Supplementary material for the article "Continual multi-label and multi-task learning for tabular data: proposal of a standardized protocol for task creation and classifier evaluation"

Anonymous Authors

A. Algorithms

- Neural networks [1] (NN1): a neural network without hidden layers, equivalent to logistic regression, was implemented using PyTorch. The output activation function employed is the sigmoid function. For the loss function, binary cross-entropy was utilized, which is commonly applied in multi-label classification problems [1]. The optimization algorithm implemented is Adam. The time complexity for processing an input instance is given by $\mathcal{O}(d*l)$, where d represents the number of input features and l denotes the number of labels. Parameters: learning rate [0.1, 0.01, 0.001].
- Neural network with a targeted loss function (NN2_TL): a variation of the NN1 model that introduces a focused approach to loss function computation. Unlike the traditional NN1, this modified version calculates the loss function exclusively on the labels that belong to the signature of the evaluated task[2]. In this approach, the losses corresponding to all other labels are set to zero. The time complexity for processing an input instance is given by $\mathcal{O}(d*l)$ where d is the number of input features and l denotes the number of labels. Parameters: learning rate [0.1, 0.01, 0.001].
- Neural network with a targeted loss function and a hidden layer (NN3_TLH): a variation of the NN2_TL model, which incorporates an additional hidden layer utilizing a ReLU activation function. The time complexity for processing an input instance is given by $\mathcal{O}(d*h+h*l)$ where d represents the number of input features, h denotes the size of the hidden layer, and l indicates the number of labels. Parameters: learning rate [0.1, 0.01, 0.001], hidden layer size [200, 2000].
- Neural network with a targeted loss function, a hidden layer, and a FIFO memory (NN4_TLH_fifo): a variation of the NN3_TLH model, which incorporates an online data replay with a FIFO memory. Each instance of the data stream is retained in a memory that deletes the oldest stored instance, thereby maintaining a fixed capacity. During the training phase, a subset of n memorized instances is randomly sampled from the FIFO memory to train the online neural network [2]. The time complexity for processing an input instance is given by $\mathcal{O}(r_f*(d*h+h*l))$ where d represents the number of input features, h denotes the size of the hidden layer, l indicates the number of labels, and r_f signifies the number of instances randomly sampled from the FIFO memory. Parameters: learning rate [0.1, 0.01, 0.001], size of hidden layer [200,2000], size of FIFO memory [100, 1000], number of instances randomly sampled from the FIFO memory [5,10].
- Neural network with targeted loss function, a hidden layer, and reservoir sampling (NN5_TLH_sampling): a variation of the NN3_TLH model, which incorporates an online data replay with a reservoir sampling. Each instance of the data stream is retained in a reservoir that deletes a random stored instance, thereby maintaining a fixed capacity. During the training phase, a subset of n memorized instances is randomly sampled from the reservoir to train the online neural network. The time complexity for processing an input instance is given by $\mathcal{O}(r_s*(d*h+h*l))$ where d represents the number of

[1]Normalization of the input feature vectors of the neural networks was considered. However, since the datasets did not require normalization and preliminary testing revealed negligible effects from its implementation, it was decided that normalization would not be employed in this study.

[2]In contrast to the other algorithms tested, neural networks with task-signature-targeted loss function computation can take advantage of the knowledge of the task in progress during learning.

input features, h denotes the size of the hidden layer, l indicates the number of labels, and r_s signifies the number of instances randomly sampled from the reservoir. Parameters: learning rate [0.1, 0.01, 0.001], size of the hidden layer [200,2000], size of the reservoir [100, 1000], number of instances randomly sampled from the reservoir [5,10].

