OAB - An Open Anomaly Benchmark Framework for Unsupervised and Semisupervised Anomaly Detection on Image and Tabular Data Sets

Supplementary Material

1 Data Sets

1.1 Tabular Data Sets

In addition to the material in the paper, Table 1 contains information about which label is considered normal and which label is anomalous.

Name	n_{normal}	n_{anomaly}	Features	Normal labels	Anomaly labels
spambase	2528	1679	57	[0]	[1]
NASA	877	315	21	[False]	[True]
wilt	4562	257	5	['n']	['w']
annthyroid	6528	534	6	[0]	[1]
page-blocks	4883	510	10	[1]	[2, 3, 4, 5]
ionosphere	225	125	33	[0]	[1]
pulsar_star	16259	1639	8	[0]	[1]
boston	255	251	13		

Table 1: Tabular data sets and their characteristics. n_{normal} is the number of normal data points, and n_{anomaly} the number of anomalies in the data set. Features indicates the number of features (or dimensionality) of each data point. Finally, normal labels and anomaly labels indicate how the data sets were transformed into anomaly data sets. Note that as boston is a regression data set, the last two columns do not apply. The process for transforming a regression data set into an anomaly data set is described in the paper.

These data sets are available at the following url:

- Spambase: `https://archive.ics.uci.edu/ml/datasets/`

- NASA: https://www.openml.org/d/1067
- Wilt: https://archive.ics.uci.edu/ml/datasets/wilt
- Annthyroid: http://odds.cs.stonybrook.edu/annthyroid-dataset/
- Page-Blocks: https://archive.ics.uci.edu/ml/datasets/
- Ionosphere: http://odds.cs.stonybrook.edu/ionosphere-dataset/
- Boston: https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html

Apart from the data sets shown in the Table and mentioned in the paper, the following datasets are also natively supported by `oab`:

- Pulsar_star: https://archive.ics.uci.edu/ml/datasets/HTRU2
- Forest_cover: https://archive.ics.uci.edu/ml/datasets/covertype
- Wine: http://odds.cs.stonybrook.edu/wine-dataset/
- Http: http://odds.cs.stonybrook.edu/http-kddcup99-dataset/
- Smtp: http://odds.cs.stonybrook.edu/smtp-kddcup99-dataset/
- Cardio: http://odds.cs.stonybrook.edu/cardiotocogrpahy-dataset/
- Thyroid: http://odds.cs.stonybrook.edu/thyroid-disease-dataset/
- Musk: http://odds.cs.stonybrook.edu/musk-dataset/
- Pima: http://odds.cs.stonybrook.edu/pima-indians-diabetes-dataset/
- Shuttle: http://odds.cs.stonybrook.edu/shuttle-dataset/
- Breastw: http://odds.cs.stonybrook.edu/breast-cancer-wisconsin-original-dataset/
- Arrhytmia: http://odds.cs.stonybrook.edu/arrhythmia-dataset/
- Optdigits: http://odds.cs.stonybrook.edu/optdigits-dataset/
- Mammography: http://odds.cs.stonybrook.edu/mammography-dataset/
- Pendigits: http://odds.cs.stonybrook.edu/pendigits-dataset/
- Vertebral: http://odds.cs.stonybrook.edu/vertebral-dataset/

1.2 Image Data Sets

Table 2 contains additional information about the data sets used in the experiment. In total, the following data sets are natively provided by `oab`:

- MNIST (`http://yann.lecun.com/exdb/mnist/`), Fashion MNIST (`https://github.com/zalandoresearch/fashion-mnist`), CIFAR10 and CIFAR100 (superclasses) (`https://www.cs.toronto.edu/~kriz/cifar.html`) are imported from `tensorflow`.
- MVTec AD (`https://www.mvtec.com/company/research/datasets/mvtec-ad`) is a collection of 15 data sets, all of which can automatically be downloaded.
- Concrete Crack Images for Classification: `https://data.mendeley.com/datasets/5y9wdsg2zt/2`

Name	n_{normal}	n_{anomaly}	Features	Normal labels	Anomaly labels
transistor	273	40	(256, 256, 3)	[0]	[1]
screw	361	119	(256, 256)	[0]	[1]
pill	293	141	(256, 256, 3)	[0]	[1]
carpet	308	89	(256, 256, 3)	[0]	[1]
hazelnut	431	70	(256, 256, 3)	[0]	[1]
mnist_0	6903	63097	(28, 28)	[0]	[1 - 9]
cifar10_0	6000	54000	(32, 32, 3)	[0]	[1 - 9]

Table 2: Image data sets. The first 5 data sets are from MVTec AD. n_{normal} is the number of normal data points, and n_{anomaly} the number of anomalies in the data set. Features indicates the number of features (or dimensionality) of each data point. Finally, normal labels and anomaly labels indicate how the data sets were transformed into anomaly data sets.

