Appendix

A.1 Datasets

To show that our method works well with various kinds of datasets, we choose the following seven image and vector datasets. A brief description is given in Tab. A1.

(1). MNIST-full [8]: The MNIST-full dataset consists of 70000 handwritten digits of 28 $\times$ 28 pixels. Each gray image is reshaped to a 784-dimensional vector.

(2). MNIST-test [8]: The MNIST-test is the testing part of MNIST dataset, which contains a total of 10000 samples.

(3). USPS: The USPS dataset is composed of 9298 gray-scale handwritten digit images with a size of 16×16 pixels.

(4). Fashion-MNIST [23]: This Fashion-MNIST dataset has the same number of images and the same image size as MNIST-full, but it is fairly more complicated. Instead of digits, it consists of various types of fashion products.

(5). REUTERS-10K: REUTERS-10K is a subset of REUTERS with 10000 samples. Four root categories (corporate/industrial, government/social, markets, and economics) are used as labels, and the tf-idf features on the 2000 most frequent words are computed.

(6). HAR: HAR is a time-series dataset consisting of 10299 sensor samples from a smartphone. It was collected from 30 people performing six activities: walking, walking upstairs, walking downstairs, sitting, standing, and laying.

(7). Pendigits: HAR is a dataset consisting of 10992 samples from pressure-sensitive plates with ten different digits, each represented by 8 coordinates of the stylus.

Table A1. Description of Datasets.

Dataset	Samples	Categories	Data Size
MNIST-full	70000	10	28×28×1
MNIST-test	10000	10	28×28×1
USPS	9298	10	16×16×1
Fashion-MNIST	70000	10	28×28×1
REUTERS-10K	10000	4	2000
HAR	10299	6	561
Pendigits	10992	10	16

A.2 Gradient Derivation

In the paper, we have emphasized time and again that $\{\mu_j\}_{j=1}^{C}$ is a set of ***learnable*** parameters, which means that we can optimize it while optimizing the network parameter θ_f. In Eq. (4) of the paper, we have presented the gradient of $L_{cluster}$ with respect to μ_j. In addition to $L_{cluster}$, both L_{rank} and L_{align} are involving μ_j. Hence, the detailed derivations for the gradient of L_{rank} and L_{align} with respect to μ_j are also provided. The gradient of L_{rank} with respect to each cluster center μ_j can be computed as:

$$\begin{aligned}\frac{\partial L_{rank}}{\partial \mu_j} &= \frac{\partial \sum_{i'=1}^{C}\sum_{j'=1}^{C}\left|d_Z\left(\mu_{i'},\mu_{j'}\right)-\kappa * d_X\left(v_{i'}^X, v_{j'}^X\right)\right|}{\partial \mu_j} \\ &= \sum_{i'=1}^{C}\sum_{j'=1}^{C}\frac{\partial\left|d_Z\left(\mu_{i'},\mu_{j'}\right)-\kappa * d_X\left(v_{i'}^X, v_{j}^X\right)\right|}{\partial \mu_j}\end{aligned}$$

The Euclidean metric is used for both the input space and the hidden layer space, i.e., $d_Z\left(\mu_{i'},\mu_{j'}\right)=\|\mu_{i'}-\mu_{j'}\|$. In addition, the symbols are somewhat abused for clear derivation, representing $\kappa * d_X\left(v_{i'}^X, v_{j'}^X\right)$ with K. Accordingly, the above equation can be further derived as follows:

