Supplementary Material of “Online Learning for Noisy Labeled Streams”

Jinjie Qiu, Shengda Zhuo, Philip S.Yu,Changdong Wang, and Shuqiang Huang

Outline. This document supplements the manuscript in the following aspects. Section 1 introduces compared methods, and Section 2 presents the complete experimental results that supplement.

1 Compared Methods

This section presents complete descriptions of methods which used for comparison analysis.

Compared Methods. We compared OLNLS with 13 related state-of-the-art online learning (OL), offline noisy labels learning (OFNL), and online noisy labels learning (ONL^2).

- FOBOS (Singer and Duchi 2009) (OL): is an online learning algorithm that integrates gradient descent with regularization, designed to update models incrementally and promote sparsity by penalizing non-essential features.
- RDA (Singer and Duchi 2009) (OL): offers a base framework for direct training of online learners on observed features, utilizing a projected subgradient approach to enforce sparsity and eliminate features with negligible coefficients.
- FTRL (Singer and Duchi 2009) (OL): is an online learning algorithm that uses regularization techniques to accumulate all past gradients for parameter updates, which balances between aggressive learning and maintaining stability, especially in sparse data environments.
- FTRL-P (FTRL-Promixal) (Singer and Duchi 2009) (OL): is an online learning algorithm that extends FTRL by incorporating proximal regularization, improving convergence on sparse features and providing robustness against feature correlation, particularly in high-dimensional spaces.
- OVFM (He et al. 2021) (OL): manages mixed data streams by using Gaussian cointegration modeling to link observed and distributed data, enhancing model convergence despite mixed data challenges.
- OLI^{2}DS (You et al. 2023) (OL): optimizing feature selection via empirical risk minimization and using dynamic cost strategies for class imbalance to improve model efficacy.
- OSLMF (Wu et al. 2023) (OL): applying Gaussian Copula for imputation and Density Peak Clustering for label reconstruction, thus enriching the model with extensive label information.
- Co-teaching+ (Yu et al. 2019) (OFNL): addresses noisy label learning in offline settings, using a dual-model approach to pick samples with small loss and discrepant predictions for mutual updates, focusing on reliable samples to mitigate label noise impact.
- PHuber (Menon et al. 2019) (OFNL): employs robustness regularization in offline label noise learning to prevent overfitting to noise by modifying gradients, minimizing noisy label interference in the model.
- Bey.Ima. (BeyondImages) (Zhu, Wang, and Liu 2022) (OFNL): tackles label noise in offline learning by correcting loss through an information-theoretic method that extracts key features from low-dimensional representations and estimates a noise transition matrix for model compensation, diminishing noisy label interference.
- AdaStream (Zhang et al. 2022) (ONL^2): innovates in data stream mining under conditions of incomplete and noisy labels, using a landmark assumption to estimate the noise transition matrix and leveraging model reuse ensembles to bolster method stability.
- PA (Crammer et al. 2006) (ONL^2): is an online learning algorithm using first-order linear learning with a threshold for sample training to lessen sensitivity to noisy labels and prevent overfitting to noise distributions.

2 Complete Experiments for noise-label data streams

This section presents additional experimental results that were precluded of the main paper because of page limitation. These results include:

1. The complete CAR results yielded from all 13 datasets, which are shown in Table 1, Table 2, and Table 3. The results are collected from experiments in the setting of symmetric noisy labeled, asymmetric noise labeled, and flip noisy labeled data streams, compared with online learning algorithms.
2. The complete CAR trends yielded from all 13 datasets on symmetric noisy labeled, asymmetric noise labeled, and flip noisy labeled data streams, shown in Figure 1, Figure 2, and Figure 3. The results are collected from experiments in the setting of symmetric noisy labeled, asymmetric noise labeled, and flip noisy labeled data streams, compared with online learning algorithms.

3. The complete CAR results yielded from all 13 datasets, which are shown in Table 4, Table 5, and Table 6. The results are collected from experiments in the setting of symmetric noisy labeled, asymmetric noise labeled, and flip noisy labeled data streams, compared with noisy labels learning algorithms.
4. The complete CAR trends yielded from all 13 datasets on symmetric noisy labeled, asymmetric noise labeled, and flip noisy labeled data streams, shown in Figure 4, Figure 5, and Figure 6. The results are collected from experiments in the setting of symmetric noisy labeled, asymmetric noise labeled, and flipped noisy labeled data streams, compared with noisy labels learning algorithms.

References

Crammer, K.; Dekel, O.; Keshet, J.; Shalev-Shwartz, S.; and Singer, Y. 2006. Online passive aggressive algorithms.

He, Y.; Dong, J.; Hou, B.-J.; Wang, Y.; and Wang, F. 2021. Online Learning in Variable Feature Spaces with Mixed Data. In *ICDM*, 181–190. IEEE.

Menon, A. K.; Rawat, A. S.; Reddi, S. J.; and Kumar, S. 2019. Can gradient clipping mitigate label noise? In *ICLR*.

Singer, Y.; and Duchi, J. C. 2009. Efficient learning using forward-backward splitting. *In NeurIPS*, 22.

Wu, D.; Zhuo, S.; Wang, Y.; Chen, Z.; and He, Y. 2023. Online semi-supervised learning with mix-typed streaming features. In *AAAI*, volume 37, 4720–4728.

You, D.; Xiao, J.; Wang, Y.; Yan, H.; Wu, D.; Chen, Z.; Shen, L.; and Wu, X. 2023. Online Learning From Incomplete and Imbalanced Data Streams. *IEEE Transactions on Knowledge and Data Engineering*.

Yu, X.; Han, B.; Yao, J.; Niu, G.; Tsang, I.; and Sugiyama, M. 2019. How does disagreement help generalization against label corruption? In *ICML*, 7164–7173. PMLR.

Zhang, Z.-Y.; Qian, Y.-Y.; Zhang, Y.-J.; Jiang, Y.; and Zhou, Z.-H. 2022. Adaptive Learning for Weakly Labeled Streams. In *KDD*, 2556–2564.

Zhu, Z.; Wang, J.; and Liu, Y. 2022. Beyond images: Label noise transition matrix estimation for tasks with lower-quality features. In *ICML*, 27633–27653. PMLR.

Table 1: The comparison results on cumulative accuracy rates. We repeated the experiment 5 times for each dataset, averaged the cumulative accuracy rate (CAR), and calculated the standard variance of the 5 times values. Experimental results (CAR ± Standard Variance) for 13 datasets in the case of Symmetric Noisy labeled data. ● indicates the cases that our method loses the comparison. * showing the total number of wins and losses for OLNLS. OOT: Out Of Time (24 hours).