- Neural network with targeted loss function, a hidden layer, FIFO memory, and reservoir sampling (NN6_TLH_memories): a variation of NN3_TLH, which incorporates an online data replay with a FIFO memory and a reservoir sampling. The time complexity for processing an input instance is given by $\mathcal{O}((r_s+r_f)*(d*h+h*l))$ where d represents the number of input features, h denotes the size of the hidden layer, l indicates the number of labels, r_s signifies the number of instances randomly sampled from the reservoir, and r_f is the number of instances randomly sampled from the FIFO memory. Parameters: learning rate [0.1, 0.01, 0.001], size of hidden layer [200,2000], size of FIFO memory [100, 1000], number of instances randomly sampled from the FIFO memory [5,10], size of the reservoir [100, 1000], number of instances randomly sampled from the reservoir [5,10].
- Neural network with targeted loss function, a hidden layer, with data replay with FIFO memory and reservoir sampling in the form of mini-batches (NN7_TLH_mini_memories): a variation of NN6_TLH_memories, which incorporates a single pass mini-batch learning for the current instance to learn and the data replay. The time complexity for processing an input instance is given by $\mathcal{O}((r_s+r_f)*(d*h+h*l))$ where d represents the number of input features, h denotes the size of the hidden layer, l indicates the number of labels, r_s signifies the number of instances randomly sampled from the reservoir, and r_f is the number of instances randomly sampled from the FIFO memory. Parameters: learning rate [0.1, 0.01, 0.001], size of hidden layer [200,2000], size of FIFO memory [100, 1000], number of instances randomly sampled from the FIFO memory [5,10], size of the reservoir [100, 1000], number of instances randomly sampled from the reservoir [5,10].
- T1_BR: the Hoeffding tree classifier, as proposed by Hulten et al. [3], adapted for the multi-label problem using the binary relevance method [4]. The time complexity for processing an input instance is given by $\mathcal{O}(l*d)$ where d represents the number of input features and l denotes the number of labels. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T2_LC: the Hoeffding tree classifier, adapted for the multi-label problem using the pruned set method [5], also called label combinations learning. The time complexity for processing an input instance is given by $\mathcal{O}(2^l*d)$ where d represents the number of input features and l denotes the number of labels. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T3_CC: the Hoeffding tree classifier, adapted for the multi-label problem using the classifier chain method [6]. The time complexity for processing an input instance is given by $\mathcal{O}(l*d)$ where d represents the number of input features and l denotes the number of labels. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T4_ARF: the Adaptive Random Forest algorithm, as proposed by Gomes et al. [7], adapted for the multi-label problem using the binary relevance method. The time complexity for processing an input instance is given by $\mathcal{O}(k*l*d)$ where d represents the number of input features, l denotes the number of labels, and k is the number of classifiers. Parameters: number of models [5, 10, 15], grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T5_SOUP: the Incremental Structured Output Prediction Tree algorithm, made for multi-target regression, as proposed by Osojnik et al. [8]. The time complexity for processing an input instance is given by $\mathcal{O}(l*d)$ where d represents the number of input features and l denotes the number of labels. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- B1_oracle: prediction of the correct label vector for each instance of the stream.
- B2_last_class: prediction of the label vector of the instance at $t-1$ for the instance at time t.
- B3_prior_distribution: prediction of most frequent value of each label in the past stream. The time complexity for processing an input instance is given by $\mathcal{O}(l)$ where l represents the number of labels.
- B4_mean: prediction of the mean value of each label in the past stream. The time complexity for processing an input instance is given by $\mathcal{O}(l)$ where l represents the number of labels.
- B5_1NN: prediction of the label vector of the instance in FIFO memory closest to the input instance in the feature space. The time complexity for processing an input instance is given by $\mathcal{O}(n*d)$ where n represents the size of the FIFO memory and d denotes the number of features. Parameters: FIFO memory size [0,10,100,1000], use of cosine or Euclidean similarity, similarity of two zero vectors to 1 or 0.

B. Synthetic datasets

To study the behavior of the algorithms in greater detail, three synthetic datasets were generated following our protocol: synth_monolab, which consists of tasks without common labels; synth_bilab, featuring tasks with a limited number of common labels to facilitate the observation of phenomena such as forward and backward transfers; and synth_rand, which presents a more complex scenario where the tasks all have the same label signature. Each dataset consists of four features and four labels, with features drawn randomly from a uniform distribution between 0 and 1. Three distinct M

matrices of dimensions 4×4 were constructed, where each row corresponds to a task within the dataset. The element $M_{i,j}$ denotes the multiplier coefficient applied to feature j in task i. After the application of the multiplier coefficient, the corresponding label is assigned a value of 1 if the feature value exceeds 0.5, and 0 otherwise.

- synth_monolab: this dataset employs a diagonal matrix.

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- synth_bilab: this dataset employs a matrix with two multiplier coefficient per task.

$$M = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- synth_rand: This dataset employs a randomly generated matrix of rank 4, to which a constant of 0.5 was added to mitigate the excessive occurrence of absent labels after thresholding at 0.5.

$$M = \begin{bmatrix} 1.1 & 1.3 & 1.4 & 1.3 \\ 0.9 & 1 & 1.5 & 1.1 \\ 1 & 1.3 & 0.6 & 1.2 \\ 0.7 & 0.6 & 0.6 & 1.4 \end{bmatrix}$$

C. Statistical tests

The Friedman test was successfully applied to all metrics, revealing significant differences among the algorithms in the results. Although the Nemenyi post hoc test did not provide conclusive results, this is not a concern, as the primary objective was to present the protocol in detail. A more thorough analysis will require conducting a new benchmark with a broader range of datasets.

D. Complementary results

Additional results are provided in this paper, including measurements for RMSE, precision@p, and estimated total consumption. Comprehensive tables of the metrics from the article, along with results for the baseline, are also included.

E. Illustration of the relationship between $FWT-$ and catastrophic forgetting

This section presents an illustration of the relationship between negative forward transfer $FWT-$ and catastrophic forgetting, utilizing a simplified scenario.

- Task 1. Labels: L1, L2, L3, L4.
- Task 2. Labels: L3, L4, L5, L6.
- Task 3. Labels: L5, L6, L7, L8.
- Task 4. Labels: L1, L2, L7, L8.

1) Task 1 appears in the stream: $FWT+$ is observed in task 2 thanks to the labels L3 and L4, and in task 4 via labels labels L1 and L2.
2) Task 2 appears in the stream: $FWT+$ is observed in task 3 thanks to the labels L5 and L6. However, there is forgetting associated with labels L1 and L2, resulting in negative forward transfer $FWT-$ for task 4.
3) etc.

This example illustrates the observations made with the synth_bilab dataset, which was developed to facilitate the examination of transfer phenomena.