$$\begin{aligned}
\frac{\partial L_{rank}}{\partial \mu_j} &= \sum_{i'=1}^{C}\sum_{j'=1}^{C}\frac{\partial\left|d_Z\left(\mu_{i'},\mu_{j'}\right)-\kappa * d_X\left(v_{i'}^X, v_{j'}^X\right)\right|}{\partial \mu_j} \\
&= \sum_{i'=1}^{C}\sum_{j'=1}^{C}\frac{\partial\left|\,\|\mu_{i'}-\mu_{j'}\|-K\right|}{\partial \mu_j} \\
&= \sum_{i'=1}^{C}\frac{\partial\left|\,\|\mu_{i'}-\mu_{j}\|-K\right|}{\partial \mu_j} + \sum_{j'=1}^{C}\frac{\partial\left|\,\|\mu_{j}-\mu_{j'}\|-K\right|}{\partial \mu_j} \\
&= \sum_{i'=1}^{C}\frac{\partial\left(\|\mu_{i'}-\mu_{j}\|-K\right)}{\partial \mu_j}\cdot\frac{\|\mu_{i'}-\mu_{j}\|-K}{\left|\,\|\mu_{i'}-\mu_{j}\|-K\right|} \\
&\quad + \sum_{j'=1}^{C}\frac{\partial\left(\|\mu_{j}-\mu_{j'}\|-K\right)}{\partial \mu_j}\cdot\frac{\|\mu_{j}-\mu_{j'}\|-K}{\left|\,\|\mu_{j}-\mu_{j'}\|-K\right|} \\
&= \sum_{i'=1}^{C}\frac{\partial\|\mu_{i'}-\mu_{j}\|}{\partial \mu_j}\cdot\frac{\|\mu_{i'}-\mu_{j}\|-K}{\left|\,\|\mu_{i'}-\mu_{j}\|-K\right|} \\
&\quad + \sum_{j'=1}^{C}\frac{\partial\|\mu_{j}-\mu_{j'}\|}{\partial \mu_j}\cdot\frac{\|\mu_{j}-\mu_{j'}\|-K}{\left|\,\|\mu_{j}-\mu_{j'}\|-K\right|} \\
&= \sum_{i'=1}^{C}\frac{\mu_j-\mu_{i'}}{\|\mu_{j}-\mu_{i'}\|}\cdot\frac{\|\mu_{j}-\mu_{i'}\|-K}{\left|\,\|\mu_{j}-\mu_{i'}\|-K\right|} \\
&\quad + \sum_{j'=1}^{C}\frac{\mu_j-\mu_{j'}}{\|\mu_{j}-\mu_{j'}\|}\cdot\frac{\|\mu_{j}-\mu_{j'}\|-K}{\left|\,\|\mu_{j}-\mu_{j'}\|-K\right|} \\
&= 2\sum_{i'=1}^{C}\frac{\mu_j-\mu_{i'}}{\|\mu_{j}-\mu_{i'}\|}\cdot\frac{\|\mu_{j}-\mu_{i'}\|-K}{\left|\,\|\mu_{j}-\mu_{i'}\|-K\right|} \\
&= 2\sum_{i'=1}^{C}\frac{\mu_j-\mu_{i'}}{\|\mu_{j}-\mu_{i'}\|}\cdot\frac{\|\mu_{j}-\mu_{i'}\|-\kappa * d_X\left(v_{i'}^X, v_{j}^X\right)}{\left|\,\|\mu_{j}-\mu_{i'}\|-\kappa * d_X\left(v_{i'}^X, v_{j}^X\right)\right|} \\
&= 2\sum_{i'=1}^{C}\frac{\mu_j-\mu_{i'}}{d_Z\left(\mu_j,\mu_{i'}\right)}\cdot\frac{d_Z\left(\mu_j,\mu_{i'}\right)-\kappa * d_X\left(v_{i'}^X, v_{j}^X\right)}{\left|d_Z\left(\mu_j,\mu_{i'}\right)-\kappa * d_X\left(v_{i'}^X, v_{j}^X\right)\right|}
\end{aligned}$$

The gradient of L_{align} with respect to each learnalbe cluster center μ_j can be computed as:

$$
\begin{aligned}
\frac{\partial L_{align}}{\partial \mu_j} &= \frac{\partial \sum_{j'=1}^{C} ||\mu_{j'} - v_{j'}^Z||}{\partial \mu_j} \\
&= \sum_{j'=1}^{C} \frac{\partial ||\mu_{j'} - v_{j'}^Z||}{\partial \mu_j} \\
&= \frac{\partial ||\mu_j - v_j^Z||}{\partial \mu_j} \\
&= \frac{\partial (\mu_j - v_j^Z)}{\partial \mu_j} \cdot \frac{\mu_j - v_j^Z}{\left\|\mu_j - v_j^Z\right\|} \\
&= \frac{\mu_j - v_j^Z}{\left\|\mu_j - v_j^Z\right\|}
\end{aligned}
$$

A.3 Definitions of Performance Metrics

The following notations are used for the definitions:

$d_X(i,j)$: the pairwise distance between x_i and x_j in input space X;

$d_Z(i,j)$: the pairwise distance between z_i and z_j in latent space Z;

$\mathcal{N}_i^{k,X}$: the set of indices to the k-nearest neighbor (kNN) of x_i in input space X;

$\mathcal{N}_i^{k,Z}$: the set of indices to the k-nearest neighbor (kNN) of z_i in latent space Z;

$r_X(i,j)$: the rank of the closeness (in Euclidean distance) of x_j to x_i in input space X;

$r_Z(i,j)$: the rank of the closeness (in Euclidean distance) of z_j to z_i in latent space Z.

