Weighted Aging Ensemble

Supplementary materials of experimental evaluation

I. Research questions

The experiments presented in this paper are designed to answer the following questions:

Q1. How do the values of individual hyperparameters of the WAE algorithm affect its ability to classify data streams with concept drift?

Q2. What set of hyperparameter values of the proposed WAE algorithm works best for classifying drifting data streams with both full labeling and restricted label access?

Q3. How does the WAE algorithm compare to the *state-of-the-art* algorithms known from the literature, designed specifically for the task of drifting data streams classification, both in the case of full and limited label access?

II. Data streams & evaluation protocol

In order to perform the experimental evaluation of the WAE algorithm, data streams – both synthetic and based on real concepts – with various characteristics were used.

Synthetic data streams were generated using the Python stream-learn library. Three balanced streams were prepared, differing in the concept drift type, and replicated 5 times based on a random state value to stabilize the results and to enable statistical analysis. These streams were characterized by the following parameters:

- *data chunks number* – 200,
- *chunk size* – 250,
- *global label noise* – 1%,
- *concept drift type – sudden, gradual, and incremental*,
- *drifts number* – 10,
- *number of features* – 8.

The second data source was a generator that allowed to create data streams based on static concepts from known benchmark static datasets available on repositories such as UCI or KEEL. The following data sets were used to use the streams (Table I). Finally, the experiments were also carried out using real streams (Table II).

All experiments were conducted using the *stream-learn* library and based on the Test-Than-Train evaluation protocol. In the case of synthetic streams, *accuracy score* values are reported, while in experiments containing streams based on real concepts – in order to eliminate the impact of possible data imbalance – a *balanced accuracy score* was used. The statistical analysis was performed using the *student's t-test*. The experiments presented in this article can be replicated using the code available in the *GitHub* repository.

TABLE I
Static datasets used for data stream generation.

Dataset	#I	#F	#C
banknote	1372	4	2
heart	270	13	2
liver	345	6	2
monkone	556	6	2
sonar	208	60	2
soybean	47	35	2
wisconsin	683	9	2

TABLE II
Real data streams characteristics.

Data stream	#I	#F	#C
NSECTS-abrupt_balanced_norm	1372	4	2
INSECTS-gradual_balanced$_{n}orm$	1372	4	2
INSECTS-incremental-abrupt_balanced$_{n}orm$	1372	4	2
NSECTS-incremental-reoccurring_balanced$_{n}orm$	1372	4	2
INSECTS-incremental_balanced_norm	1372	4	2
airlines	1372	4	2
covtype	1372	4	2
elec	1372	4	2
poker-lsn	1372	4	2

III. Full results of experiments

A. Experiment 1 – Hyperparametrization

The experiment presents streams from the stream-learn package. The impact of the values of individual WAE hyperparameter pairs on the quality of classification in the event of a given type of concept drift is presented.

Finally, the found pseudo-optimal hyperparameter values for WAE are presented for scenarios with full label access (Tables III-IX and Figure 1) and with limited labeling (only final hyperparameterization in Table X).

TABLE III
Experiment 1 — Pruning type (rows) vs aging method (columns).

		GNB			HT			MLP		
		PROP	CONST	GAUSS	PROP	CONST	GAUSS	PROP	CONST	GAUSS
SUDDEN	POST	**0.839**	0.815	0.736	**0.839**	0.815	0.74	**0.908**	0.879	0.812
	PRE	**0.838**	0.814	0.727	**0.839**	0.814	0.724	**0.908**	0.878	0.804
GRAD.	POST	**0.829**	0.811	0.752	**0.829**	0.811	0.755	**0.918**	0.895	0.851
	PRE	**0.827**	0.810	0.752	**0.827**	0.810	0.751	**0.916**	0.893	0.841
INC.	POST	**0.862**	0.841	0.778	**0.862**	0.841	0.773	**0.925**	0.900	0.829
	PRE	**0.861**	0.84	0.775	**0.860**	0.839	0.768	**0.922**	0.898	0.823

TABLE IV
Experiment 1 — Weight calculation method (rows) vs aging method (columns).