2 Results

This section provides more in-depth results from the experiments section in the paper. For each of the four settings (tabular unsupervised, tabular semisupervised, image unsupervised, image semisupervised), the ROC AUC score, Adjusted Average Precision (Adjusted AP), and Area under Precision Recall Curve are printed.

2.1 Tabular Data: Unsupervised

For the experiments in the unsupervised setting on tabular data, Table 3 shows the ROC AUC scores (which can also be found in the paper), Table 4 shows the Adjusted Average Precision (AP) scores, and Table 5 shows the Area under Precision Recall Curve results.

	spambase	wilt	NASA	annthyroid	page-blocks	ionosphere	boston	Avg
kNN	0.633±0.025	*0.612±0.005*	0.653±0.034	*0.947±0.002*	*0.935±0.004*	*0.950±0.030*	0.733±0.008	*0.780*
LOF	0.588±0.025	0.573±0.006	0.652±0.032	**0.950±0.002**	0.927±0.004	0.907±0.040	0.715±0.010	0.758
IForest	**0.778±0.014**	0.414±0.027	*0.671±0.031*	0.826±0.008	0.914±0.004	0.947±0.026	**0.811±0.018**	0.766
ABOD	0.721±0.016	**0.699±0.005**	**0.698±0.029**	0.927±0.005	**0.954±0.004**	**0.970±0.022**	0.632±0.042	**0.800**
AE	*0.753±0.019*	0.334±0.008	0.593±0.040	0.688±0.011	0.915±0.005	0.917±0.035	*0.796±0.012*	0.714
AELOF	0.504±0.005	0.500±0.000	0.506±0.008	0.503±0.004	0.524±0.032	0.503±0.013	0.470±0.004	0.501
Avg	0.663	0.522	0.629	0.807	0.861	0.866	0.693	

Table 3: Tabular data in unsupervised setting, ROC AUC with standard deviations.

	spambase	wilt	NASA	annthyroid	page-blocks	ionosphere	boston	Avg
kNN	0.031±0.014	*0.008±0.001*	0.076±0.026	*0.505±0.013*	*0.498±0.020*	*0.803±0.088*	0.038±0.003	**0.280**
LOF	0.019±0.012	0.003±0.001	0.074±0.025	**0.515±0.013**	0.483±0.019	0.634±0.102	0.032±0.003	0.251
IForest	**0.122±0.012**	-0.012±0.002	**0.085±0.025**	0.209±0.017	0.342±0.014	0.696±0.083	**0.163±0.024**	0.229
ABOD	0.075±0.017	**0.025±0.001**	*0.082±0.021*	0.365±0.010	**0.561±0.022**	**0.806±0.100**	0.024±0.010	*0.277*
AE	*0.101±0.014*	-0.017±0.000	0.069±0.023	0.120±0.008	0.378±0.030	0.628±0.118	*0.093±0.010*	0.196
AELOF	0.008±0.009	0.000±0.000	0.009±0.010	0.005±0.008	0.043±0.056	0.009±0.027	0.000±0.000	0.011
Avg	0.059	0.001	0.066	0.286	0.384	0.596	0.058	

Table 4: Tabular data in unsupervised setting, Adjusted AP with standard deviations.

	spambase	wilt	NASA	annthyroid	page-blocks	ionosphere	boston	Avg
kNN	0.078±0.013	0.058±0.001	0.111±0.023	*0.529±0.013*	*0.522±0.019*	**0.810±0.085**	0.083±0.003	*0.313*
LOF	0.067±0.012	0.053±0.001	0.110±0.022	**0.538±0.012**	0.507±0.018	0.647±0.098	0.077±0.003	0.286
IForest	*0.163±0.011*	0.039±0.002	*0.121±0.023*	0.246±0.017	0.373±0.013	0.703±0.082	**0.196±0.023**	0.263
ABOD	0.118±0.016	*0.074±0.001*	0.119±0.019	0.395±0.010	**0.581±0.021**	*0.809±0.099*	0.069±0.009	0.309
AE	0.143±0.013	0.034±0.000	0.106±0.021	0.162±0.007	0.408±0.029	0.640±0.115	*0.128±0.009*	0.232
AELOF	**0.295±0.223**	**0.525±0.000**	**0.238±0.214**	0.528±0.004	0.415±0.094	0.400±0.202	0.026±0.000	**0.347**
Avg	0.144	0.130	0.134	0.400	0.468	0.668	0.096	

Table 5: Tabular data in unsupervised setting, Area under Precision Recall Curve with standard deviations.

