

A Stacking-based Ensemble Learning Method for Outlier Detection

AA. ABRO, E. TAŞCI, A. UĞUR

Abstract—Outlier detection is considered as one of the crucial research areas for data mining. Many methods have been studied widely and utilized for achieving better results in outlier detection from existing literature; however, the effects of these few ways are inadequate. In this paper, a stacking-based ensemble classifier has been proposed along with four base learners (namely, Rotation Forest, Random Forest, Bagging and Boosting) and a Meta-learner (namely, Logistic Regression) to progress the outlier detection performance. The proposed mechanism is evaluated on five datasets from the ODDS library by adopting five performance criteria. The experimental outcomes demonstrate that the proposed method outperforms than the conventional ensemble approaches concerning the accuracy, AUC (Area Under Curve), precision, recall and F-measure values. This method can be used for image recognition and machine learning problems, such as binary classification.

Index Terms— Outlier detection, Ensemble learning, Machine Learning, Classification, Data Mining.

I. INTRODUCTION

OUTLIER IS defined as an observation that deviates from other observations or suspicious events that are generated by different mechanisms. Outliers are anomalous, irregular, or outlying reflections, the distortion of estimations in statistical models [1].

This is one of the best approaches of data analysis to deal with observations having numerous datasets, as automated tools are being used in it to find patterns and relationships. In recent years, outlier detection has been widely used in several industries, such as medical, to detect credit card frauds and sensors (IoT).

ABDUL AHAD ABRO, is with the Department of Computer Engineering, Ege University, Izmir, Turkey, (e-mail: abdualahadabro1@gmail.com).



<https://orcid.org/0000-0002-3591-9231>

ERDAL TAŞCI, is with the Department of Computer Engineering, Ege University, Izmir, Turkey, (e-mail: arif.erdal.tasci@ege.edu.tr).



<https://orcid.org/0000-0001-6754-2187>

AYBARS UĞUR, is with the Department of Computer Engineering, Ege University, Izmir, Turkey, (e-mail: aybars.ugur@ege.edu.tr).



<https://orcid.org/0000-0003-3622-7672>

Manuscript received January 24, 2020; accepted April 14, 2020.
DOI: [10.17694/bajece.679662](https://doi.org/10.17694/bajece.679662)

Ensemble Learning is a machine learning technique that aggregates various base models to generate a single predictive model. Numerous methods are used in Ensemble Learning to reduce bias (boosting), variance (bagging), or to progress predictions (stacking) [2]. It also means that the concept provides a promissory field of future research.

While Random Forest was developed approximately two decades ago, it gives a powerful performance, simplicity in implementation and interpretability [3].

On the other hand, Rotation Forest, which is proposed by Pardo [4-5], provides favourable outcomes when compared to AdaBoost, Random Subspaces, Bagging and Iterated Bagging. The principal contribution of this paper is a) A stacking-based ensemble learning method which improves the outlier detection performance ii) A comparative analysis of four base learners and one Meta-ensemble learner on five datasets from the ODDS library in terms of five evaluation criteria; accuracy (Acc), AUC, precision, recall and F-measure.

This paper is structured with several different sections. In section II, related work presents ideas about ensemble methods. Section III discusses the proposed method in detail. Section IV, provides experimental work, detail of datasets and outcomes. Section V, is related to the evaluation of performance and results. Lastly, conclusion and future work are suggested in Section VI.

II. RELATED WORK

Outliers are mainly segregated into three main areas: Collective outliers, global outliers and contextual outliers [3]. Global Outliers consider that outliers are associated with all the available data points. Contextual Outliers consider that data separated from other data points in the context. However, Collective Outlier values are different data groups that are inaccurate according to a complete dataset. Outlier values are also known as abnormal as they examine the change to identify unexpected behaviour [4].

A static ensemble shows the base learner and the fusion rule is fixed for each single test point [5]. Generally, Bagging and Random subspace methods are employed in these processes. For instance, methods used to generate numerous diverse training subsets for base learners are combined in bagging and random subspace. Many ensemble approach are also applied over clustering algorithms. Therefore, the aggregation and structure standard of the ensemble is set for each single test point in this form of outlier detection strategy [6].