		GNB			HT			MLP		
		PROP	CONST	GAUSS	PROP	CONST	GAUSS	PROP	CONST	GAUSS
SUDDEN	SAME	**0.839**	0.813	0.629	**0.839**	0.815	0.636	**0.908**	0.875	0.714
	KUNCH.	**0.839**	**0.842**	0.825	**0.839**	**0.842**	0.823	**0.908**	**0.915**	0.902
	PTA	**0.839**	0.832	0.721	**0.839**	0.832	0.719	**0.908**	0.900	0.801
	BELL	**0.839**	0.769	0.749	**0.839**	0.770	0.75	**0.908**	0.823	0.815
GRAD.	SAME	**0.828**	0.820	0.697	**0.828**	0.820	0.702	**0.917**	**0.911**	0.81
	KUNCH.	**0.828**	**0.831**	0.788	**0.828**	**0.831**	0.787	**0.917**	**0.915**	0.861
	PTA	**0.828**	**0.825**	0.736	**0.828**	**0.825**	0.735	**0.917**	**0.914**	0.829
	BELL	**0.828**	0.767	0.787	**0.828**	0.766	0.789	**0.917**	0.836	0.883
INC.	SAME	**0.862**	0.851	0.714	**0.861**	0.851	0.703	**0.924**	0.909	0.755
	KUNCH.	**0.862**	**0.862**	0.819	**0.861**	**0.861**	0.817	**0.924**	**0.925**	0.877
	PTA	**0.862**	**0.857**	0.76	**0.861**	**0.857**	0.75	**0.924**	0.918	0.811
	BELL	**0.862**	0.792	0.812	**0.861**	0.791	0.813	**0.924**	0.843	0.861

TABLE V

EXPERIMENT 1 — WEIGHT CALCULATION METHOD (ROWS) VS PRUNING TYPE (COLUMNS).

		GNB		**HT**		**MLP**	
		POST	PRE	POST	PRE	POST	PRE
SUDDEN	SAME	0.764	0.757	0.768	0.758	0.834	0.830
	KUNCH.	**0.836**	**0.835**	**0.836**	**0.834**	**0.908**	**0.908**
	PTA	0.799	0.796	0.799	0.794	0.871	0.869
	BELL	0.788	0.784	0.789	0.783	0.851	0.846
GRAD.	SAME	0.782	0.781	0.785	0.782	0.883	0.876
	KUNCH.	**0.816**	**0.815**	**0.816**	**0.815**	**0.899**	**0.896**
	PTA	0.796	0.796	0.797	0.795	0.889	0.884
	BELL	0.795	0.793	0.796	0.793	0.881	0.877
INC.	SAME	0.810	0.808	0.808	0.802	0.864	0.861
	KUNCH.	**0.848**	**0.847**	**0.847**	**0.846**	**0.910**	**0.907**
	PTA	0.827	0.825	0.823	0.822	0.886	0.883
	BELL	0.824	0.820	0.823	0.820	0.879	0.873

TABLE VI

EXPERIMENT 1 — THETA (ROWS) VS AGING METHOD (COLUMNS).

		GNB			**HT**			**MLP**		
		PROP	CONST	GAUSS	PROP	CONST	GAUSS	PROP	CONST	GAUSS
SUDDEN	0%	**0.839**	0.811	0.731	**0.839**	0.812	0.732	**0.908**	0.872	0.808
	2.5%	**0.839**	0.828	0.731	**0.839**	0.829	0.732	**0.908**	0.894	0.808
	5%	**0.839**	**0.837**	0.731	**0.839**	**0.837**	0.732	**0.908**	**0.906**	0.808
	7.5%	**0.839**	**0.839**	0.731	**0.839**	**0.839**	0.732	**0.908**	**0.910**	0.808
	10%	**0.839**	0.755	0.731	**0.839**	0.755	0.732	**0.908**	0.808	0.808
GRADUAL	0%	**0.828**	0.813	0.752	**0.828**	0.813	0.753	**0.917**	0.903	0.846
	2.5%	**0.828**	**0.825**	0.752	**0.828**	**0.825**	0.753	**0.917**	**0.913**	0.846
	5%	**0.828**	**0.830**	0.752	**0.828**	**0.830**	0.753	**0.917**	**0.917**	0.846
	7.5%	**0.828**	**0.833**	0.752	**0.828**	**0.833**	0.753	**0.917**	**0.919**	0.846
	10%	**0.828**	0.752	0.752	**0.828**	0.752	0.753	**0.917**	0.817	0.846
INCREMENTAL	0%	**0.862**	0.844	0.776	**0.861**	0.844	0.771	**0.924**	0.900	0.826
	2.5%	**0.862**	0.857	0.776	**0.861**	**0.856**	0.771	**0.924**	0.918	0.826
	5%	**0.862**	**0.862**	0.776	**0.861**	**0.861**	0.771	**0.924**	**0.924**	0.826
	7.5%	**0.862**	**0.865**	0.776	**0.861**	**0.864**	0.771	**0.924**	**0.929**	0.826
	10%	**0.862**	0.775	0.776	**0.861**	0.775	0.771	**0.924**	0.824	0.826

TABLE VII
EXPERIMENT 1 — THETA (ROWS) VS PRUNING METHOD (COLUMNS).