The eight evaluation metrics are defined below:

(1) **ACC** (Accuracy) measures the accuracy of clustering:

$$
ACC = \max_m \frac{\sum_{i=1}^{N} 1\left\{l_i = m\left(s_i\right)\right\}}{N}
$$

where l_i and s_i are the true and predicted labels for data point x_i, respectively, and $m(\cdot)$ is all possible one-to-one mappings between clusters and label categories.

(2) **NMI** (Normalized Mutual Information) NMI calculates the normalized measure of similarity between two labels of the same data

$$
NMI = \frac{I(l;s)}{\max\{H(l), H(s)\}}
$$

where $I(l,s)$ is the mutual information between the real label l and predicted label s, and $H(\cdot)$ represents their entropy.

(3) **RRE** (Relative Rank Change) measures the average of changes in neighbor ranking between two spaces X and Z:

$$
RRE = \frac{1}{(k_2 - k_1 + 1)} \sum_{k=k_1}^{k_2} \left\{MR_{X\to Z}^k + MR_{Z\to X}^k\right\}
$$

where k_1 and k_2 are the lower and upper bounds of the k-NN.

$$
MR_{X\to Z}^k = \frac{1}{H_k} \sum_{i=1}^{N} \sum_{j\in\mathcal{N}_i^{k,Z}} \left(\frac{|r_X(i,j) - r_Z(i,j)|}{r_Z(i,j)}\right)
$$

$$
MR_{Z\to X}^k = \frac{1}{H_k} \sum_{i=1}^{N} \sum_{j\in\mathcal{N}_i^{k,X}} \left(\frac{|r_X(i,j) - r_Z(i,j)|}{r_X(i,j)}\right)
$$

where H_k is the normalizing term, defined as

$$
H_k = N \sum_{l=1}^{k} \frac{|N - 2l|}{l}
$$

(4) **Trust** (Trustworthiness) measures to what extent the k nearest neighbors of a point are preserved when going from the input space to the latent space:

$$
Trust = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \left\{1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{j\in\mathcal{N}_i^{k,Z}, j\notin\mathcal{N}_i^{k,X}} (r_X(i,j) - k)\right\}
$$

where k_1 and k_2 are the bounds of the number of nearest neighbors.

(5) **Cont** (Continuity) is defined analogously to $Trust$, but checks to what extent neighbors are preserved when going from the latent space to the input space:

$$
Cont = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \left\{1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{j\notin\mathcal{N}_i^{k,Z}, j\in\mathcal{N}_i^{k,X}} (r_Z(i,j) - k)\right\}
$$

Table A2. Parameter Sensitivity with different parameters k and κ on the MNIST-test dataset.

Parameters	ACC/NMI↑	RRE↓	Trust↑	Cont↑	d-RMSE↓	LGD↓	CRA↑
k=1, κ=3	**0.975/0.936**	0.0125	0.9944	0.9756	5.757	**0.8868**	1.00
k=3, κ=3	0.973/0.931	0.0114	0.9970	0.9757	5.805	0.9207	1.00
k=5, κ=3	0.972/0.930	0.0109	0.9981	0.9761	5.800	0.9339	1.00
k=8, κ=3	0.972/0.929	**0.0104**	0.9989	**0.9765**	5.810	0.9476	1.00
k=10, κ=3	0.972/0.929	0.0105	**0.9990**	0.9764	**5.704**	0.9487	1.00
k=5, κ=1	0.967/0.912	**0.0068**	**0.9993**	**0.9845**	**5.409**	**0.2524**	1.00
k=5, κ=3	**0.972/0.930**	0.0109	0.9981	0.9761	5.800	0.9339	1.00
k=5, κ=5	0.972/0.929	0.0146	0.9964	0.9691	15.0653	1.5719	1.00
k=5, κ=8	0.972/0.929	0.0190	0.9943	0.9615	29.4607	2.5410	1.00
k=5, κ=10	0.972/0.929	0.0195	0.9951	0.9597	37.7661	3.1434	1.00

where k_1 and k_2 are the bounds of the number of nearest neighbors.