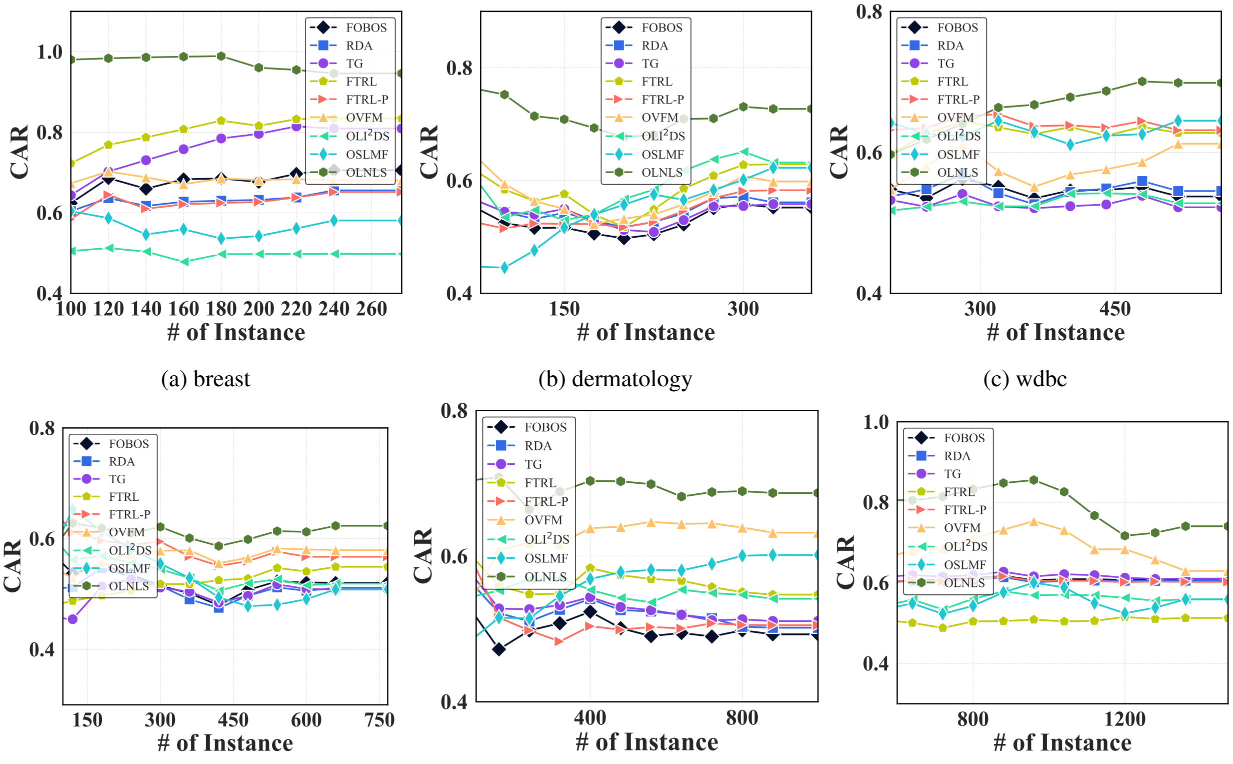
	Symetric Noise $\rho_{-1} = \rho_{+1} = 0.4$								
Datasets	**FOBOS**	**RDA**	**TG**	**FTRL**	**FTRL-P**	**OSLMF**	**OVFM**	**OLI^{2}DS**	**OLNLS**
breast	.705 ± .023	.634 ± .023	.801 ± .036	.803 ± .027	.638 ± .011	.596 ± .015	.653 ± .041	.522 ± .043	**.892 ± .026**
dermatology	.531 ± .016	.559 ± .014	.547 ± .017	.684 ± .025	.623 ± .026	.604 ± .054	.603 ± .012	.649 ± .020	**.721 ± .010**
wdbc	.537 ± .012	.542 ± .013	.549 ± .015	.573 ± .033	.607 ± .029	.598 ± .031	.579 ± .029	.544 ± .009	**.671 ± .023**
diabetes	.528 ± .013	.532 ± .014	.530 ± .022	.528 ± .025	.533 ± .019	.580 ± .037	.571 ± .023	.520 ± .006	**.598 ± .022**
german	.508 ± .017	.515 ± .013	.513 ± .009	.570 ± .017	.534 ± .015	.578 ± .017	.607 ± .022	.534 ± .024	**.694 ± .006**
contraceptive	.600 ± .002	.596 ± .011	.597 ± .008	.530 ± .017	.601 ± .003	.609 ± .033	.613 ± .020	.545 ± .016	**.787 ± .023**
splice	.603 ± .002	.610 ± .006	.604 ± .005	.694 ± .009	.599 ± .003	.500 ± .030	.576 ± .035	.554 ± .008	**.778 ± .001**
mushroom	.577 ± .005	.653 ± .037	.621 ± .006	.794 ± .017	.618 ± .004	.860 ± .026●	.763 ± .011	.674 ± .018	.833 ± .020
marketing	.580 ± .003	.608 ± .011	.613 ± .009	.767 ± .008	.592 ± .003	.717 ± .038	.696 ± .006	.565 ± .010	**.818 ± .005**
hapt	.593 ± .001	.687 ± .009	.623 ± .006	.671 ± .007	.600 ± .009	.575 ± .019	.596 ± .005	.555 ± .004	**.773 ± .008**
ring	.526 ± .008	.529 ± .008	.525 ± .008	.530 ± .006	.528 ± .008	.599 ± .012	.566 ± .016	.525 ± .006	**.656 ± .005**
magic04	.598 ± .001	.595 ± .001	.599 ± .002	.718 ± .005	.599 ± .003	.777 ± .071	.837 ± .013	.553 ± .002	**.859 ± .015**
a8a	.490 ± .003	.671 ± .008	.568 ± .006	.671 ± .004	.576 ± .003	.664 ± .006	.617 ± .005	.543 ± .005	**.696 ± .019**
loss/win	0/13	0/13	0/13	0/13	0/13	1/12	0/13	0/13	**1/103***
p-value	.0001	.0001	.0001	.0001	.0001	.0004	.0001	.0001	−−
F-rank	2.6538	4.3462	4.0000	6.1538	4.6923	5.8462	5.9231	2.4615	8.9231

Table 2: The comparison results on cumulative accuracy rates. We repeated the experiment 5 times for each dataset, averaged the cumulative accuracy rate (CAR), and calculated the standard variance of the 5 times values. Experimental results (CAR ± Standard Variance) for 13 datasets in the case of ASymmetric Noisy labeled data. ● indicates the cases that our method loses the comparison. * showing the total number of wins and losses for OLNLS. OOT: Out Of Time (24 hours).

	ASymetric Noise $\rho_{-1} = 0.3, \rho_{+1} = 0.5$								
Datasets	**FOBOS**	**RDA**	**TG**	**FTRL**	**FTRL-P**	**OSLMF**	**OVFM**	**OLI^{2}DS**	OLNLS
breast	.752 ± .019	.662 ± .007	.800 ± .017	.781 ± .014	.678 ± .018	.591 ± .037	.656 ± .013	.533 ± .033	**.812 ± .020**
dermatology	.588 ± .010	.598 ± .008	.588 ± .015	.717 ± .031	.659 ± .031	.610 ± .027	.673 ± .023	.657 ± .024	**.749 ± .018**
wdbc	.569 ± .005	.568 ± .012	.578 ± .008	.602 ± .038	.589 ± .031	.588 ± .047	.616 ± .016	.556 ± .011	**.677 ± .027**
diabetes	.486 ± .014	.484 ± .016	.493 ± .009	.483 ± .015	.493 ± .011	.510 ± .035	.492 ± .016	.510 ± .015●	.510 ± .021
german	.491 ± .018	.495 ± .016	.497 ± .014	.470 ± .007	.503 ± .013	.476 ± .033	.487 ± .039	.536 ± .019	**.556 ± .074**
contraceptive	.613 ± .009	.618 ± .010	.618 ± .011	.579 ± .011	.614 ± .005	.622 ± .047	.645 ± .038	.569 ± .017	**.730 ± .044**
splice	.655 ± .002	.671 ± .005	.660 ± .005	.684 ± .009	.650 ± .002	.607 ± .025	.639 ± .021	.547 ± .010	**.759 ± .000**
mushroom	.609 ± .004	.644 ± .014	.643 ± .005	.732 ± .013	.632 ± .004	.736 ± .071	.743 ± .006	.679 ± .016	**.771 ± .008**
marketing	.683 ± .003	.745 ± .002	.720 ± .002	.820 ± .001	.673 ± .003	.767 ± .017	.783 ± .006	.574 ± .017	**.826 ± .002**
hapt	.655 ± .002	.758 ± .007	.696 ± .003	.763 ± .004	.666 ± .004	.653 ± .009	.653 ± .017	.561 ± .005	**.818 ± .007**
ring	.504 ± .005	.500 ± .003	.502 ± .005	.505 ± .006	.503 ± .006	.502 ± .012	.514 ± .006	.528 ± .003	**.547 ± .018**
magic04	.628 ± .002	.626 ± .002	.627 ± .001	.732 ± .002	.630 ± .003	.631 ± .066	.796 ± .007	.566 ± .004	**.811 ± .022**
a8a	.460 ± .003	.671 ± .013	.575 ± .007	.740 ± .012	.638 ± .005	.740 ± .005	.684 ± .005	.547 ± .004	**.757 ± .000**
loss/win	0/13	0/13	0/13	0/13	0/13	0/13	0/13	1/12	**1/103***
p-value	.0001	.0001	.0001	.0001	.0001	.0013	.0001	.0013	−−
F-rank	3.3462	4.0385	4.4615	5.8846	4.5769	4.7308	5.7308	3.3077	8.9231

Table 3: The comparison results on cumulative accuracy rates. We repeated the experiment 5 times for each dataset, averaged the cumulative accuracy rate (CAR), and calculated the standard variance of the 5 times values. Experimental results (CAR ± Standard Variance) for 13 datasets in the case of Flip Noisy labeled data. ● indicates the cases that our method loses the comparison. * showing the total number of wins and losses for OLNLS. OOT: Out Of Time (24 hours).