In other studies, Rotboost is a classifier of an ensemble, inferred by combining the AdaBoost and Rotation forest. There are various datasets from the UCI ML repository,

among which a classification tree that is being utilized as the base learning algorithm. It has been shown by their results that Rotboost could generate a lower prediction error in an ensemble classifier in comparison to Rotation Forest or AdaBoost [5]. The ensemble learning approaches such as bagging mainly emphasis to get an ensemble model with less variance than its components; whereas, boosting and stacking generally try to generate strong models less biased than their components even if variance can also be condensed. Random Forest (a subprocess of the Meta-ensemble method) is used as a base learner in the rotation forest. This approach has enhanced performances [7].

In [8], polarized images have been classified using Random Forest and Rotation Forest and it is concluded that Rotation Forest provides more accurate results than SVM and Random Forest; however, Random Forest provides faster results than Rotation Forest.

It is examined whether Rotation Forest is the best classifier that assists in resolving problems with continuity or not. [6].

In [9], A-Stacking and A-Bagging, the adaptive versions of ensemble learning approaches are proposed. A-Bagging method has been applied by using the same base learners over numerous subsets of data and the predictions are aggregated by using weighted majority voting.

In [10], it is shown that ML algorithms provide satisfactory performance for the prediction of the outcomes in comparison with logistic regression.

III. DETAILS OF THE PROPOSED METHOD

In this paper, we have proposed a framework of a Stacking-based ensemble learning method, including rotation forest, random forest, bagging, boosting and logistic regression. There are numerous phases of the system, such as related with datasets, base and stacking-based ensemble learners. In order to obtain the generalization performance of the system, 10-fold cross-validation is used for all learners and datasets. The ranges of the values in data pre-processing may be high when compared to non-outlier datasets. In this scenario, classification algorithms could be affected significantly or negatively by some features.

In this work, four base learners and one Meta-learner are employed with one Stacking-based Meta classifier. Rotation Forest classifier depends upon feature extraction for ensembles. Typically, it provides more authentic results than AdaBoost and Random Forest. The Random Forest classifier is based on several collections of tree classifiers and randomly selected sub-spaces of data are being used to create each classifier independently.

Ensemble Learning such as bagging and boosting assist in diminishing various influences such as classification error. Furthermore, combinations of many classifiers drop variance, particularly in the case of unstable classifiers, which may generate a more reliable classification than a single classifier. The main idea of this study is to establish and provide data comprised of detecting outliers to present new methods related to outlier detection in classification with logistic regression.

Whereas, logistic regression predicts to analyze, explain and indicate the interrelation between one nominal and a dependent binary variable, ratio-level independent or interval

variables. Weka (Waikato Environment for Knowledge Analysis)[11], could examine and test multiple outliers[12], without losing the impact of swamping and masking. We demonstrated the behaviour of our method through simulation with different percentages of outliers and sample sizes. In this process, the different datasets have been utilized referred to from the ODDS library.

IV. EXPERIMENTAL WORK

In the experimental process, five datasets have been used from the ODDS library for classifications [13].

The characteristics of datasets are analyzed concerning the attributes and the number of instances. These datasets are generally used to solve issues related to machine learning. There are no missing values in these datasets and there are various numerical attribute descriptions, which are illustrated in Table I. As it can be observed from Table I, various datasets, the number of samples, dims and outliers are presented for each dataset. Datasets are chosen according to their distinct parameters from the ODDS library source. It is determined by investigating the appropriate data or datasets which are being utilized in the findings of outliers. The proposed stacking-based ensemble learning method has been introduced for this process. This method utilized the imbalanced classification problems of binary (two-class) where the positive case such as (class 1) is taken as an outlier and negative case (class 0) is taken as normal.

TABLE I
DATASETS DESCRIPTIONS

Datasets	Samples	Dims	Outliers
Class	214	9	9 (4.2%)
Letter Recognition	1600	32	100(6.25%)
Shuttle	49097	9	3511 (7%)
Forest Cover	286048	10	2747(0.9%)
Vertebral	240	6	30 (12.5%)

In this work, four different ensemble learning approaches have been carried out along with the ensemble learning method, which is considered suitable for the detection of outliers. However, the performance metrics are calculated based on outlier detection according to binary classification problems. In this method, a technique has been used, which is known as logistic regression from the field of statistics and it is being used to solve binary classification issues. A stacking-based ensemble method, along with logistic regression and four different baseline methods have been presented in Fig. 1.