(6) d-**RMSE** (Root Mean Square Error) measures to what extent the two distributions of **distances** coincide:

$$d-RMSE=\sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(d_X(i,j)-d_Z(i,j)\right)^2}$$

(7) **LGD** (Locally Geometric Distortion) measures how much corresponding distances between neighboring points differ in two metric spaces and is the primary metric for isometry, defined as:

$$LGD=\sum_{k=k_1}^{k_2}\sqrt{\sum_{i}^{M}\frac{\sum_{j\in\mathcal{N}_i^{k,(l)}}(d_l(i,j)-d_{l'}(i,j))^2}{\left(k_2-k_1+1\right)^2 M(\#\mathcal{N}_i)}}$$

where k_1 and k_2 are the lower and upper bounds of the k-NN.

(8) **CRA** (Cluster Rank Accuracy) measures the changes in *ranks* of cluster centers from the input space X and to the latent space Z:

$$CRA=\frac{\sum_{i=1}^{C}\sum_{j=1}^{C}\mathbf{1}(r_X(v_i^X,v_j^X)=r_Z(v_i^Z,v_j^Z))}{C^2}$$

where C is the number of clusters, v_j^X is the cluster center of the jth cluster in the input space X, v_j^Z is the cluster center of the jth cluster in the latent space Z, $r_X(v_i^X,v_j^X)$ denotes the rank of the closeness (in terms of Euclidean distance) of v_i^X to v_j^X in space X in the input space X, and $r_Z(v_i^Z,v_j^Z)$ denotes the rank of the closeness (in terms of Euclidean distance) of v_i^Z to v_j^Z in space Z.

A.4 Parameter Sensitivity

We also evaluated the sensitivity of parameters k and κ on the MNIST-test dataset and the results are shown in Tab. A2. It is found that the clustering performance is not sensitive to parameters k and κ, and some combinations of k and κ even produce better clustering performance than the metrics reported in the main paper. However, the effect of k and κ on multi-manifold learning is more pronounced, and different combinations of k and κ may increase or decrease performance. In general, this paper focuses on the design of algorithms and implementations, and no hyperparameter search is performed to find the best performance metrics.

A.5 Evaluation of Multi-Manifold Learning

GCML is compared with other methods on six manifold-related *quantitative metrics* to demonstrate its performance advantage for multi-manifold learning, and the complete results on seven datasets are shown on the left side of Tab. A3. In addition, the right side of Tab. A3 compares GCML with other methods on all datasets to see if these methods can really learn embeddings that are useful for downstream tasks. As shown in the table, GCML outperforms all the other methods with MLP, RFC, and LR as downstream tasks.

A.6 More Ablation Study Experiments

The results of the ablation study on the MNIST-full dataset have been presented in Tab. 4 in Sec 4.4. Here, we provide complete ablation study results on all datasets in Tab. A4. The conclusion is similar (note that the clustering performance of the model without clustering-oriented losses is very poorly, so the "best" metric numbers are not meaningful and are marked in gray color): (1) CL is very important for obtaining good clustering. (2) SL is beneficial for both clustering and multi-manifold learning. (3) The training strategies (WC and AT) contribute to improving performance metrics such as RRE, Trust, Cont, and CRA.

Table A3. Performance for multi-manifold learning (left) and downstream tasks (right) on all seven datasets.