	Filp Noise Rate = 0.4								
Datasets	**FOBOS**	**RDA**	**TG**	**FTRL**	**FTRL-P**	**OSLMF**	**OVFM**	**OLI^{2}DS**	**OLNLS**
breast	.677 ± .020	.617 ± .032	.736 ± .055	.790 ± .037	.641 ± .021	.582 ± .070	.656 ± .013	.561 ± .056	**.896 ± .052**
dermatology	.537 ± .029	.554 ± .026	.551 ± .033	.697 ± .032	.616 ± .024	.534 ± .053	.601 ± .013	.640 ± .032	**.745 ± .019**
wdbc	.541 ± .009	.539 ± .007	.537 ± .010	.564 ± .016	.579 ± .026	.574 ± .018	.576 ± .020	.543 ± .012	**.673 ± .023**
diabetes	.528 ± .010	.529 ± .009	.522 ± .015	.516 ± .007	.516 ± .012	.573 ± .053	.577 ± .012	.523 ± .004	**.579 ± .018**
german	.521 ± .016	.526 ± .016	.520 ± .022	.561 ± .018	.524 ± .010	.557 ± .039	.570 ± .028	.521 ± .016	**.677 ± .024**
contraceptive	.597 ± .004	.606 ± .008	.600 ± .009	.788 ± .025	.602 ± .004	.560 ± .044	.637 ± .028	.562 ± .020	**.790 ± .028**
splice	.600 ± .001	.605 ± .004	.604 ± .004	.690 ± .004	.596 ± .003	.526 ± .031	.573 ± .027	.551 ± .009	**.779 ± .001**
mushroom	.582 ± .007	.669 ± .017	.623 ± .010	.799 ± .006	.620 ± .008	.906 ± .009●	.765 ± .014	.677 ± .015	.827 ± .006
marketing	.578 ± .007	.626 ± .022	.619 ± .005	.771 ± .009	.592 ± .006	.736 ± .019	.694 ± .012	.567 ± .008	**.815 ± .006**
hapt	.592 ± .002	.700 ± .017	.622 ± .004	.673 ± .004	.599 ± .005	.568 ± .011	.587 ± .012	.552 ± .003	**.772 ± .014**
ring	.528 ± .004	.531 ± .004	.532 ± .004	.531 ± .004	.530 ± .005	.605 ± .008	.562 ± .005	.526 ± .005	**.660 ± .002**
magic04	.599 ± .001	.597 ± .001	.602 ± .003	.719 ± .002	.600 ± .002	.745 ± .039	.830 ± .007	.554 ± .003	**.869 ± .014**
a8a	.489 ± .003	.670 ± .017	.569 ± .003	.681 ± .008	.577 ± .001	.676 ± .008	.614 ± .004	.544 ± .003	**.711 ± .002**
loss/win	0/13	0/13	0/13	0/13	0/13	1/12	0/13	0/13	**1/103***
p-value	.0001	.0001	.0001	.0001	.0001	.0009	.0001	.0001	−−
F-rank	3.0385	4.8077	4.0000	6.6154	4.1154	4.9231	6.0000	2.5769	8.9231