References

[1] A. Maxwell, R. Li, B. Yang, H. Weng, A. Ou, H. Hong, Z. Zhou, P. Gong, and C. Zhang, "Deep learning architectures for multi-label classification of intelligent health risk prediction," *BMC bioinformatics*, vol. 18, pp. 121–131, 2017.

[2] P. Buzzega, M. Boschini, A. Porrello, and S. Calderara, "Rethinking experience replay: a bag of tricks for continual learning," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 2180–2187.

[3] G. Hulten, L. Spencer, and P. Domingos, "Mining time-changing data streams," in *SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 97–106.

[4] M.-L. Zhang, Y.-K. Li, X.-Y. Liu, and X. Geng, "Binary relevance for multi-label learning: an overview," *Frontiers of Computer Science*, vol. 12, pp. 191–202, 2018.

[5] J. Read, B. Pfahringer, and G. Holmes, "Multi-label classification using ensembles of pruned sets," in *2008 eighth IEEE international conference on data mining*. IEEE, 2008, pp. 995–1000.

[6] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," in *European Conference on Machine Learning*. Springer, 2009, pp. 254–269.

[7] H. M. Gomes, A. Bifet, J. Read, J. P. Barddal, F. Enembreck, B. Pfharinger, G. Holmes, and T. Abdessalem, "Adaptive random forests for evolving data stream classification," *Machine Learning*, vol. 106, pp. 1469–1495, 2017.

[8] A. Osojnik, P. Panov, and S. Džeroski, "Tree-based methods for online multi-target regression," *Journal of Intelligent Information Systems*, vol. 50, pp. 315–339, 2018.