		GNB		HT		MLP	
		POST	PRE	POST	PRE	POST	PRE
SUDDEN	0%	**0.796**	0.791	**0.798**	0.791	0.864	0.861
	2.5%	**0.801**	**0.798**	**0.803**	0.797	**0.871**	**0.869**
	5%	**0.804**	**0.800**	**0.806**	**0.800**	**0.875**	**0.873**
	7.5%	**0.805**	**0.801**	**0.806**	**0.801**	**0.877**	**0.874**
	10%	0.777	0.773	0.778	0.773	0.843	0.840
GRADUAL	0%	**0.799**	**0.797**	**0.800**	0.797	**0.892**	0.886
	2.5%	**0.802**	**0.801**	**0.803**	**0.801**	**0.895**	**0.890**
	5%	**0.804**	**0.803**	**0.805**	**0.802**	**0.896**	**0.891**
	7.5%	**0.805**	**0.804**	**0.806**	**0.804**	**0.896**	**0.892**
	10%	0.778	0.777	0.779	0.777	0.862	0.858
INCREMENTAL	0%	**0.829**	0.826	**0.827**	0.823	0.886	0.881
	2.5%	**0.833**	**0.830**	**0.831**	**0.828**	**0.891**	**0.887**
	5%	**0.834**	**0.832**	**0.832**	**0.829**	**0.893**	**0.889**
	7.5%	**0.835**	**0.833**	**0.833**	**0.831**	**0.894**	**0.891**
	10%	0.805	0.804	0.803	0.801	0.859	0.856

TABLE VIII
EXPERIMENT 1 — THETA (ROWS) VS WEIGHT CALCULATION METHOD (COLUMNS).

		GNB				HT				MLP			
		SAME	KUN.	PTA	BELL	SAME	KUN.	PTA	BELL	SAME	KUN.	PTA	BELL
SUDDEN	0%	0.755	**0.835**	0.793	0.792	0.758	**0.835**	0.792	0.793	0.825	**0.908**	0.864	0.854
	2.5%	0.757	**0.835**	0.795	0.809	0.760	**0.835**	0.795	0.810	0.828	**0.908**	0.868	0.877
	5%	0.761	**0.835**	0.797	0.815	0.763	**0.835**	0.797	0.816	0.832	**0.908**	0.870	0.885
	7.5%	0.763	**0.836**	0.799	0.815	0.766	**0.835**	0.799	0.815	0.836	**0.908**	0.873	0.884
	10%	0.766	**0.836**	0.801	0.697	0.769	**0.835**	0.800	0.697	0.841	**0.908**	0.875	0.742
GRADUAL	0%	0.778	**0.814**	0.793	0.807	0.780	**0.813**	0.792	0.808	0.876	0.896	0.883	0.900
	2.5%	0.78	**0.815**	0.795	**0.818**	0.781	**0.815**	0.794	**0.819**	0.877	0.897	0.885	**0.909**
	5%	0.782	**0.816**	0.796	**0.819**	0.783	**0.816**	0.796	**0.819**	0.879	0.898	0.886	**0.909**
	7.5%	0.784	**0.817**	0.798	**0.819**	0.785	**0.817**	0.798	**0.819**	0.881	0.899	0.888	**0.908**
	10%	0.786	**0.818**	0.800	0.706	0.788	**0.817**	0.799	0.707	0.883	0.899	0.890	0.768
INCREMENTAL	0%	0.805	**0.846**	0.822	0.837	0.801	**0.844**	0.819	0.837	0.857	**0.907**	0.879	0.890
	2.5%	0.807	**0.847**	0.824	**0.849**	0.803	**0.845**	0.821	**0.848**	0.859	**0.908**	0.882	**0.908**
	5%	0.809	**0.848**	0.826	**0.850**	0.805	**0.846**	0.823	**0.849**	0.862	**0.909**	0.885	**0.910**
	7.5%	0.811	**0.848**	0.828	**0.850**	0.807	**0.847**	0.824	**0.848**	0.866	**0.909**	0.887	**0.909**
	10%	0.813	**0.849**	0.83	0.726	0.809	**0.848**	0.826	0.726	0.869	**0.910**	0.889	0.763