	spambase	wilt	NASA	annthyroid	page-blocks	ionosphere	boston	Avg
OCSVM	0.658±0.009	**0.463±0.008**	*0.618±0.014*	**0.953±0.002**	**0.944±0.003**	0.902±0.021	0.738±0.023	*0.754*
IForest	**0.821±0.011**	*0.452±0.026*	**0.663±0.013**	*0.904±0.011*	0.929±0.004	**0.923±0.027**	**0.838±0.014**	**0.790**
PCA	0.810±0.007	0.327±0.040	0.587±0.015	0.792±0.013	0.929±0.004	0.903±0.024	0.812±0.017	0.737
AE	0.809±0.007	0.351±0.006	0.565±0.019	0.822±0.012	*0.934±0.004*	*0.914±0.021*	*0.819±0.012*	0.745
VAE	*0.810±0.007*	0.341±0.006	0.585±0.017	0.823±0.012	0.933±0.004	0.902±0.022	0.814±0.015	0.744
Avg	0.782	0.387	0.604	0.859	0.934	0.909	0.804	

Table 6: Tabular data in semisupervised setting, ROC AUC with standard deviations.

	spambase	wilt	NASA	annthyroid	page-blocks	ionosphere	boston	Avg
OCSVM	0.324±0.031	**-0.027±0.002**	*0.268±0.017*	**0.771±0.014**	**0.810±0.008**	0.843±0.032	0.116±0.020	0.444
IForest	**0.632±0.029**	*-0.032±0.006*	**0.327±0.020**	*0.606±0.029*	0.715±0.016	**0.854±0.060**	**0.401±0.057**	**0.501**
PCA	0.614±0.021	-0.057±0.007	0.242±0.013	0.493±0.017	0.736±0.013	0.831±0.046	0.287±0.044	0.450
AE	0.613±0.020	-0.051±0.001	0.216±0.016	0.523±0.018	*0.749±0.013*	*0.847±0.041*	*0.299±0.036*	*0.457*
VAE	*0.614±0.020*	-0.054±0.001	0.239±0.015	0.524±0.018	0.746±0.013	0.829±0.041	0.289±0.043	0.455
Avg	0.559	-0.044	0.258	0.583	0.751	0.841	0.279	

Table 7: Tabular data in semisupervised setting, Adjusted AP with standard deviations.

2.2 Tabular Data: Semisupervised

For the experiments in the semisupervised setting on tabular data, Table 6 shows the ROC AUC scores (which can also be found in the paper), Table 7 shows the Adjusted Average Precision scores, and Table 8 shows the Area under Precision Recall Curve results.

2.3 Image Data: Unsupervised

For the experiments in the unsupervised setting on image data, Table 9 shows the ROC AUC scores (which can also be found in the paper), Table 10 shows the Adjusted Average Precision (AP) scores, and Table 11 shows the Area under Precision Recall Curve results.

2.4 Image Data: Semisupervised

For the experiments in the semisupervised setting on image data, Table 12 shows the ROC AUC scores (which can also be found in the paper), Table 13 shows the Adjusted Average Precision scores, and Table 14 shows the Area under Precision Recall Curve results. Remeber that for the MVTec AD data sets, the original

	spambase	wilt	NASA	annthyroid	page-blocks	ionosphere	boston	Avg
OCSVM	0.662±0.015	**0.134±0.002**	*0.633±0.009*	**0.820±0.011**	**0.862±0.005**	0.921±0.016	0.265±0.016	*0.614*
IForest	**0.815±0.015**	*0.131±0.005*	**0.662±0.010**	*0.689±0.023*	0.788±0.012	**0.927±0.030**	**0.494±0.049**	**0.644**
PCA	0.807±0.010	0.110±0.006	0.619±0.007	0.601±0.013	0.804±0.010	0.915±0.023	0.396±0.035	0.607
AE	0.806±0.010	0.114±0.001	0.607±0.008	0.625±0.014	*0.813±0.009*	*0.923±0.020*	*0.406±0.028*	0.613
VAE	*0.807±0.010*	0.112±0.001	0.618±0.007	0.626±0.014	0.811±0.009	0.914±0.021	0.398±0.034	0.612
Avg	0.779	0.120	0.628	0.672	0.816	0.920	0.392	

Table 8: Tabular data in semisupervised setting, Area under Precision Recall Curve with standard deviations.

	mnist	cifar	transistor	screw	pill	carpet	hazelnut	Avg
CAEKNN	*0.992*	0.639	0.649	0.486	*0.686*	0.556	*0.554*	0.652
CAELOF	**0.995**	0.632	*0.657*	*0.526*	**0.790**	0.541	0.548	*0.670*
CAEABOD	0.935	*0.657*	**0.686**	**0.567**	0.676	**0.569**	**0.671**	**0.680**
CAEIForest	0.967	0.645	0.638	0.507	0.575	0.521	0.545	0.628
CAE	0.952	**0.737**	0.635	0.360	0.643	0.381	0.401	0.587
Avg	0.968	0.662	0.653	0.489	0.674	0.514	0.544	