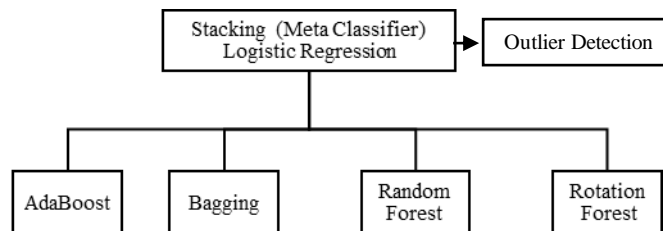


Fig.1. Ensemble Learning baseline methods

Bagging is a modest and very influential ensemble process. It is considered as the Bootstrap procedure to a high-variance ML algorithm. Simultaneously, Boosting denotes a group of

algorithms that employ weighted averages to interchange the weak learners into stronger learners. The random forest consists of multiple random decision trees [14]. Rotation forest is a tree-based ensemble that performs and transforms on subsets of attributes before constructing each tree.

V. PERFORMANCE EVALUATION

A. Evaluation Measures

This section describes the five performance evaluation measures of the proposed method, consisting of accuracy, AUC, precision, recall and F-measure.

Accuracy represents how near a measurement is to an identified or accepted figure. It is further defined in Eq.1.

$$Acc = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

In equation 1, TN, FN, FP and TP show the number of True Negatives, False Negatives, False Positives and True Positives.

AUC represents the Area under the ROC Curve. AUC calculates the whole two-dimensional area beneath the whole ROC curve from (0,0) to (1,1).

Precision is a positive analytical value [15]. Precision defines how reliable measurements are, although they are farther from the accepted value.

The equation of precision is shown in Eq.2.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

The Recall is the hit rate [15]. The recall is the reverse of precision; it calculates false negatives against true positives. The equation is illustrated in Eq. 3.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F-measure can be defined as the weighted average [16] of precision and recall. This rating considers both false positives and false negatives. The equation is illustrated in Eq. 4.

$$F = 2x \frac{Precision+Recall}{Precision+Recall} \quad (4)$$

Tables II-VII present accuracy, AUC, precision, recall and F-measure individual values with ensemble methods for all datasets.

To sum up, Tables II-VI, have been designed according to the diverse data sets concerning the numerous approaches of ensemble learning in terms of different specifications. In Table II, logistic regression has better outcomes, which provides 99.5327% Acc in comparison to others. Likely, in Table III, rotation forest indicates 95.1875% Acc adequate consequences. Similarly, in Table IV, the random forest presents 99.9939% Acc effective results. Likewise, in Table

V, the random forest illustrates the 99.9857% Acc productive outcomes. However, in the end, logistic regression shows a 92.5% Acc result in Table VI.

TABLE II
RESULTS OF ENSEMBLE LEARNING METHODS BY UTILIZING THE GLASS DATASET

Glass					
Methods	Acc (%)	AUC	Precision	Recall	F-Measure
Bagging	96.2617	0.988	0.954	0.963	0.954
AdaBoost	99.0654	0.996	0.991	0.991	0.990
Random Forest	97.6636	0.997	0.975	0.977	0.974
Rotation Forest	97.1963	0.993	0.969	0.972	0.968
Logistic Regression	99.5327	0.999	0.996	0.995	0.995

TABLE III
RESULTS OF ENSEMBLE LEARNING METHODS BY UTILIZING THE LETTER RECOGNITION DATASET

Letter Recognition					
Methods	Acc (%)	AUC	Precision	Recall	F-Measure
Bagging	94.8750	0.944	0.949	0.949	0.932
AdaBoost	93.7500	0.744	0.938	0.938	0.968
Random Forest	95.0000	0.987	0.953	0.950	0.934
Rotation Forest	95.1875	0.930	0.952	0.952	0.938
Logistic Regression	94.0625	0.813	0.925	0.941	0.926

TABLE IV
RESULTS OF ENSEMBLE LEARNING METHODS BY UTILIZING THE SHUTTLE DATASET

Shuttle					
Methods	Acc (%)	AUC	Precision	Recall	F-Measure
Bagging	99.9919	0.999	1.000	1.000	1.000
AdaBoost	99.8330	1.000	0.998	0.998	0.998
Random Forest	99.9939	1.000	1.000	1.000	1.000
Rotation Forest	99.9817	1.000	1.000	1.000	1.000
Logistic Regression	99.6497	0.988	0.997	0.996	0.996