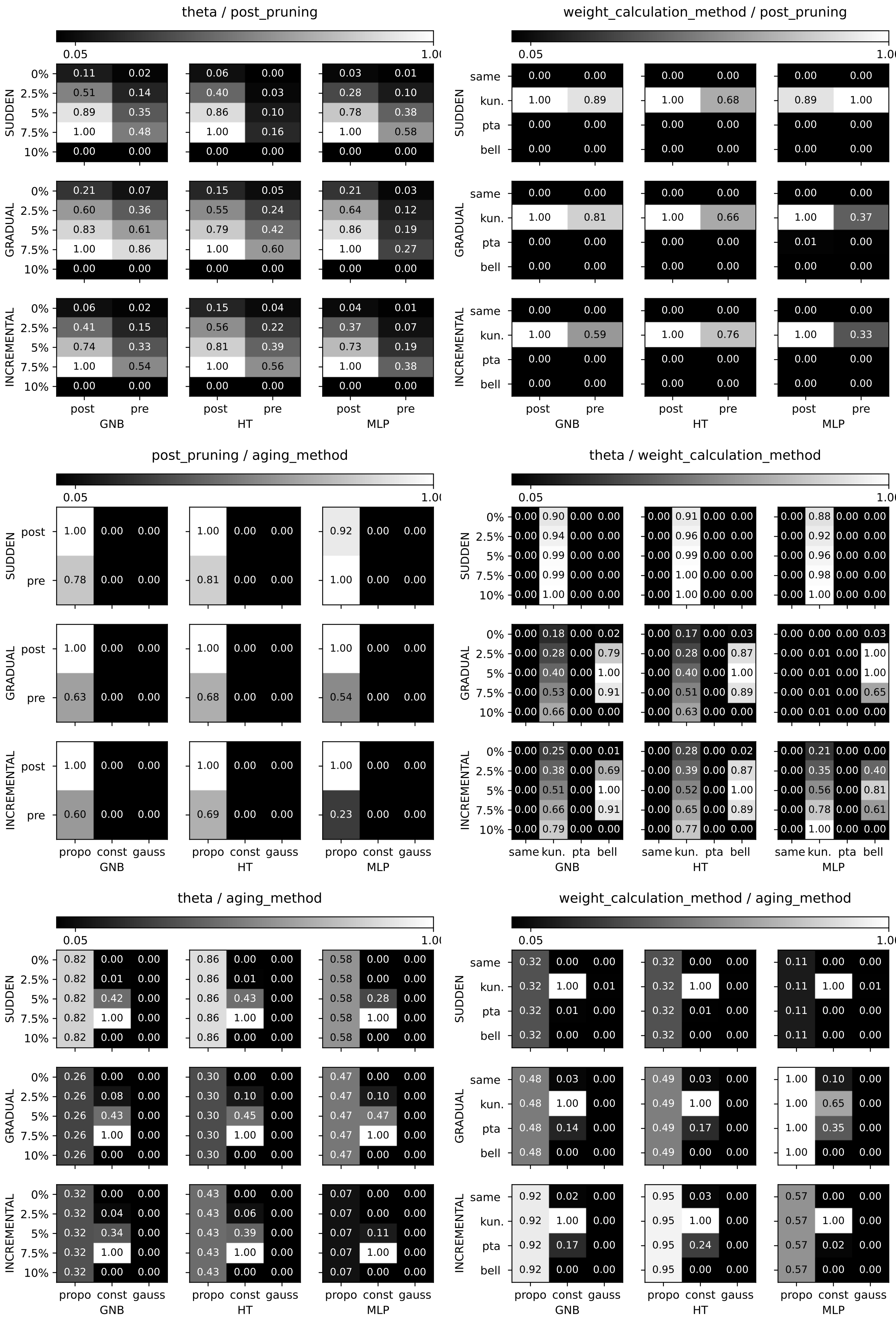

Fig. 1. EXPERIMENT 1 — Graphs of p-values achieved by various configurations of hyperparameter optimization.

TABLE IX
EXPERIMENT 1 — WAE OPTIMIZATION FOR REGULAR SCENARIOS.