Table 9: Image data in unsupervised setting, ROC AUC.

	mnist	cifar	transistor	screw	pill	carpet	hazelnut	Avg
CAEKNN	*0.856*	0.026	0.245	0.009	0.156	0.031	0.125	0.207
CAELOF	**0.907**	0.026	*0.300*	*0.016*	**0.238**	0.032	*0.198*	**0.245**
CAEABOD	0.662	*0.042*	**0.334**	**0.018**	*0.211*	*0.041*	**0.250**	*0.223*
CAEIForest	0.619	0.030	0.204	0.013	0.116	**0.043**	0.112	0.162
CAE	0.509	**0.064**	0.170	-0.012	0.063	-0.003	-0.009	0.112
Avg	0.711	0.038	0.251	0.009	0.157	0.029	0.135	

Table 10: Tabular data in unsupervised setting, Adjusted AP scores.

	mnist	cifar	transistor	screw	pill	carpet	hazelnut	Avg
CAEKNN	*0.863*	0.074	0.272	0.052	0.189	0.069	0.157	0.239
CAELOF	**0.912**	0.074	*0.328*	*0.060*	**0.265**	0.071	*0.235*	**0.278**
CAEABOD	0.679	*0.089*	**0.360**	**0.063**	*0.240*	**0.079**	**0.283**	*0.256*
CAEIForest	0.637	0.078	0.229	0.059	0.153	*0.078*	0.146	0.197
CAE	0.532	**0.110**	0.198	0.038	0.103	0.043	0.040	0.152
Avg	0.724	0.085	0.277	0.054	0.190	0.068	0.172	

Table 11: Tabular data in unsupervised setting, Area under Precision Recall Curve scores.

train-test split is used in the semisupervised setting. As this fixed split is used, there is only one sampling operation, and therefore the standard deviation is 0.

	mnist	cifar10	transistor	screw	pill	carpet	hazelnut	Avg
CAEOCSVM	**0.991±0.001**	0.639±0.009	**0.721±0.000**	**0.720±0.000**	0.502±0.000	**0.614±0.000**	0.632±0.000	**0.688**
CAEIForest	*0.980±0.003*	*0.641±0.014*	*0.718±0.000*	*0.269±0.000*	**0.524±0.000**	*0.519±0.000*	*0.661±0.000*	*0.616*
CAE	0.943±0.005	**0.714±0.008**	0.675±0.000	0.001±0.000	*0.509±0.000*	0.492±0.000	**0.841±0.000**	0.597
Avg	0.971	0.665	0.705	0.330	0.512	0.542	0.711	

Table 12: Image data in semisupervised setting, ROC AUC with standard deviations. In text: Note that mvtec's only have one sampling operation in semisupervised case.

	mnist	cifar10	mvtec_ad_transistor	mvtec_ad_screw	mvtec_ad_pill	mvtec_ad_carpet	mvtec_ad_hazelnut	Avg
CAEOCSVM	**0.977±0.002**	0.205±0.016	**0.541±0.000**	**0.544±0.000**	*0.115±0.000*	**0.368±0.000**	0.419±0.000	**0.453**
CAEIForest	*0.951±0.008*	*0.212±0.026*	*0.511±0.000*	*-0.359±0.000*	**0.147±0.000**	-0.094±0.000	*0.469±0.000*	*0.263*
CAE	0.847±0.014	**0.339±0.016**	0.422±0.000	-0.818±0.000	-0.205±0.000	*-0.064±0.000*	**0.766±0.000**	0.184
Avg	0.925	0.252	0.491	-0.211	0.019	0.070	0.551	

Table 13: Image data in semisupervised setting, Adjusted AP with standard deviations.

	mnist	cifar10	mvtec_ad_transistor	mvtec_ad_screw	mvtec_ad_pill	mvtec_ad_carpet	mvtec_ad_hazelnut	Avg
CAEOCSVM	**0.989±0.001**	0.602±0.008	**0.721±0.000**	**0.882±0.000**	*0.861±0.000*	**0.846±0.000**	0.787±0.000	**0.813**
CAEIForest	*0.976±0.004*	*0.606±0.013*	*0.703±0.000*	*0.648±0.000*	**0.866±0.000**	0.730±0.000	*0.805±0.000*	*0.762*
CAE	0.923±0.007	**0.669±0.008**	0.648±0.000	0.531±0.000	0.810±0.000	*0.738±0.000*	**0.914±0.000**	0.748
Avg	0.963	0.625	0.691	0.687	0.846	0.771	0.835	

Table 14: Image data in semisupervised setting, Area under Precision Recall Curve with standard deviations.