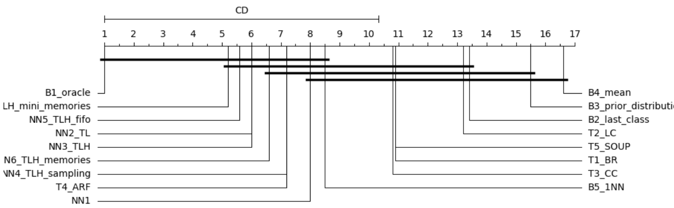

Fig. 1. BA_{macro}

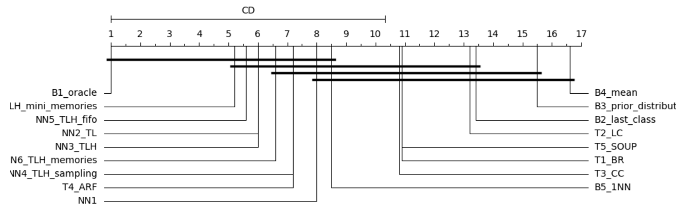

Fig. 2. Precision@p

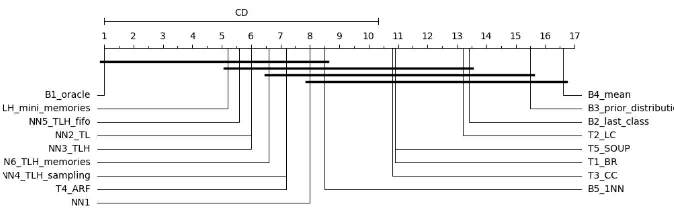

Fig. 3. RMSE

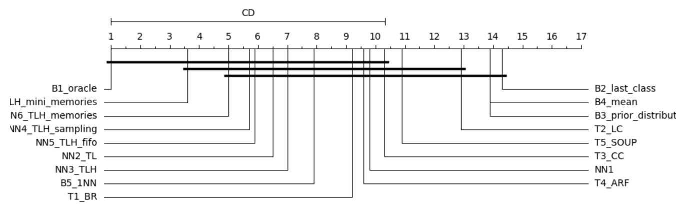

Fig. 4. Average accuracy

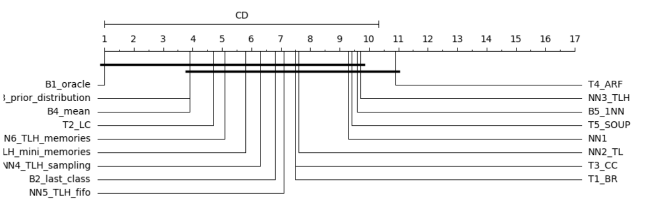

Fig. 5. $BWT-$

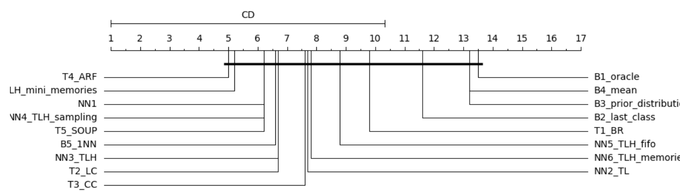

Fig. 6. $BWT+$

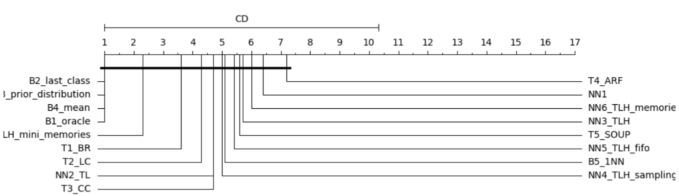

Fig. 7. $FWT-$

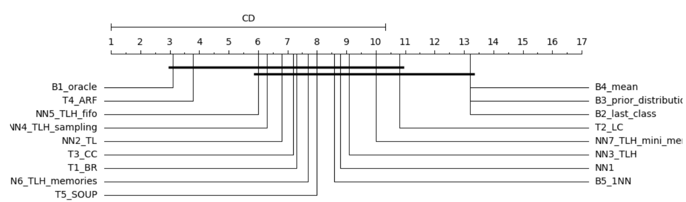

Fig. 8. $FWT+$

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.880	0.910	0.938	0.767	0.539	0.667	0.514	0.857	0.518	0.732	8.000
NN2_TL	0.899	0.926	0.940	0.781	0.673	**0.673**	0.512	0.864	0.509	0.753	6.000
NN3_TLH	0.908	0.945	0.952	0.865	**0.687**	0.647	0.525	0.847	0.500	0.764	6.000
NN4_TLH_sampling	0.911	0.961	0.889	0.734	0.621	0.658	0.523	0.871	0.501	0.741	7.200
NN5_TLH_fifo	0.949	0.959	0.952	**0.884**	0.623	0.666	0.607	0.552	0.515	0.745	5.600
NN6_TLH_memories	0.919	**0.964**	0.924	0.744	0.616	0.657	0.526	0.851	0.504	0.745	6.600
NN7_TLH_mini_memories	0.930	0.960	0.938	0.773	0.654	0.657	**0.627**	**0.880**	0.500	**0.769**	**5.200**
T1_BR	0.722	0.892	0.889	0.664	0.593	0.546	0.518	0.644	0.518	0.665	10.900
T2_LC	0.613	0.668	0.771	0.696	0.560	0.493	0.515	0.648	0.500	0.607	13.200
T3_CC	0.651	0.769	0.887	0.696	0.609	0.532	0.524	0.640	0.519	0.647	10.800
T4_ARF	**0.952**	0.937	**0.954**	0.742	0.626	0.547	0.523	0.607	**0.526**	0.713	7.200
T5_SOUP	0.768	0.787	0.846	0.670	0.624	0.566	0.510	0.744	0.506	0.669	10.900
B1_oracle	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
B2_last_class	0.490	0.504	0.556	0.603	0.640	0.562	0.519	0.510	0.500	0.543	13.400
B3_prior_distribution	0.496	0.495	0.523	0.478	0.522	0.507	0.510	0.500	0.500	0.503	15.500
B4_mean	0.326	0.327	0.439	0.379	0.413	0.337	0.347	0.475	0.500	0.394	16.600
B5_1NN	0.613	0.747	0.858	0.717	0.651	0.656	0.605	0.813	0.630	0.699	8.500

TABLE I
FINAL BA_{macro} OF 17 STRATEGIES ON 9 DATASETS.

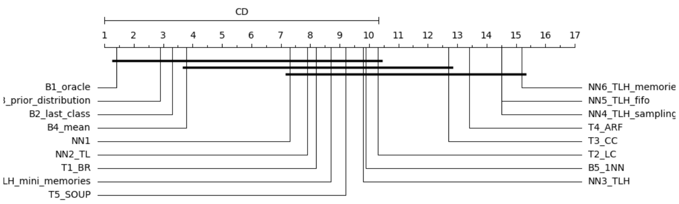

Fig. 9. Energy consumed

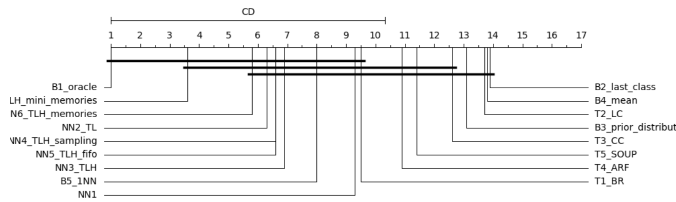

Fig. 10. Frugality score

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.675	0.768	0.880	0.641	0.531	0.608	0.641	0.657	0.565	0.663	9.800
NN2_TL	0.943	0.943	0.880	0.678	0.541	0.614	0.643	0.678	0.565	0.721	6.500
NN3_TLH	0.929	0.948	0.889	0.656	0.539	0.617	0.645	0.664	0.564	0.717	7.000
NN4_TLH_sampling	0.943	0.977	0.915	0.701	**0.581**	0.609	0.643	0.661	0.564	0.733	5.700
NN5_TLH_fifo	0.960	0.971	0.896	0.689	0.575	0.609	0.668	0.645	0.564	0.731	5.900
NN6_TLH_memories	0.950	0.977	0.911	0.704	**0.581**	0.615	0.639	0.684	0.564	0.736	5.000
NN7_TLH_mini_memories	**0.968**	**0.981**	0.899	**0.720**	0.570	**0.635**	**0.683**	**0.695**	0.564	**0.746**	**3.600**
T1_BR	0.642	0.867	**0.918**	0.640	0.535	0.588	0.639	0.620	**0.568**	0.668	9.200
T2_LC	0.641	0.679	0.801	0.655	0.557	0.429	0.632	0.586	0.436	0.602	12.900
T3_CC	0.638	0.791	0.917	0.632	0.527	0.572	0.641	0.623	0.566	0.656	10.300
T4_ARF	0.689	0.789	0.909	0.630	0.522	0.568	0.644	0.634	0.567	0.661	9.600
T5_SOUP	0.650	0.768	0.875	0.582	0.523	0.592	0.644	0.628	0.565	0.647	10.900
B1_oracle	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
B2_last_class	0.500	0.500	0.500	0.522	0.505	0.558	0.635	0.625	0.564	0.546	14.300
B3_prior_distribution	0.500	0.500	0.500	0.511	0.504	0.571	0.640	0.625	0.564	0.546	13.900
B4_mean	0.500	0.500	0.500	0.511	0.504	0.571	0.640	0.625	0.564	0.546	13.900
B5_1NN	0.666	0.763	0.892	0.676	0.534	0.579	0.653	0.665	0.614	0.671	7.900