	Drift type	Accuracy	Pruning	θ	WCM	Aging
GNB	SUDDEN	0.858	—	5%	BELL	CONST
	GRADUAL	0.843	—	5%	BELL	CONST
	INCREMENTAL	0.876	—	5%	BELL	CONST
HT	SUDDEN	0.858	—	5%	BELL	CONST
	GRADUAL	0.841	—	5%	BELL	CONST
	INCREMENTAL	0.873	—	5%	BELL	CONST
MLP	SUDDEN	0.934	—	5%	BELL	CONST
	GRADUAL	0.928	—	5%	BELL	CONST
	INCREMENTAL	0.944	—	5%	BELL	CONST

TABLE X
EXPERIMENT 1 — WAE OPTIMIZATION FOR ACTIVE LEARNING SCENARIOS.

	Drift type	Accuracy	Pruning	θ	WCM	Aging
GNB	SUDDEN	0.806	—	5%	BELL	CONST
	GRADUAL	0.788	—	5%	BELL	CONST
	INCREMENTAL	0.821	—	5%	BELL	CONST
HT	SUDDEN	0.791	POST	5%	PTA	CONST
	GRADUAL	0.778	PRE	7.5%	SAME	CONST
	INCREMENTAL	0.812	POST	7.5%	SAME	CONST
MLP	SUDDEN	0.871	—	5%	BELL	CONST
	GRADUAL	0.859	—	5%	BELL	CONST
	INCREMENTAL	0.887	—	5%	BELL	CONST

B. Experiment 2 – Comparison with state-of-the-art on sythetic streams

The averaged results of the WAE comparison with the default and optimized hyperparameterization with reference methods for a given drift type and base classifier are presented, supplemented with statistical analysis using the student's pair T-test. Results for scenarios with full (Table XI) and limited labeling (Table XII).

TABLE XI
EXPERIMENT 1 — FINAL RESULTS ON SYNTHETIC PROBLEMS FOR REGULAR SCENARIOS.

	SUDDEN			GRADUAL			INCREMENTAL		
	GNB	HT	MLP	GNB	HT	MLP	GNB	HT	MLP
(1) SEA	0.801	0.802	0.859	0.813	0.813	0.904	0.843	0.843	0.896
	–	–	5	–	–	5	–	–	5
(2) AWE	0.802	0.803	0.860	0.813	0.813	0.904	0.843	0.840	0.896
	–	–	5	–	–	5	–	–	5
(3) AUE	0.832	0.872	0.874	0.833	0.858	0.909	0.865	0.892	0.904
	1,2	*all*	1,2,5	1,2,5	1,2,4,6,7	5	1,2,5	*all*	5
(4) NSE	0.860	0.860	0.922	0.841	0.841	0.920	0.872	0.871	0.929
	1,2,3,5,6	1,2,6	1,2,3,5,6	1,2,5,6	1,2,6	1,2,3,5	1,2,5,6	1,2,6	1,2,3,5
(5) OALE	0.827	0.861	0.703	0.817	0.854	0.722	0.850	0.874	0.712
	1,2	1,2,6	–	–	1,2,4,6,7	–	–	1,2,6	–
(6) WAE^D	0.839	0.839	0.908	0.829	0.829	0.918	0.862	0.862	0.925
	1,2,5	1,2	1,2,3,5	1,2,5	1,2	1,2,3,5	1,2,5	1,2	1,2,3,5
(7) WAE^O	0.858	0.858	0.934	0.843	0.841	0.928	0.876	0.873	0.944
	1,2,3,5,6	1,2,6	*all*	1,2,3,5,6	1,2,6	*all*	1,2,3,5,6	1,2,6	*all*

TABLE XII
EXPERIMENT 2 — FINAL RESULTS ON SYNTHETIC PROBLEMS FOR ACTIVE LEARNING SCENARIOS.