(a) breast

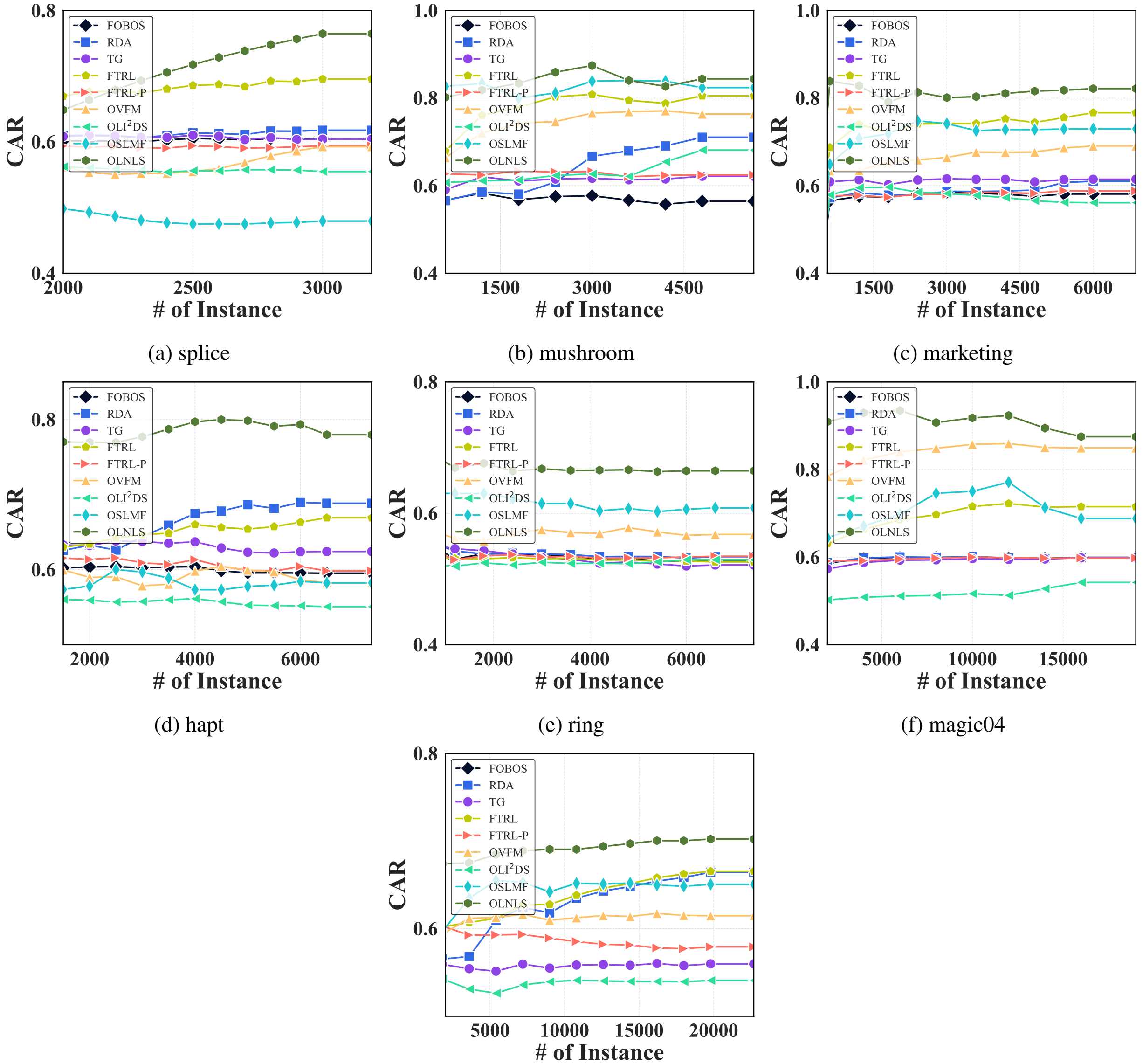
(b) dermatology

(c) wdbc

(d) diabetes

(e) german

(f) contraceptive



(a) splice

(b) mushroom

(c) marketing

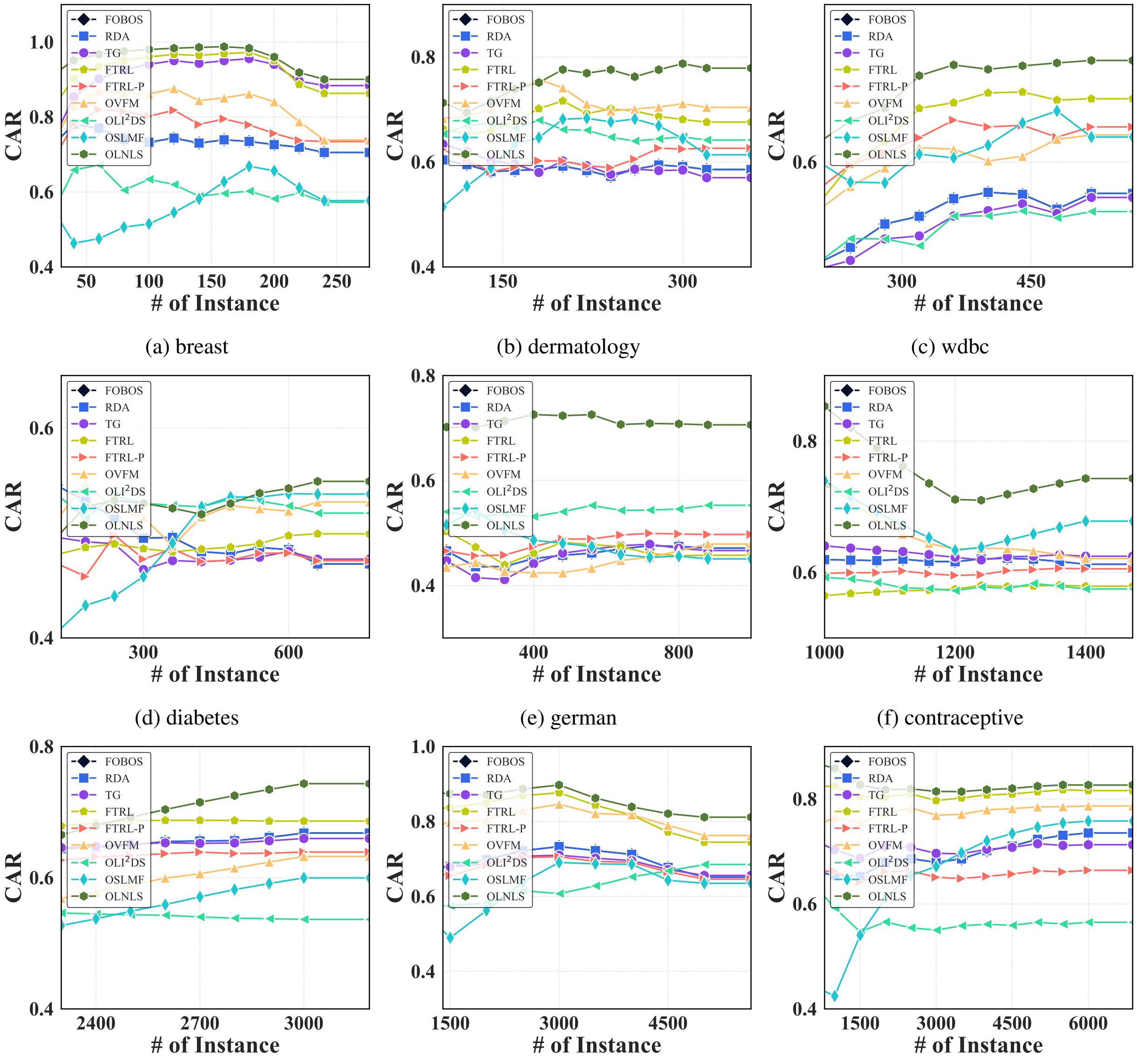
(d) hapt

(e) ring

(f) magic04

(g) a8a

Figure 1: The cumulative accuracy rate (CAR) trends of FOBOS, RDA, TG, FTRL, FTRL-P, OSLMF, OVFM, OL^{2}DS, and OLNLS in all 13 symmetric noisy labeled data streams.



(a) breast

(b) dermatology

(c) wdbc

(d) diabetes

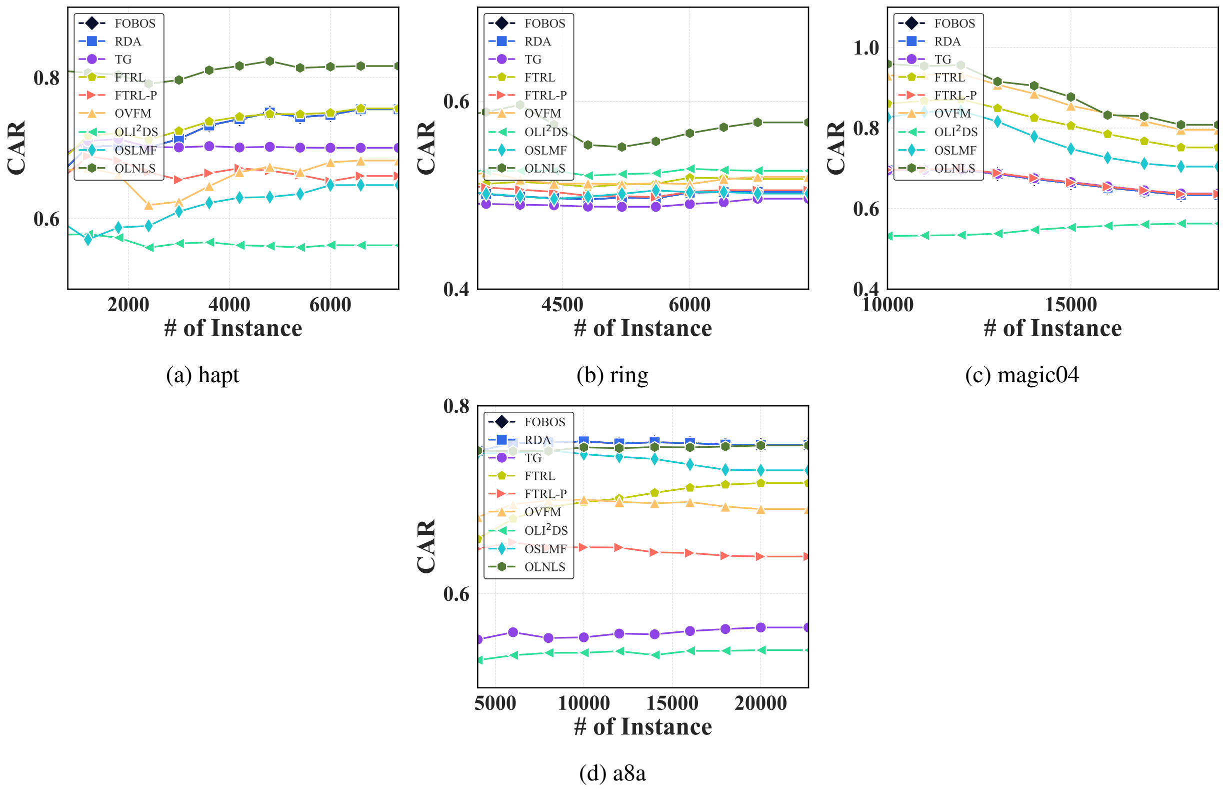
(e) german

(f) contraceptive

(g) splice

(h) mushroom

(i) marketing



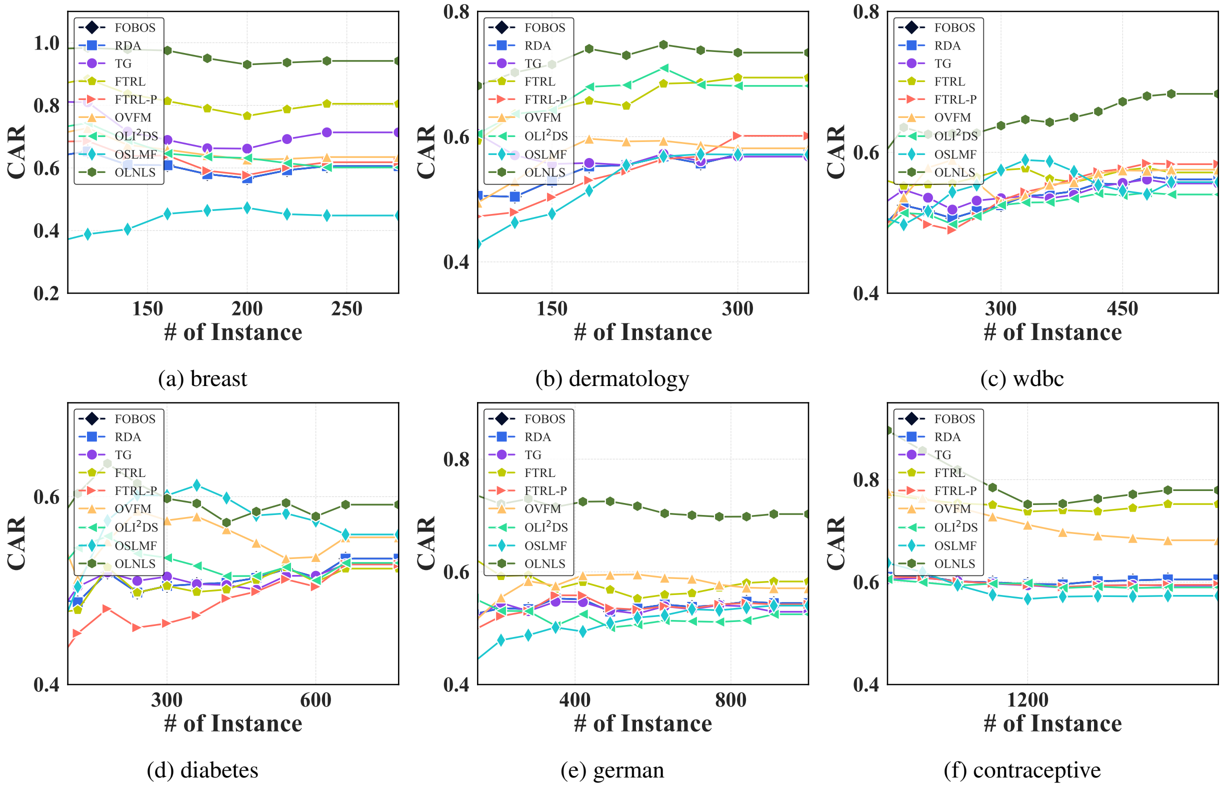
(a) hapt

(b) ring

(c) magic04

(d) a8a

Figure 2: The cumulative accuracy rate (CAR) trends of FOBOS, RDA, TG, FTRL, FTRL-P, OSLMF, OVFM, OL^{2}DS, and OLNLS in all 13 asymmetric noisy labeled data streams.