TABLE II

AVERAGE ACCURACY OF 17 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	-0.460	-0.293	-0.077	-0.177	-0.016	-0.039	-0.004	-0.096	-0.002	-0.129	9.300
NN2_TL	-0.028	-0.019	-0.077	-0.128	-0.067	-0.036	-0.004	-0.078	-0.001	-0.049	7.600
NN3_TLH	-0.042	-0.019	-0.083	-0.118	-0.087	-0.044	-0.009	-0.123	**0.000**	-0.058	9.700
NN4_TLH_sampling	-0.031	-0.013	-0.040	-0.054	-0.024	-0.040	-0.015	-0.084	$\mathbf{-0.000}$	-0.033	6.300
NN5_TLH_fifo	-0.018	-0.011	-0.092	-0.117	-0.019	-0.043	-0.025	-0.026	-0.003	-0.039	7.100
NN6_TLH_memories	-0.023	$\mathbf{-0.008}$	-0.055	$\mathbf{-0.033}$	-0.026	-0.053	-0.007	-0.069	**0.000**	$\mathbf{-0.030}$	5.100
NN7_TLH_mini_memories	$\mathbf{-0.012}$	-0.010	-0.077	-0.046	-0.026	-0.040	-0.018	-0.070	**0.000**	-0.033	5.800
T1_BR	-0.259	-0.144	-0.045	-0.081	-0.043	-0.013	-0.005	-0.051	-0.005	-0.072	7.500
T2_LC	-0.243	-0.144	$\mathbf{-0.029}$	-0.038	$\mathbf{-0.010}$	**0.000**	-0.004	$\mathbf{-0.014}$	**0.000**	-0.054	**4.700**
T3_CC	-0.217	-0.225	-0.042	-0.071	-0.026	-0.025	-0.009	-0.064	-0.003	-0.076	7.500
T4_ARF	-0.490	-0.350	-0.098	-0.111	-0.058	-0.027	-0.006	-0.095	-0.006	-0.138	10.900
T5_SOUP	-0.387	-0.256	-0.079	-0.119	-0.035	-0.038	$\mathbf{-0.003}$	-0.075	-0.006	-0.111	9.400
B1_oracle	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
B2_last_class	0.000	0.000	0.000	-0.200	-0.080	-0.064	-0.033	0.000	0.000	-0.042	6.800
B3_prior_distribution	0.000	0.000	0.000	-0.200	-0.080	0.000	0.000	0.000	0.000	-0.031	3.900
B4_mean	0.000	0.000	0.000	-0.200	-0.080	0.000	0.000	0.000	0.000	-0.031	3.900
B5_1NN	-0.094	-0.050	-0.014	-0.023	-0.080	-0.076	-0.034	-0.130	-0.010	-0.057	9.600

TABLE III

AVERAGE NEGATIVE BACKWARD TRANSFER (BWT-) OF 17 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.000	0.124	0.074	**0.073**	0.010	0.022	0.001	0.072	0.003	**0.042**	6.200
NN2_TL	0.000	0.014	**0.075**	0.012	0.027	0.039	0.001	0.016	0.001	0.020	7.700
NN3_TLH	0.021	0.022	0.069	0.014	0.017	0.039	0.005	0.031	0.000	0.024	6.700
NN4_TLH_sampling	0.026	0.011	0.031	0.050	0.016	**0.063**	0.007	0.028	0.000	0.026	6.200
NN5_TLH_fifo	0.015	0.010	0.061	0.016	0.016	0.022	0.011	0.011	0.001	0.018	8.800
NN6_TLH_memories	0.016	0.007	0.043	0.030	0.019	0.040	0.005	0.022	0.000	0.020	7.800
NN7_TLH_mini_memories	0.018	0.009	0.067	0.042	0.022	0.032	0.019	**0.077**	0.000	0.032	5.200
T1_BR	0.000	0.071	0.044	0.007	0.015	0.016	0.003	0.006	0.006	0.019	9.800
T2_LC	**0.094**	0.053	0.012	0.017	**0.038**	0.000	**0.020**	0.012	0.000	0.027	6.700
T3_CC	0.000	0.077	0.024	0.007	0.013	0.022	0.015	0.060	0.004	0.025	7.600
T4_ARF	0.000	0.124	0.067	0.037	0.027	0.028	0.002	0.035	**0.010**	0.036	**5.000**
T5_SOUP	0.000	**0.144**	0.063	0.015	0.025	0.035	0.001	0.020	0.007	0.035	6.200
B1_oracle	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	13.500
B2_last_class	0.000	0.000	0.000	0.000	0.019	0.018	0.003	0.000	0.000	0.004	11.600
B3_prior_distribution	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.001	13.200
B4_mean	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.001	13.200
B5_1NN	0.004	0.004	0.006	0.018	0.029	0.051	0.006	0.066	0.003	0.021	6.600