	SUDDEN			GRADUAL			INCREMENTAL		
	GNB	HT	MLP	GNB	HT	MLP	GNB	HT	MLP
(1) SEA	0.842	0.843	0.871	0.856	0.855	0.909	0.886	0.885	0.920
	4,5,7	4,5,7	4,5	4,5,7	4,5,7	4,5,7	4,5,7	4,5,7	4,5,7
(2) AWE	0.841	0.842	0.870	0.855	0.854	0.908	0.885	0.886	0.919
	4,5,7	4,5,7	4,5	4,5,7	4,5,7	4,5,7	4,5,7	4,5,7	4,5,7
(3) AUE	0.874	0.884	0.871	0.863	0.873	0.909	0.898	0.903	0.920
	1,2,4,5,7	*all*	4,5	4,5,7	1,2,4,5,7	4,5,7	1,2,4,5,7	1,2,4,5,7	4,5,7
(4) NSE	0.536	0.540	0.586	0.540	0.550	0.595	0.539	0.541	0.598
	–	–	–	–	–	–	–	–	–
(5) OALE	0.788	0.789	0.604	0.773	0.780	0.608	0.812	0.814	0.608
	4	4	–	4	4	–	4	4	
(6) WAE^D	0.866	0.868	0.912	0.866	0.866	0.923	0.896	0.896	0.941
	1,2,4,5,7	1,2,4,5,7	*all*	4,5,7	4,5,7	*all*	2,4,5,7	4,5,7	*all*
(7) WAE^O	0.810	0.807	0.891	0.798	0.795	0.877	0.823	0.826	0.904
	4,5	4,5	1,2,3,4,5	4,5	4,5	4,5	4,5	4,5	4,5

IV. Experiment 3 – Comparison with state-of-the-art on real streams

Results comparing the efficiency on (a) real streams (Figure 2) and (b) streams with real concepts (Figures 3-4), presented using the accumulative sum of the flow visualization.