(a) breast

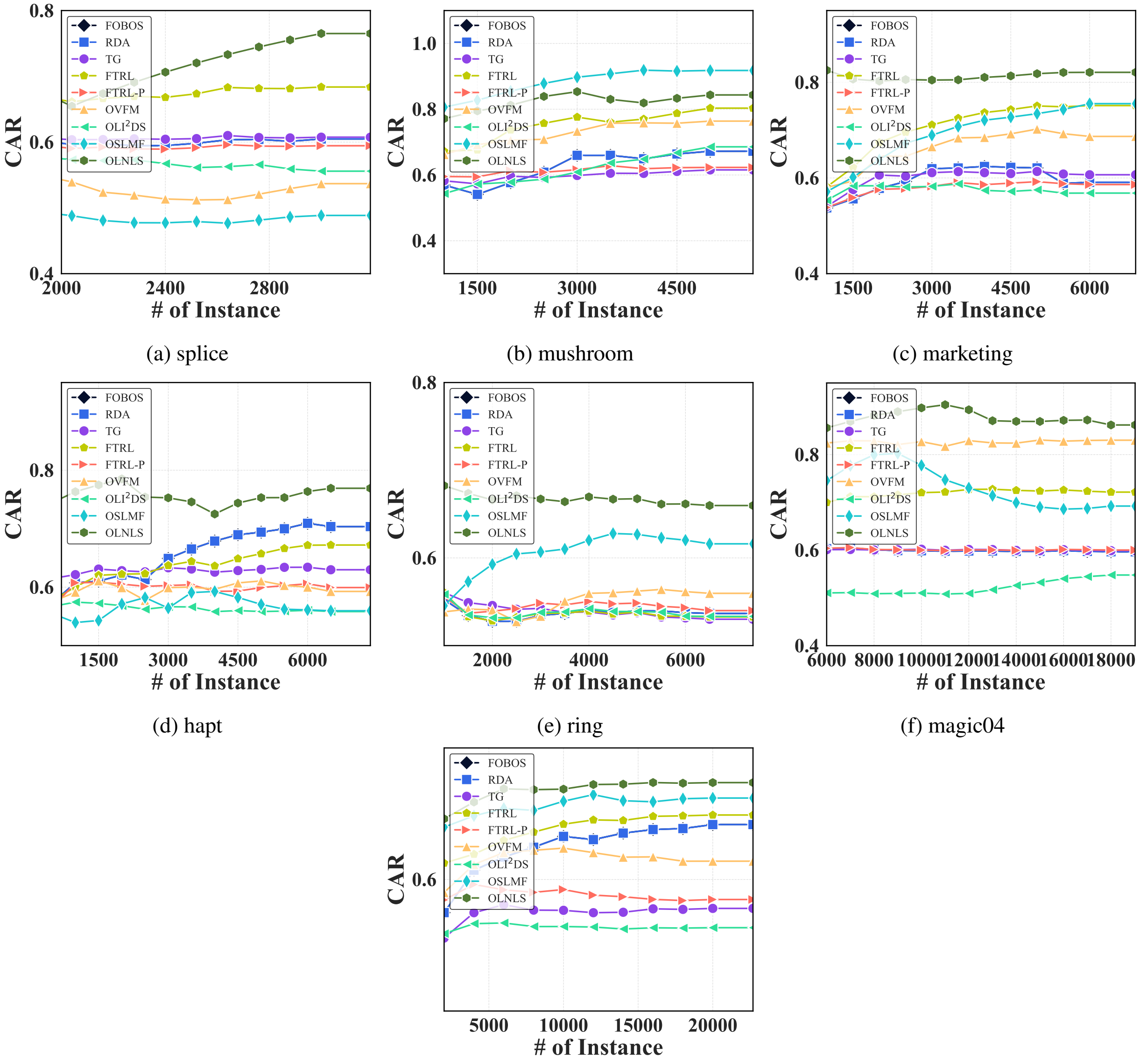
(b) dermatology

(c) wdbc

(d) diabetes

(e) german

(f) contraceptive



(a) splice

(b) mushroom

(c) marketing

(d) hapt

(e) ring

(f) magic04

(g) a8a

Figure 3: The cumulative accuracy rate (CAR) trends of FOBOS, RDA, TG, FTRL, FTRL-P, OSLMF, OVFM, OL^{2}DS, and OLNLS in all 13 flipped noisy labeled data streams.

Table 4: The comparison results on cumulative accuracy rates. We repeated the experiment 5 times for each dataset, averaged the cumulative accuracy rate (CAR), and calculated the standard variance of the 5 times values. Experimental results (CAR ± Standard Variance) for 13 datasets in the case of Symmetric Noisy labeled data. ● indicates the cases that our method loses the comparison. * showing the total number of wins and losses for OLNLS. OOT: Out Of Time (24 hours).

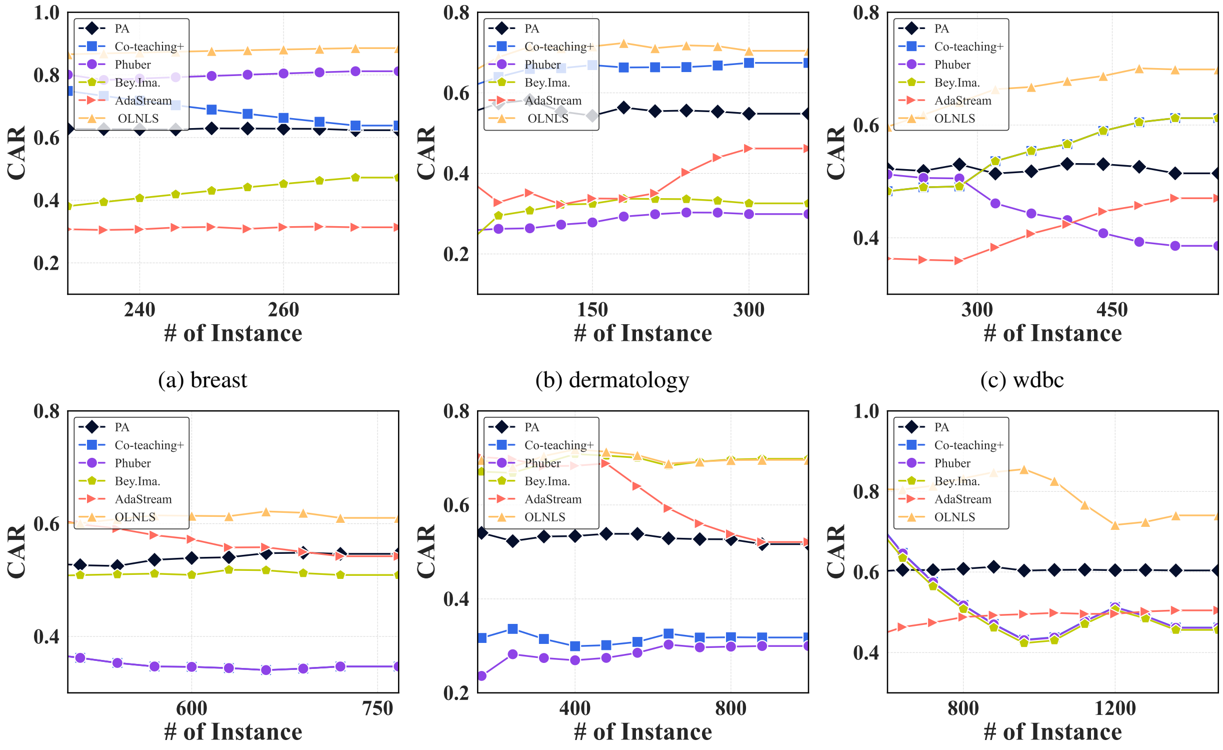
	Symetric Noise $\rho_{-1} = \rho_{+1} = 0.4$					
Datasets	**Co-teaching+**	**Phuber**	**Bey.Ima.**	**PA**	**AdaStream**	**OLNLS**
breast	.442 ± .116	.765 ± .073	.558 ± .150	.605 ± .017	.518 ± .136	**.892 ± .026**
dermatology	.460 ± .185	.615 ± .151	.466 ± .190	.541 ± .016	.499 ± .171	**.721 ± .010**
wdbc	.473 ± .126	.473 ± .126	.424 ± .102	.536 ± .013	.530 ± .080	**.671 ± .023**
diabetes	.469 ± .144	.470 ± .149	.424 ± .079	.529 ± .012	.480 ± .066	**.598 ± .022**
german	.449 ± .172	.539 ± .195	.606 ± .155	.509 ± .013	.377 ± .094	**.694 ± .006**
contraceptive	.456 ± .059	.426 ± .000	.493 ± .080	.600 ± .001	.507 ± .021	**.787 ± .023**
splice	.609 ± .187	.656 ± .207	.478 ± .219	.597 ± .004	.548 ± .071	**.778 ± .001**
mushroom	.565 ± .167	.816 ± .024	.465 ± .133	.580 ± .003	.533 ± .079	**.833 ± .020**
marketing	.691 ± .254	.794 ± .018	.699 ± .229	.562 ± .002	.517 ± .121	**.818 ± .005**
hapt	.658 ± .258	.550 ± .195	.567 ± .326	.599 ± .001	.649 ± .058	**.773 ± .008**
ring	.478 ± .062	.567 ± .089	.544 ± .099	.537 ± .005	.490 ± .052	**.656 ± .005**
magic04	.331 ± .163	.529 ± .146	.426 ± .203	.600 ± .002	.642 ± .008	**.859 ± .015**
a8a	.657 ± .098	.689 ± .009	.467 ± .232	.574 ± .002	.687 ± .004	**.696 ± .019**
loss/win	0/13	0/13	0/13	0/13	0/13	**0/65***
p-value	.0001	.0001	.0001	.0001	.0001	$--$
F-rank	2.3462	3.8077	2.3077	3.6154	2.9231	6.0000