TABLE IV

AVERAGE POSITIVE BACKWARD TRANSFER (BWT+) OF 17 STRATÉGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	**0.000**	−0.226	−0.049	−0.062	−0.021	−0.065	**0.000**	**0.000**	**0.000**	−0.047	6.400
NN2_TL	**0.000**	−0.005	−0.049	−0.008	−0.017	−0.014	**0.000**	−0.081	−0.002	−0.019	4.700
NN3_TLH	−0.018	−0.027	−0.047	−0.137	−0.025	**0.000**	**0.000**	−0.037	**0.000**	−0.032	5.700
NN4_TLH_sampling	−0.057	−0.012	−0.011	−0.025	−0.026	**0.000**	−0.001	−0.076	**0.000**	−0.023	5.000
NN5_TLH_fifo	**0.000**	**0.000**	−0.044	−0.131	−0.012	−0.023	−0.022	−0.004	−0.003	−0.026	5.400
NN6_TLH_memories	−0.138	−0.034	−0.011	−0.196	−0.021	**0.000**	**0.000**	−0.038	**0.000**	−0.049	6.000
NN7_TLH_mini_memories	−0.016	**0.000**	−0.006	**0.000**	**0.000**	**0.000**	−0.029	−0.021	**0.000**	**−0.008**	**2.300**
T1_BR	**0.000**	−0.133	−0.009	−0.034	−0.003	**−0.000**	−0.004	**0.000**	−0.001	−0.020	3.600
T2_LC	−0.030	−0.235	−0.022	**0.000**	−0.003	**0.000**	−0.002	**0.000**	**0.000**	−0.032	4.300
T3_CC	**0.000**	−0.015	−0.023	−0.029	−0.003	**0.000**	−0.008	−0.051	−0.004	−0.015	4.700
T4_ARF	**0.000**	−0.126	−0.019	−0.169	−0.022	−0.018	−0.003	−0.049	−0.004	−0.046	7.200
T5_SOUP	**0.000**	−0.210	**0.000**	−0.125	−0.025	−0.010	**0.000**	−0.036	−0.003	−0.046	5.600
B1_oracle	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
B2_last_class	0.000	0.000	0.000	0.000	0.000	0.000	80.000	0.000	0.000	0.000	1.000
B3_prior_distribution	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
B4_mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
B5_1NN	0.000	−0.109	−0.010	0.000	−0.004	−0.041	−0.016	−0.032	−0.021	−0.026	5.100

TABLE V

AVERAGE NEGATIVE FORWARD TRANSFER (FWT-) OF 17 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.167	0.292	0.314	0.501	0.265	**0.317**	0.202	0.225	0.190	0.275	8.800
NN2_TL	0.185	0.360	0.314	0.520	0.265	0.296	0.192	0.299	0.287	0.302	6.800
NN3_TLH	0.166	0.374	0.310	**0.625**	0.183	0.219	0.191	0.286	0.189	0.282	9.100
NN4_TLH_sampling	0.298	0.376	0.301	0.508	**0.275**	0.227	0.286	0.299	0.190	0.307	6.300
NN5_TLH_fifo	0.215	0.358	0.322	0.545	0.272	0.216	0.288	0.245	0.289	0.305	6.000
NN6_TLH_memories	**0.318**	0.374	0.315	0.576	0.261	0.215	0.192	0.281	0.189	0.302	7.700
NN7_TLH_mini_memories	0.214	0.333	0.306	0.273	0.180	0.202	0.286	0.289	0.194	0.253	10.000
T1_BR	0.167	0.329	**0.456**	0.500	0.253	0.194	0.303	0.196	**0.571**	0.330	7.300
T2_LC	0.250	0.290	0.436	0.277	0.198	0.140	0.228	0.186	0.143	0.239	10.800
T3_CC	0.167	0.388	0.309	0.500	0.249	0.204	0.304	0.309	0.287	0.302	7.200
T4_ARF	0.167	**0.455**	0.324	0.548	0.260	0.293	**0.607**	**0.314**	0.289	**0.362**	**3.800**
T5_SOUP	0.167	0.289	0.292	0.509	0.248	0.219	0.202	0.308	0.570	0.311	8.000
B1_oracle	0.333	0.333	0.333	0.500	0.333	0.333	0.333	0.333	0.333	0.352	3.100
B2_last_class	0.167	0.167	0.167	0.250	0.162	0.193	0.202	0.202	0.190	0.189	13.200
B3_prior_distribution	0.167	0.167	0.167	0.250	0.162	0.193	0.202	0.202	0.190	0.189	13.200
B4_mean	0.167	0.167	0.167	0.250	0.162	0.193	0.202	0.202	0.190	0.189	13.200
B5_1NN	0.167	0.274	0.294	0.252	0.266	0.307	0.206	0.321	0.284	0.263	8.600

TABLE VI

AVERAGE POSITIVE FORWARD TRANSFER (FWT+) OF 17 STRATEGIES ON 9 DATASETS.