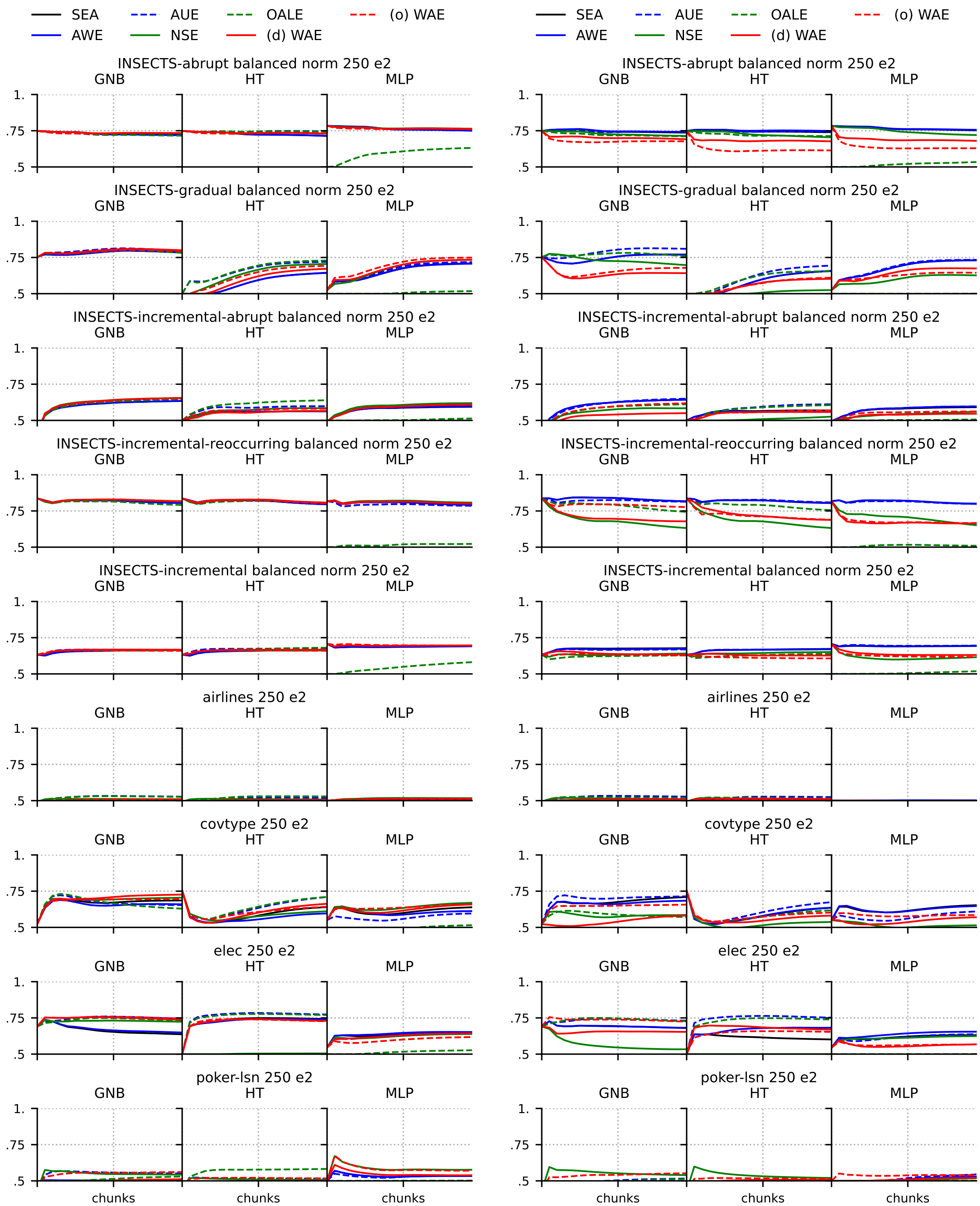

Fig. 2. Experiment 3 — final results on real streams for regular (left) and active learning scenarios (right).

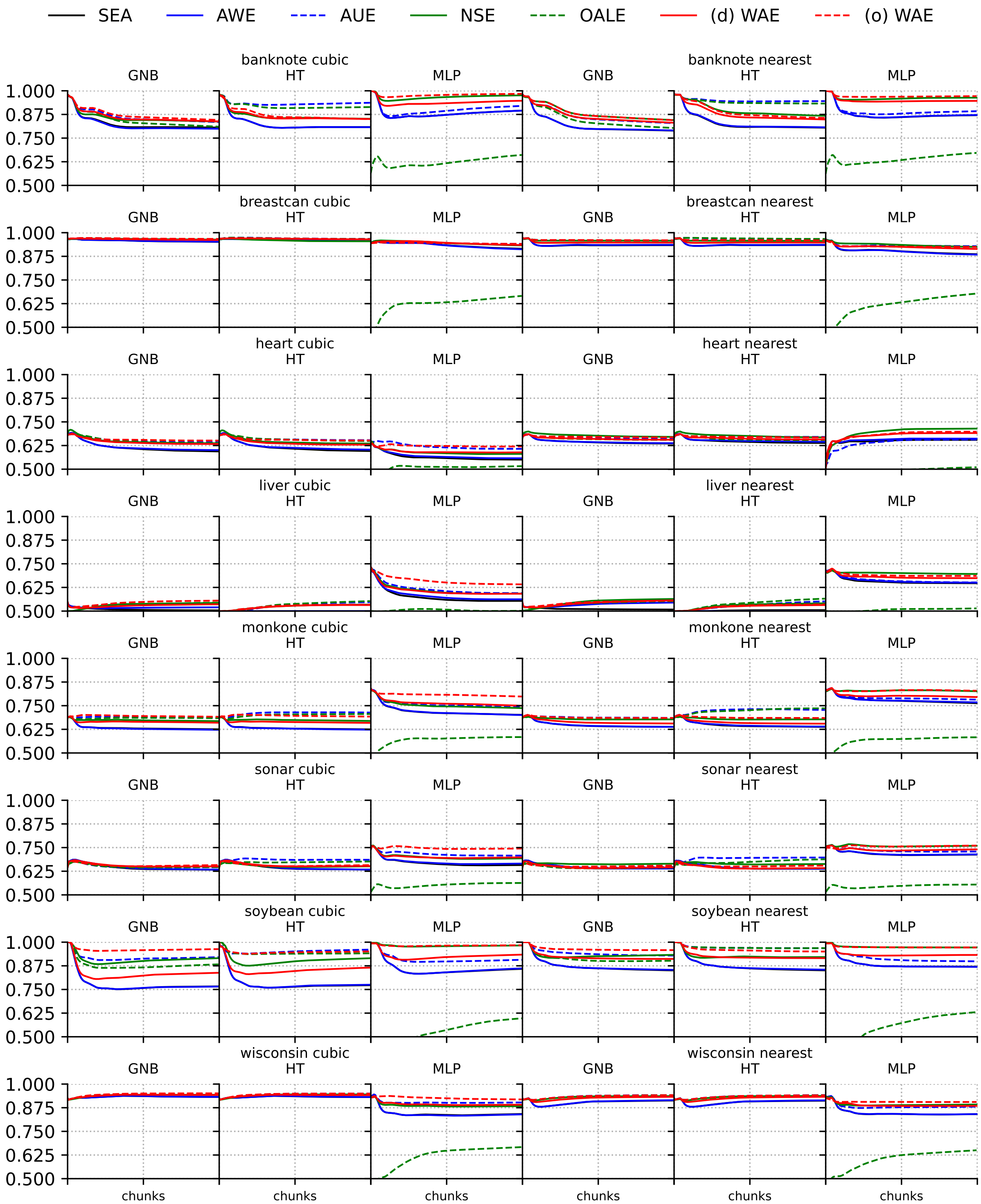

Fig. 3. EXPERIMENT 3 — final results on streams with real concepts for regular scenarios.