Table 5: The comparison results on cumulative accuracy rates. We repeated the experiment 5 times for each dataset, averaged the cumulative accuracy rate (CAR), and calculated the standard variance of the 5 times values. Experimental results (CAR ± Standard Variance) for 13 datasets in the case of Symmetric Noisy labeled data. ● indicates the cases that our method loses the comparison. * showing the total number of wins and losses for OLNLS. OOT: Out Of Time (24 hours).

	ASymetric Noise $\rho_{-1} = 0.3, \rho_{+1} = 0.5$					
Datasets	**Co-teaching+**	**Phuber**	**Bey.Ima.**	**PA**	**AdaStream**	**OLNLS**
breast	.479 ± .153	.705 ± .004	.512 ± .170	.658 ± .010	.657 ± .098	**.812 ± .020**
dermatology	.598 ± .189	.615 ± .151	.539 ± .186	.607 ± .009	.614 ± .152	**.749 ± .018**
wdbc	.473 ± .126	.371 ± .000	.536 ± .134	.560 ± .005	.568 ± .073	**.677 ± .027**
diabetes	.403 ± .125	.470 ± .149	.578 ± .116●	.502 ± .017	.349 ± .000	.510 ± .021
german	.572 ± .164●	.539 ± .195	.503 ± .170	.486 ± .017	.417 ± .096	.556 ± .074
contraceptive	.514 ± .072	.485 ± .072	.507 ± .067	.614 ± .001	.513 ± .012	**.730 ± .044**
splice	.628 ± .154	.759 ± .000	.564 ± .200	.649 ± .004	.480 ± .092	**.759 ± .000**
mushroom	.529 ± .096	.670 ± .059	.505 ± .103	.604 ± .005	.513 ± .068	**.771 ± .008**
marketing	.507 ± .277	.816 ± .002	.553 ± .298	.629 ± .003	.488 ± .085	**.826 ± .002**
hapt	.700 ± .266	.746 ± .028	.439 ± .322	.666 ± .001	.762 ± .028	**.818 ± .007**
ring	.501 ± .051	.522 ± .109	.553 ± .114●	.503 ± .006	.462 ± .021	.547 ± .018
magic04	.545 ± .145	.679 ± .045	.415 ± .116	.631 ± .002	.642 ± .007	**.811 ± .022**
a8a	.727 ± .028	.706 ± .009	.460 ± .238	.641 ± .003	.716 ± .005	**.757 ± .000**
loss/win	1/12	0/13	2/11	0/13	0/13	**3/62***
p-value	.0002	.0013	.0009	.0001	.0001	$--$
F-rank	2.8462	3.8846	2.4615	3.3846	2.6923	5.7308

Table 6: The comparison results on cumulative accuracy rates. We repeated the experiment 5 times for each dataset, averaged the cumulative accuracy rate (CAR), and calculated the standard variance of the 5 times values. Experimental results (CAR ± Standard Variance) for 13 datasets in the case of Symmetric Noisy labeled data. • indicates the cases that our method loses the comparison. * showing the total number of wins and losses for OLNLS. OOT: Out Of Time (24 hours)

	Flip Noise rate = 0.4					
Datasets	**Co-teaching+**	**Phuber**	**Bey.Ima.**	**PA**	**AdaStream**	**OLNLS**
breast	.492 ± .094	.692 ± .070	.551 ± .194	.609 ± .009	.508 ± .127	**.896 ± .052**
dermatology	.530 ± .182	.464 ± .185	.574 ± .153	.563 ± .028	.538 ± .186	**.745 ± .019**
wdbc	.576 ± .102	.525 ± .126	.475 ± .126	.533 ± .015	.576 ± .063	**.673 ± .023**
diabetes	.410 ± .123	.470 ± .149	.500 ± .130	.540 ± .013	.406 ± .071	**.579 ± .018**
german	.484 ± .179	.380 ± .160	.557 ± .170	.527 ± .014	.335 ± .070	**.677 ± .024**
contraceptive	.539 ± .103	.456 ± .059	.485 ± .072	.600 ± .001	.500 ± .023	**.790 ± .028**
splice	.483 ± .202	.345 ± .207	.626 ± .196	.596 ± .005	.488 ± .086	**.779 ± .001**
mushroom	.437 ± .092	.667 ± .178	.526 ± .112	.582 ± .003	.499 ± .060	**.827 ± .006**
marketing	.502 ± .276	.804 ± .012	.673 ± .247	.564 ± .004	.447 ± .105	**.815 ± .006**
hapt	.688 ± .260	.766 ± .088	.757 ± .124	.597 ± .001	.713 ± .044	**.772 ± .014**
ring	.484 ± .029	.388 ± .000	.512 ± .054	.535 ± .005	.508 ± .042	**.660 ± .002**
magic04	.652 ± .058	.689 ± .053	.546 ± .124	.599 ± .002	.642 ± .008	**.869 ± .014**
a8a	.636 ± .198	.678 ± .012	.460 ± .191	.576 ± .003	.684 ± .004	**.711 ± .002**
loss/win	0/13	0/13	0/13	0/13	0/13	**0/65***
p-value	.0001	.0001	.0001	.0001	.0001	−−
F-rank	2.5000	3.0769	3.2308	3.5385	2.6538	6.000