Datasets	**Algorithms**	**RRE**$\downarrow$	**Trust**$\uparrow$	**Cont**$\uparrow$	d**-RMSE**$\downarrow$	**LGD**$\downarrow$	**CRA**$\uparrow$	**MLP**$\uparrow$	**RFC**$\uparrow$	**SVM**$\uparrow$	**LR**$\uparrow$
MNIST-full	DEC	0.09988	0.84499	0.94805	44.8535	4.37986	0.28	0.8647	0.8706	0.8707	0.8566
	IDEC	0.00984	0.99821	0.97936	24.5803	1.71484	0.33	0.9797	0.9737	0.9852	0.9650
	JULE	0.02657	0.93675	0.98321	28.3412	2.12955	0.27	0.9802	0.9825	0.9787	0.9743
	DSC	0.09785	0.87315	0.92508	6.98098	1.19886	0.23	0.9622	0.9501	0.9837	0.9752
	N2D	0.01002	0.99243	0.98466	5.7162	0.69946	0.21	0.9796	0.9803	0.9799	0.9792
	GCML (ours)	**0.00567**	**0.99978**	**0.98716**	**5.4986**	**0.69168**	**1.00**	**0.9851**	**0.9874**	**0.9869**	**0.9841**
MNIST-test	DEC	0.12800	0.81841	0.91767	14.6113	2.29499	0.19	0.8525	0.8605	0.8725	0.8685
	IDEC	0.01505	0.99403	0.97082	7.4599	1.08350	0.38	0.9740	0.9725	0.9845	0.9655
	JULE	0.04122	0.92971	0.97208	9.4768	1.17176	0.42	0.9775	0.9845	0.9800	0.9825
	DSC	0.10728	0.85498	0.92254	7.1689	1.19239	0.26	0.9535	0.9740	0.9825	0.9795
	N2D	0.01565	0.98764	0.97572	**5.0120**	0.97454	0.33	0.9715	0.9760	0.9725	0.9725
	GCML (ours)	**0.01090**	**0.99811**	**0.97612**	5.8000	**0.93394**	**1.00**	**0.9855**	**0.9875**	**0.9865**	**0.9855**
USPS	DEC	0.07911	0.88871	0.94628	16.4355	1.77848	0.31	0.8289	0.8668	0.8289	0.8294
	IDEC	0.01043	0.99726	0.97960	13.0573	1.11689	0.30	0.9482	0.9556	0.9656	0.9125
	JULE	0.02972	0.98763	0.98810	14.6324	1.43426	0.33	0.9576	0.9617	**0.9703**	0.9476
	DSC	0.06319	0.9151	0.93988	8.4412	1.02131	0.27	0.9351	0.9572	0.9612	0.9342
	N2D	0.01337	0.98769	0.98135	8.1961	0.54967	0.37	0.9569	0.9569	0.9569	0.9541
	GCML (ours)	**0.00577**	**0.99979**	**0.98701**	**6.4980**	**0.53180**	**1.00**	**0.9656**	**0.9651**	0.9604	**0.9551**
Fashion-MNIST	DEC	0.04787	0.93896	0.95450	39.3274	3.87731	0.37	0.6268	0.9853	0.6377	0.6245
	IDEC	0.01089	0.99683	0.97797	25.4024	1.91385	0.27	0.8367	0.9918	**0.8607**	0.7514
	JULE	0.03013	0.97732	0.97923	15.2213	1.43642	0.43	0.8541	0.9892	0.8566	0.7723
	DSC	0.05168	0.95013	0.96121	17.2201	1.42091	0.36	0.8084	0.9823	0.8618	0.7676
	N2D	0.00894	0.99062	0.98054	14.49079	**1.28180**	0.26	0.8412	0.9493	0.8230	0.7753
	GCML (ours)	**0.00836**	**0.99868**	**0.98203**	**13.3788**	1.33893	**1.00**	**0.8642**	**0.9942**	0.8468	**0.7768**
REUTERS-10K	DEC	0.26192	0.65518	0.80477	40.4671	4.00423	0.63	0.7985	0.7880	0.8105	0.7450
	IDEC	0.05981	0.95840	0.90550	43.9556	2.01365	0.75	0.9225	0.8930	0.9280	0.7705
	JULE	-	-	-	-	-	-	-	-	-	-
	DSC	-	-	-	-	-	-	-	-	-	-
	N2D	0.03827	0.97385	0.93412	36.1042	**1.69013**	0.31	0.9205	0.9080	0.9240	0.8335
	GCML (ours)	**0.03206**	**0.98380**	**0.93802**	**34.5478**	2.72096	**1.00**	**0.9360**	**0.9185**	**0.9390**	**0.8475**
HAR	DEC	0.09060	0.89097	0.91766	10.0222	1.58691	0.30	0.7696	0.7847	0.7628	0.7634
	IDEC	0.01031	0.99433	0.98132	9.9155	0.93736	0.39	0.8973	0.9031	0.9041	0.8822
	JULE	-	-	-	-	-	-	-	-	-	-
	DSC	-	-	-	-	-	-	-	-	-	-
	N2D	0.00841	0.99281	0.97695	**8.2326**	0.64296	0.33	0.9138	0.9083	0.9174	0.8799
	GCML (ours)	**0.00665**	**0.99895**	**0.98634**	15.2876	**0.46189**	**1.00**	**0.9235**	**0.9193**	**0.9293**	**0.8996**
pendigits	DEC	0.03932	0.95149	0.96300	21.6608	1.32970	0.24	0.7215	0.8432	0.7376	0.7429
	IDEC	0.00384	0.99879	0.99255	19.7243	0.86137	0.28	0.8870	0.9461	**0.9595**	0.8636
	JULE	-	-	-	-	-	-	-	-	-	-
	DSC	-	-	-	-	-	-	-	-	-	-
	N2D	0.00262	0.99919	0.99473	20.7052	0.76941	0.42	0.9131	0.9669	0.9528	0.8516
	GCML (ours)	**0.00091**	**0.99994**	**0.99808**	**2.5184**	**0.08223**	**1.00**	**0.9538**	**0.9705**	0.9532	**0.8953**

Table A4. Ablation study of loss items and training strategies used in the proposed GCML framework.