TABLE V
RESULTS OF ENSEMBLE LEARNING METHODS BY UTILIZING THE FOREST COVER DATASET

Forest Cover					
Methods	Acc (%)	AUC	Precision	Recall	F-Measure
Bagging	99.9790	1.000	1.000	1.000	1.000
AdaBoost	99.8133	0.999	0.998	0.998	0.998
Random Forest	99.9857	1.000	1.000	1.000	1.000
Rotation Forest	99.9773	1.000	1.000	1.000	1.000
Logistic Regression	99.8941	1.000	0.999	0.999	0.999

TABLE VI
RESULTS OF ENSEMBLE LEARNING METHODS BY UTILIZING THE
VERTEBRAL DATASET

Vertebral					
Methods	Acc (%)	AUC	Precision	Recall	F-Measure
Bagging	91.6667	0.887	0.912	0.917	0.903
AdaBoost	87.0833	0.879	0.844	0.871	0.852
Random Forest	91.6667	0.889	0.910	0.917	0.905
Rotation Forest	91.6667	0.929	0.909	0.917	0.909
Logistic Regression	92.5000	0.930	0.919	0.925	0.920

In general, bagging has more successive consequences than boosting, whereas, the random forest provides more effective outputs than rotation forest in most of the datasets. On the other hand, logistic regression has also provided satisfactory results to some extent, which is illustrated in Tables II and VI. In Table VII, a stacking-based ensemble learning method has been applied, in which the model is trained with the combined prediction preceding model. The logistic regression has been set as a Meta classifier and experienced the diverse datasets with numerous methods like rotation forest, random forest, boosting and bagging in the given order.

The letter recognition, forest cover and vertebral datasets have significant outputs concerning the accuracy, AUC, precision, recall and F-measure parameters in Table VII; however, glass and shuttle datasets show similar outcomes for Tables II and IV.

Table VII demonstrates the comparison of all datasets results, with respect to our proposed stacking-based meta-ensemble learning method. As it is clearly shown in Table VII, a Meta-ensemble classifier, stacking with four base learners (namely, Rotation Forest, Random Forest, Bagging and Boosting) and one Meta-learner (namely, Logistic Regression) provide highly accurate outcomes as compare to others.

TABLE VII
OUR STACKING-BASED ENSEMBLE LEARNING METHOD

Proposed Stacking Meta Classifier Logistic Regression						
Classifier	Acc (%)	Impr. (%)	AUC	Precision	Recall	F-Measure
Glass	*99.5327	0.0000	0.997	*0.996	*0.995	*0.995
Letter Recognition	97.4375	2.2500	*0.987	0.973	0.974	0.974
Shuttle	*99.9939	0.0000	*1.000	*1.000	*1.000	*1.000
Forest Cover	99.9860	0.0003	*1.000	*1.000	*1.000	*1.000
Vertebral	93.3333	0.8333	0.903	0.931	0.933	0.926

- * Indicates the similar performance results concerning base learner.
- High Acc, AUC, Precision, Recall and F-measure is shown in Bold, while the greyed shows insufficient results.
- Impr. represents improvement according to best results of Tables II-VI.

Moreover, in Table VII, it is analyzed that when stacking based ensemble learning method combines with logistic regression, it provides more accurate outcomes than logistic

regression; whereas, logistic regression does not provide better outcomes when applied individually.

VI. CONCLUSION AND FUTURE WORK

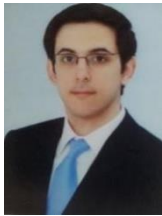
This research work has proposed an approach of ensemble classifiers in multiple datasets efficiently for outlier detection. However, these different methods such as Random forest, Rotation forest, Bagging and Boosting (base learners) and Meta-learner logistic regression under stacking classifiers occupy more space and consume more time for computations. This method enables us to provide more productive and effective outputs by using the advantages of these algorithms. We believe that this scenario is suitable for both research and commercial applications. The performance of the classifier models can be different depending on the datasets that will be chosen. In the future, other hybridization of ensemble learning methods will be employed for performance improvement.