(a) breast

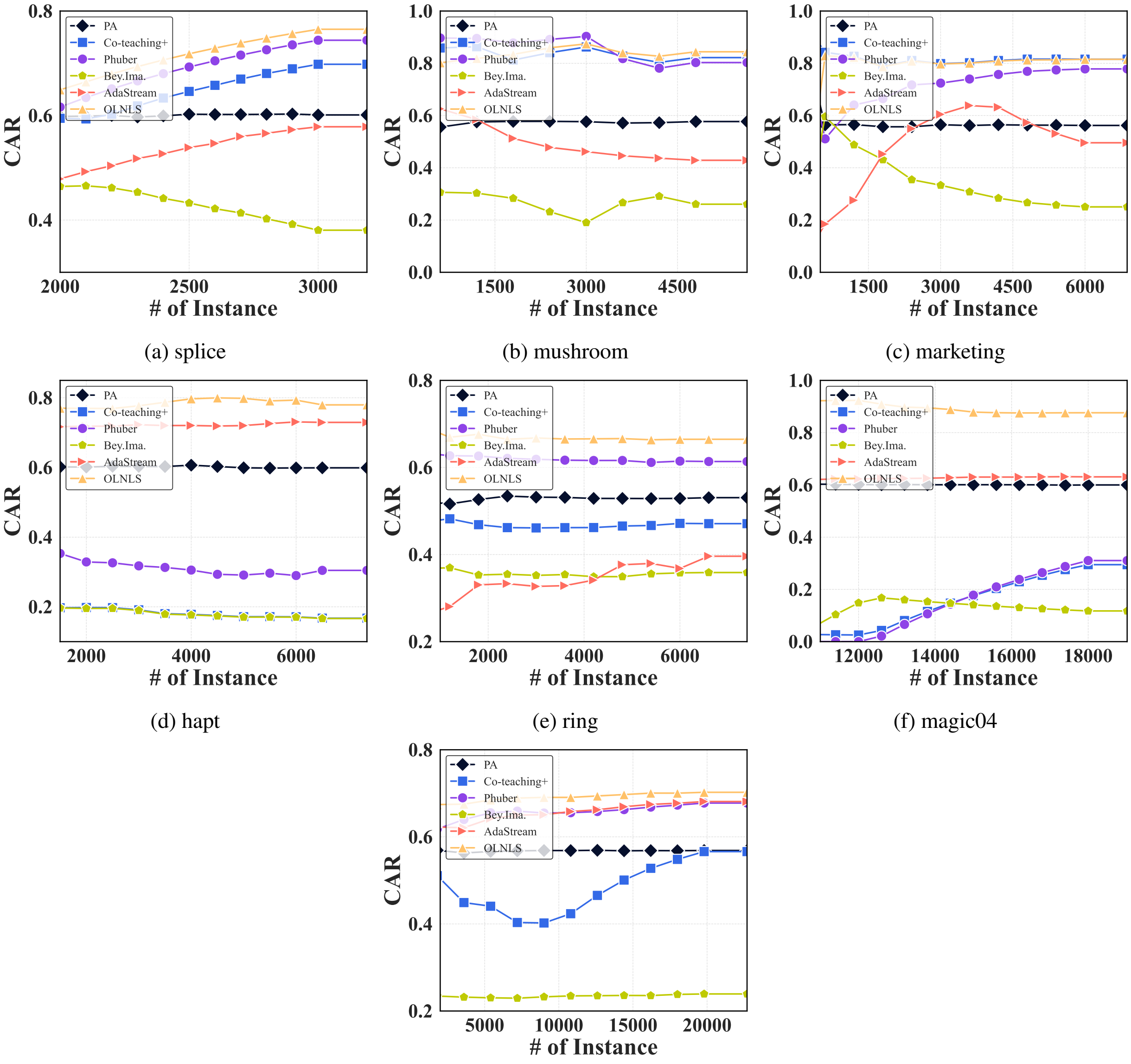
(b) dermatology

(c) wdbc

(d) diabetes

(e) german

(f) contraceptive



(a) splice

(b) mushroom

(c) marketing

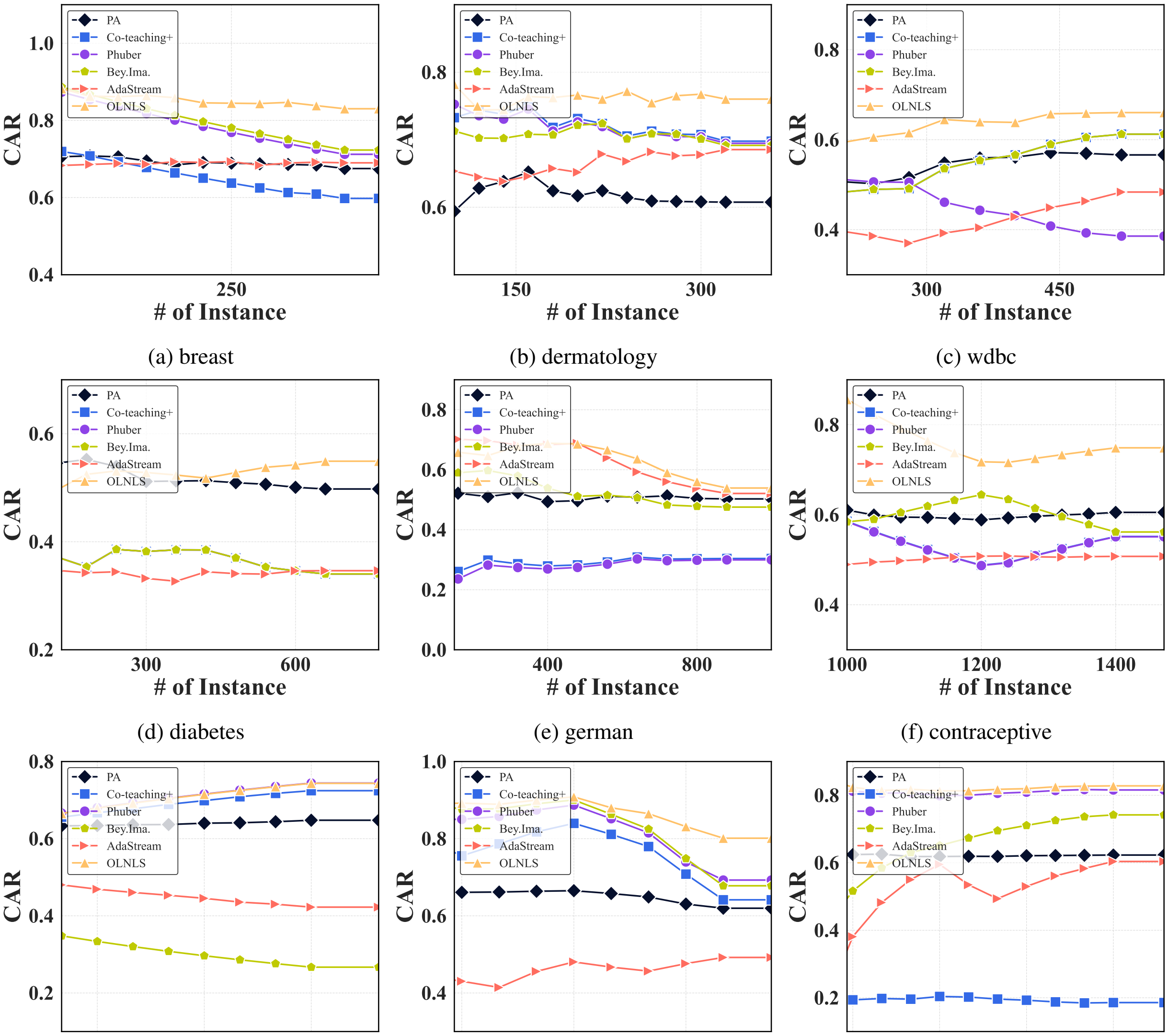
(d) hapt

(e) ring

(f) magic04

(g) a8a

Figure 4: The cumulative accuracy rate (CAR) trends of Co-teaching+, Phuber, BeyondImages, PA, AdaStream, and OLNLS in all 13 symmetric noisy labeled data streams.