	Task 1	Task 2	Avg. value	Avg. rank
NN1	90.1	7.4	48.75	4.5
NN2_TL	131.7	8	69.85	6.5
NN3_TLH	221.8	11.4	116.6	9
NN4_TLH_sampling	32	10.7	21.35	4.5
NN5_TLH_fifo	54.4	18	36.2	7.5
NN6_TLH_memories	48.4	29.9	39.15	7.5
NN7_TLH_mini_memories	84.1	29.1	56.6	8.5
T1_BR	**14.1**	**1.4**	**7.75**	**1**
T2_LC	215.8	487.6	351.7	11.5
T3_CC	26.5	8.8	17.65	3.5
T4_ARF	34.7	5.9	20.3	3
T5_SOUP	339.5	40.6	190.05	11.5
B5_1NN	475.1	320.8	397.95	12.5

TABLE VII

AVERAGE NUMBER OF INSTANCES NEEDED TO ATTAIN THE TARGETED BA_{macro} FOR 13 ALGORITHMS OVER 10 RUNS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.675	0.768	0.880	0.641	0.531	0.602	0.641	0.655	0.559	0.661	9.300
NN2_TL	0.943	0.943	0.880	0.678	0.541	0.607	0.642	0.677	**0.560**	0.719	6.300
NN3_TLH	0.929	0.948	0.889	0.655	0.538	0.610	0.644	0.663	0.558	0.715	6.900
NN4_TLH_sampling	0.942	0.977	0.914	0.701	**0.581**	0.595	0.638	0.649	0.551	0.728	6.600
NN5_TLH_fifo	0.960	0.971	0.896	0.689	0.574	0.599	0.665	0.622	0.543	0.724	6.600
NN6_TLH_memories	0.949	0.977	0.910	0.703	**0.581**	0.603	0.636	0.678	0.537	0.730	5.800
NN7_TLH_mini_memories	**0.968**	**0.981**	0.899	**0.720**	0.570	**0.630**	**0.682**	**0.694**	0.558	**0.744**	**3.600**
T1_BR	0.642	0.867	**0.918**	0.639	0.535	0.581	0.638	0.615	0.558	0.666	9.500
T2_LC	0.641	0.679	0.801	0.655	0.556	0.424	0.632	0.564	0.378	0.592	13.700
T3_CC	0.638	0.791	0.917	0.632	0.527	0.497	0.489	0.611	0.540	0.627	12.600
T4_ARF	0.688	0.789	0.909	0.630	0.522	0.542	0.642	0.628	0.539	0.654	10.900
T5_SOUP	0.650	0.768	0.875	0.582	0.523	0.586	0.643	0.622	0.553	0.645	11.400
B1_oracle	1.000	1.000	1.000	1.000	1.000	0.995	0.999	1.000	0.996	0.999	1.000
B2_last_class	0.500	0.500	0.500	0.522	0.505	0.553	0.635	0.625	0.560	0.544	13.900
B3_prior_distribution	0.500	0.500	0.500	0.511	0.504	0.566	0.639	0.625	0.560	0.545	13.100
B4_mean	0.500	0.500	0.500	0.511	0.504	0.566	0.639	0.625	0.560	0.545	13.800
B5_1NN	0.666	0.763	0.891	0.676	0.534	0.574	0.652	0.663	0.572	0.666	8.000

TABLE VIII

FRUGALITY SCORE OF 17 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.006	**0.001**	0.002	**0.005**	**0.002**	**7.300**
NN2_TL	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.007	**0.001**	**0.001**	0.006	**0.002**	7.900
NN3_TLH	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.007	**0.001**	**0.001**	0.006	**0.002**	9.800
NN4_TLH_sampling	0.001	**0.000**	0.001	**0.000**	0.001	0.014	0.005	0.011	0.013	0.005	14.500
NN5_TLH_fifo	**0.000**	0.001	**0.000**	**0.000**	0.001	0.011	0.002	0.023	0.021	0.007	14.500
NN6_TLH_memories	**0.000**	0.001	0.001	0.001	0.001	0.012	0.003	0.007	0.028	0.006	15.200
NN7_TLH_mini_memories	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.006	**0.001**	**0.001**	0.006	**0.002**	8.700
T1_BR	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.007	**0.001**	0.006	0.010	0.003	8.200
T2_LC	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.005**	**0.001**	0.023	0.062	0.010	10.300
T3_CC	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.081	0.180	0.012	0.027	0.033	12.700
T4_ARF	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.026	0.002	0.005	0.029	0.007	13.400
T5_SOUP	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.006	**0.001**	0.007	0.012	0.003	9.200
B1_oracle	0.000	0.000	0.000	0.000	0.000	0.005	0.001	0.000	0.004	0.001	1.400
B2_last_class	0.000	0.000	0.000	0.000	0.000	0.005	0.001	0.000	0.004	0.001	3.300
B3_prior_distribution	0.000	0.000	0.000	0.000	0.000	0.006	0.001	0.000	0.004	0.001	2.900
B4_mean	0.000	0.000	0.000	0.000	0.000	0.006	0.001	0.000	0.004	0.001	3.800
B5_1NN	0.000	0.000	0.000	0.000	0.000	0.005	0.001	0.002	0.043	0.006	9.900