REFERENCES

- [1] Ö. G. Alma, S. Kurt and U. Aybars, "Genetic algorithms for outlier detection in multiple regression with different information criteria," vol. 9655, 2011.
- [2] C. Pardo, J. F. Diez-Pastor, C. Garcia-Osorio and J. J. Rodríguez, "Rotation Forests for regression," *Appl. Math. Comput.*, vol. 219, no. 19, pp. 9914–9924, 2013.
- [3] L. Chen, S. Gao and X. Cao, "Research on real-time outlier detection over big data streams," *Int. J. Comput. Appl.*, vol. 7074, pp. 1–9, 2017.
- [4] N. Simidjievski, "Predicting long-term population dynamics with bagging and boosting of process-based models," vol. 42, pp. 8484–8496, 2015.
- [5] C. Zhang and J. Zhang, "RotBoost : A technique for combining Rotation Forest and AdaBoost," vol. 29, pp. 1524–1536, 2008.
- [6] A. Bagnall, M. Flynn, J. Large, J. Line, A. Bostrom and G. Cawley, "Is rotation forest the best classifier for problems with continuous features?," 2018.
- [7] E. Taşçı, "A Meta-Ensemble Classifier Approach: Random Rotation Forest," *Balk. J. Electr. Comput. Eng.*, vol. 7, no. 2, pp. 182–187, 2019.
- [8] P. Du, A. Samat, B. Waske, S. Liu and Z. Li, "Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 38–53, 2015.
- [9] S. Agarwal and C. R. Chowdary, "A-Stacking and A-Bagging: Adaptive versions of ensemble learning algorithms for spoof fingerprint detection," *Expert Syst. Appl.*, vol. 146, p. 113160, 2020.
- [10] J. Zhou Feng, Y. Wang, J. Peng, M. wei Sun, J. Zeng and H. Jiang, "Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries," *J. Crit. Care*, vol. 54, pp. 110–116, 2019.
- [11] Eibe Frank, Mark A. Hall and Ian H. Witten (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, Morgan Kaufmann, Fourth Edition, 2016.
- [12] T. A. Engel, A. S. Charão, M. Kirsch-Pinheiro and L. A. Steffanel, "Performance improvement of data mining in weka through GPU acceleration," *Procedia Comput. Sci.*, vol. 32, pp. 93–100, 2014.
- [13] Shebuti Rayana (2016). *ODDS Library* [<http://odds.cs.stonybrook.edu>]. Stony Brook, NY: Stony Brook University, Department of Computer Science.
- [14] Y. Zhou and G. Qiu, "Random forest for label ranking," *Expert Syst. Appl.*, vol. 112, pp. 99–109, 2018.
- [15] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
- [16] L. A. Bull, K. Worden, R. Fuentes, G. Manson, E. J. Cross, and N. Dervilis, "Outlier ensembles: A robust method for damage detection and unsupervised feature extraction from high-dimensional data," *J. Sound Vib.*, vol. 453, pp. 126–150, 2019.



ABDUL AHAD ABRO was born in 1988. He received the BS degree from the University of Sindh, MSc degree from Mohammad Ali Jinnah University and enrolled in a PhD degree at the Ege University, Computer Engineering Department, Izmir, Turkey. From 2014 to 2015, he was an ad-hoc lecturer in Sindh Madressatul Islam University. He also serves as a reviewer for numerous indexed journals & conferences. His research interests are machine learning, data mining and ensemble learning.



ERDAL TAŞCI was born in 1989. He received the BS, MSc and PhD degrees from the Ege University, Computer Engineering Department, Izmir, Turkey, in 2011, 2013 and 2018, respectively. He has been working as a teaching assistant at the Department of Computer Engineering of Ege University, Izmir, Turkey. He has been reviewing for various journals including IEEE Transactions on Systems, Man and Cybernetics: Systems, Computer Methods and Programs in Biomedicine and IET Image Processing. His research interests are pattern recognition, machine learning, image processing and data mining.



AYBARS UĞUR is a full-time professor in the Department of Computer Engineering at Ege University, Izmir, Turkey. He received his BS, MSc and Ph.D. degrees in computer engineering from Ege University, Izmir, Turkey, in 1993, 1996, 2001, respectively. His research interests are artificial intelligence, machine learning, deep learning, swarm intelligence, computer vision, optimization, intelligent systems and computer graphics.