Datasets	Methods	ACC/NMI↑	RRE↓	Trust↑	Cont↑	d-RMSE↓	LGD↓	CRA↑
MNIST-full	w/o SL	0.976/0.939	0.0093	0.9967	0.9816	24.589	1.6747	0.32
	w/o CL	0.814/0.736	0.0004	0.9998	0.9990	7.458	0.0487	1.00
	w/o WC	0.977/0.943	0.0065	0.9987	0.9860	5.576	0.6968	0.98
	w/o AT	0.978/0.944	0.0069	0.9986	0.9851	5.617	0.7037	0.96
	full model	**0.980/0.946**	**0.0056**	**0.9997**	**0.9871**	**5.498**	**0.6916**	**1.00**
MNIST-test	w/o SL	**0.973/0.932**	0.0146	0.9928	0.9727	7.701	1.0578	0.31
	w/o CL	0.773/0.747	0.0020	0.9994	0.9954	7.229	0.0809	1.00
	w/o WC	0.956/0.904	0.0132	0.9955	0.9735	**5.470**	0.9364	**1.00**
	w/o AT	0.970/0.929	0.0118	0.9974	0.9747	5.567	0.9404	**1.00**
	full model	0.972/0.930	**0.0109**	**0.9981**	**0.9761**	5.800	**0.9339**	**1.00**
USPS	w/o SL	0.957/0.902	0.0095	0.9967	0.9812	14.609	0.9847	0.29
	w/o CL	0.664/0.658	0.0020	0.9996	0.9952	2.934	0.0687	1.0
	w/o WC	0.956/0.896	0.0060	0.9991	0.9868	6.572	0.5335	**1.00**
	w/o AT	0.947/0.885	0.0080	0.9979	0.9833	**5.960**	**0.4967**	**1.00**
	full model	**0.958/0.902**	**0.0057**	**0.9997**	**0.9870**	6.498	0.5318	**1.00**
Fashion-MNIST	w/o SL	0.706/0.682	0.0108	0.9964	0.9781	25.954	1.8936	0.30
	w/o CL	0.576/0.569	0.0004	0.9994	0.9995	7.654	0.0523	1.00
	w/o WC	0.702/0.695	0.0084	0.9972	0.9814	**13.238**	1.3474	**1.00**
	w/o AT	0.708/0.694	0.0097	0.9975	0.9798	13.354	1.3611	**1.00**
	full model	**0.710/0.685**	**0.0083**	**0.9986**	**0.9820**	13.378	**1.3389**	**1.00**
REUTERS-10K	w/o SL	0.819/0.564	0.0529	0.9610	0.9185	44.481	1.9090	0.38
	w/o CL	0.542/0.279	0.0277	0.9868	0.9456	37.018	2.2294	1.00
	w/o WC	0.830/0.583	0.0420	0.9667	0.9361	35.302	2.8286	**1.00**
	w/o AT	0.825/0.563	0.0440	0.9650	0.9330	39.275	2.9146	**1.00**
	full model	**0.836/0.590**	**0.0320**	**0.9838**	**0.9380**	**34.547**	**2.7209**	**1.00**
HAR	w/o SL	0.835/0.746	0.0116	0.9944	0.9792	**8.168**	0.8882	0.33
	w/o CL	0.744/0.615	0.0024	0.9986	0.9948	15.060	0.2193	1.00
	w/o WC	0.786/0.701	0.0130	0.9950	0.9756	15.398	0.6171	**1.00**
	w/o AT	0.834/0.745	0.0089	0.9965	0.9835	15.726	0.4734	**1.00**
	full model	**0.844/0.862**	**0.0066**	**0.9989**	**0.9863**	15.287	**0.4618**	**1.00**
pendigits	w/o SL	0.843/0.803	0.0027	0.9989	0.9948	20.692	0.7587	0.31
	w/o CL	0.773/0.705	0.0004	0.9998	0.9992	2.177	0.0385	1.00
	w/o WC	0.840/0.790	0.0014	0.9995	0.9972	2.690	0.1124	**1.00**
	w/o AT	0.842/0.804	0.0018	0.9991	0.9965	2.756	0.1148	**1.00**
	full model	**0.855/0.814**	**0.0009**	**0.9999**	**0.9981**	**2.518**	**0.0822**	**1.00**