TABLE IX

RMSE OF 17 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.880	0.910	0.938	0.767	0.539	0.667	0.514	0.857	0.518	0.732	8.000
NN2_TL	0.899	0.926	0.940	0.781	0.673	**0.673**	0.512	0.864	0.509	0.753	6.000
NN3_TLH	0.908	0.945	0.952	0.865	**0.687**	0.647	0.525	0.847	0.500	0.764	6.000
NN4_TLH_sampling	0.911	0.961	0.889	0.734	0.621	0.658	0.523	0.871	0.501	0.741	7.200
NN5_TLH_fifo	0.949	0.959	0.952	**0.884**	0.623	0.666	0.607	0.552	0.515	0.745	5.600
NN6_TLH_memories	0.919	**0.964**	0.924	0.744	0.616	0.657	0.526	0.851	0.504	0.745	6.600
NN7_TLH_mini_memories	0.930	0.960	0.938	0.773	0.654	0.657	**0.627**	**0.880**	0.500	**0.769**	**5.200**
T1_BR	0.722	0.892	0.889	0.664	0.593	0.546	0.518	0.644	0.518	0.665	10.900
T2_LC	0.613	0.668	0.771	0.696	0.560	0.493	0.515	0.648	0.500	0.607	13.200
T3_CC	0.651	0.769	0.887	0.696	0.609	0.532	0.524	0.640	0.519	0.647	10.800
T4_ARF	**0.952**	0.937	**0.954**	0.742	0.626	0.547	0.523	0.607	**0.526**	0.713	7.200
T5_SOUP	0.768	0.787	0.846	0.670	0.624	0.566	0.510	0.744	0.506	0.669	10.900
B1_oracle	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
B2_last_class	0.490	0.504	0.556	0.603	0.640	0.562	0.519	0.510	0.500	0.543	13.400
B3_prior_distribution	0.496	0.495	0.523	0.478	0.522	0.507	0.510	0.500	0.500	0.503	15.500
B4_mean	0.326	0.327	0.439	0.379	0.413	0.337	0.347	0.475	0.500	0.394	16.600
B5_1NN	0.613	0.747	0.858	0.717	0.651	0.656	0.605	0.813	0.630	0.699	8.500

TABLE X

PRECISION@P OF 17 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.028	0.042	0.051	**0.034**	**0.067**	5.770	0.747	1.656	**5.418**	**1.535**	**7.300**
NN2_TL	0.040	0.044	0.054	0.041	**0.067**	6.751	0.848	**0.999**	5.532	1.597	7.900
NN3_TLH	0.069	0.079	0.094	0.110	0.128	7.000	0.886	1.321	6.422	1.790	9.800
NN4_TLH_sampling	0.519	0.315	0.522	0.486	0.595	14.172	4.847	11.382	13.326	5.129	14.500
NN5_TLH_fifo	0.309	0.551	0.361	0.389	0.551	10.536	2.342	23.425	20.512	6.553	14.500
NN6_TLH_memories	0.491	0.778	0.820	1.109	0.766	11.655	2.934	6.723	27.663	5.882	15.200
NN7_TLH_mini_memories	0.073	0.057	0.099	0.107	0.109	5.604	0.816	1.246	6.483	1.621	8.700
T1_BR	0.019	0.025	0.033	0.138	0.186	6.788	**0.583**	5.572	9.722	2.563	8.200
T2_LC	**0.015**	**0.022**	0.036	0.195	0.814	**4.848**	0.629	23.312	61.610	10.165	10.300
T3_CC	0.023	0.029	0.038	0.217	0.227	81.481	180.254	11.739	26.825	33.426	12.700
T4_ARF	0.121	0.069	0.221	0.227	0.330	26.234	2.302	5.266	28.961	7.081	13.400
T5_SOUP	0.024	**0.022**	**0.030**	0.146	0.130	6.412	0.898	6.503	11.951	2.902	9.200
B1_oracle	0.011	0.015	0.022	0.012	0.034	5.146	0.583	0.411	3.663	1.100	1.400
B2_last_class	0.014	0.016	0.024	0.012	0.035	5.447	0.739	0.423	3.761	1.164	3.300
B3_prior_distribution	0.013	0.016	0.023	0.012	0.037	5.587	0.599	0.320	3.789	1.155	2.900
B4_mean	0.013	0.017	0.025	0.013	0.039	5.696	0.591	0.337	4.116	1.205	3.800
B5_1NN	0.187	0.328	0.354	0.106	0.049	5.124	0.605	1.744	43.362	5.762	9.900

TABLE XI

ESTIMATED TOTAL CONSUMPTION (WH) OF 17 STRATEGIES ON 9 DATASETS.