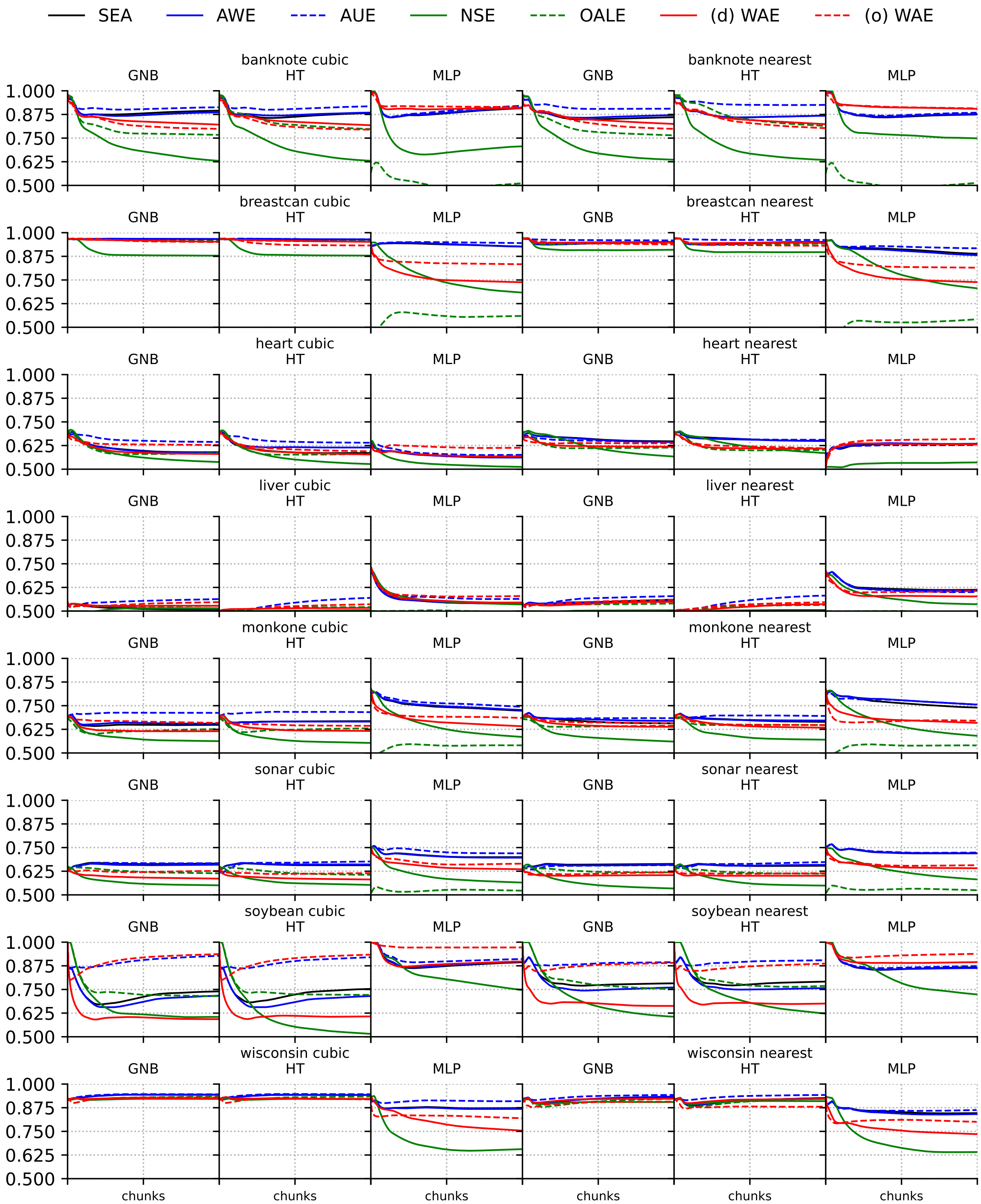

Fig. 4. EXPERIMENT 3 — final results on streams with real concepts for active learning scenarios.