(a) breast

(b) dermatology

(c) wdbc

(d) diabetes

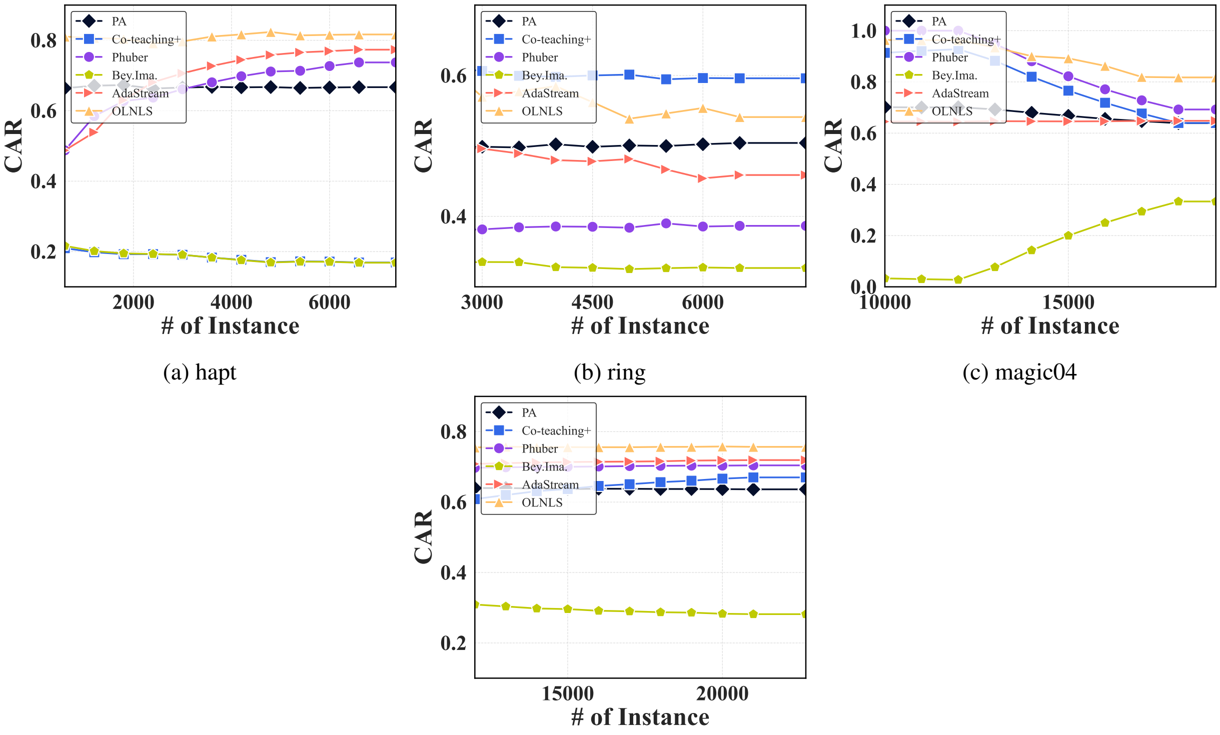
(e) german

(f) contraceptive

(g) splice

(h) mushroom

(i) marketing



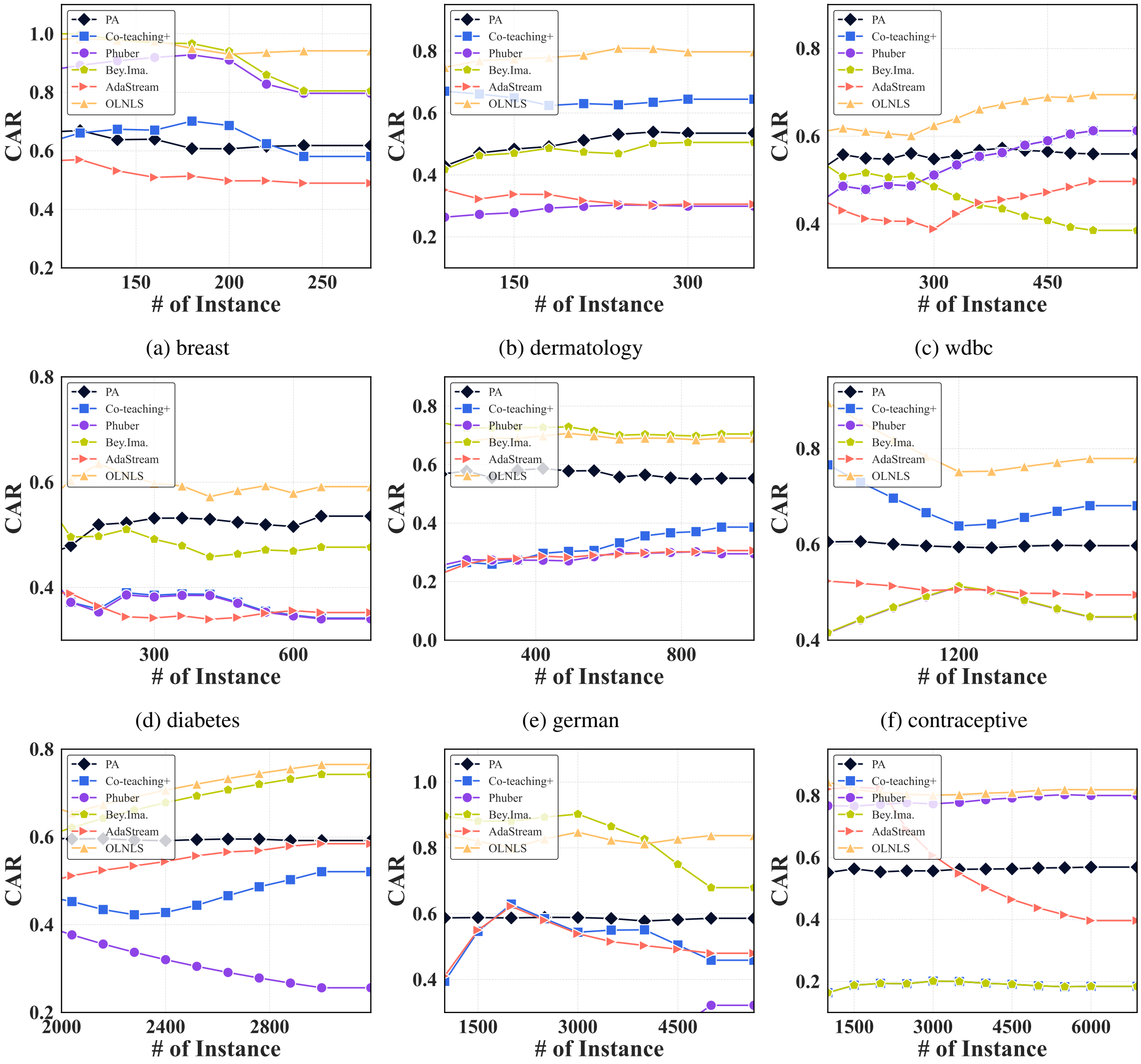
(a) hapt

(b) ring

(c) magic04

(d) a8a

Figure 5: The cumulative accuracy rate (CAR) trends of Co-teaching+, Phuber, BeyondImages, PA, AdaStream, and OLNLS in all 13 asymmetric noisy labeled data streams.



(a) breast

(b) dermatology

(c) wdbc

(d) diabetes

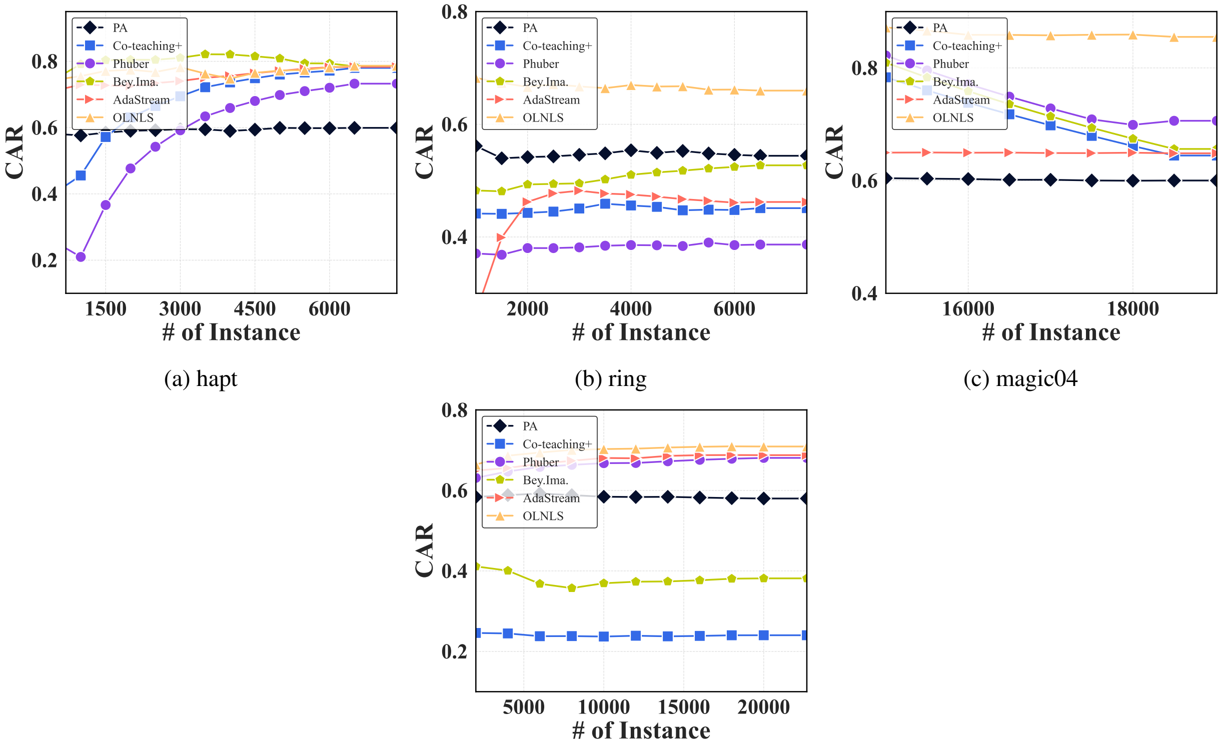
(e) german

(f) contraceptive

(g) splice

(h) mushroom

(i) marketing



(a) hapt

(b) ring

(c) magic04

(d) a8a

Figure 6: The cumulative accuracy rate (CAR) trends of Co-teaching+, Phuber, BeyondImages, PA, AdaStream, and OLNLS in all 13 flipped noisy labeled data streams.