multi-instance multi-label learning for relation extraction,” Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Stroudsburg, PA, USA, pp. 455-465, July, 2012.

[4] Y. Xu, T. Mo, Q. Fengi, P. Zhong, M. Lai, and E. C. Chang, “Deep learning of feature representation with multiple instance learning for medical image analysis,” IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 1626-1630, May, 2014.

[5] G. Quellec et al., “A multiple-instance learning framework for diabetic retinopathy screening,” Med. Image Anal., vol. 16, no. 6, pp. 1228-1240, 2012.

[6] L. Chen et al., “Identification of cerebral small vessel disease using multiple instance learning”, in MICCAI, N. Navab et al., Eds. Switzerland: Springer, 2015, pp. 523-530.

[7] Y. Chen and J. Z. Wang, “Image categorization by learning and reasoning with regions,” J. Mach. Learn. Res., vol. 5, pp. 913-939, 2004.

[8] O. Kraus, J. Ba, and B. Frey, “Classifying and segmenting microscopy images with deep multiple instance learning,” Bioinform., vol. 32, no. 12, pp. i52-i59, 2016.

[9] F. Minhas and A. Ben-Hur, “Multiple instance learning of Calmodulin binding sites,” Bioinform., vol. 28, no. 18, pp. i416-i422, 2012.

[10] L. Kok, S-L Chua, C-K Ho, L. K. Foo, and M. R. Bi M. Ramly, “Improving network service fault prediction performance with multi-instance learning”, in Computational Science and Technology, Lecture Notes in Electrical Engineering, R. Alfread et al., Eds. Singapore: Springer, 2019, pp 249-257.

[11] D. Tarragó, C. Cornelis, R. Bello, and F. Herrera, “A multi-instance learning wrapper based on the Rocchio classifier for web index recommendation,” Knowl-Based Syst., vol. 59, pp. 173-181, 2014.

[12] G. Naqvi, “A hybrid filter-wrapper approach for feature selection,” M.S. thesis, Department of Technology, Örebro Univ., Örebro, 2012.

[13] S. Kauschke, M. Mühlhauser, and J. Fürnkranz, “Leveraging reproduction-error representations for multi-instance classification”, in DS 2018, L. Soldatova et al., Eds. Switzerland: Springer, 2018, pp. 83-95.

[14] S. Sastrawaha and P. Horata, “Ensemble extreme learning machine for multi-instance learning”, Proceedings of the 9th International Conference on Machine Learning and Computing (ICMLC), Singapore, pp. 56-60, February, 2017.

[15] A. Cano, “An ensemble approach to multi-view multi-instance learning”, Knowl-Based Syst., vol. 136, pp. 46-57, 2017.

[16] X. Song, L. C. Jiao, S. Yang, X. Zhang, and F. Shang, “Sparse coding and classifier ensemble based multi-instance learning for image categorization”, Sig. P., vol. 93, pp. 1-11, 2012.

[17] J. Amores, “Multiple instance classification: Review, taxonomy and comparative study,” Artif. Intell., vol. 201, pp. 81-105, 2013.

[18] E.Frank, X. Xu, “Applying propositional learning algorithms to multi-instance data,” Tech. rep., Department of Computer Science, University of Waikato, 2003.

[19] L. Dong, “A comparison of multi-instance learning algorithms,” M.S. thesis, Department of Computer Science, The University of Waikato, Hamilton, NewZealand, February, 2006.

[20] “Weka 3 - Data Mining with Open Source Machine Learning Software in Java,” Cs.waikato.ac.nz, 2019. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/. [Accessed: 09- Mar- 2019].

[21] “Weka - Browse Files at SourceForge.net,” Sourceforge.net, 2019. [Online]. Available: https://sourceforge.net/projects/weka/files/. [Accessed: 09- Mar- 2019].