WHERE **INNOVATION** BEGINS

DESIGN AUTOMATION CONFERENCE

JULY 9–13, 2023

MOSCONE WEST CENTER
SAN FRANCISCO, CA, USA

WHERE **INNOVATION** BEGINS

DESIGN AUTOMATION CONFERENCE

DistHD: A Learner-Aware Dynamic Encoding Method for Hyperdimensional Classification

Junyao Wang, Sitao Huang, Mohsen Imani

University of California, Irvine

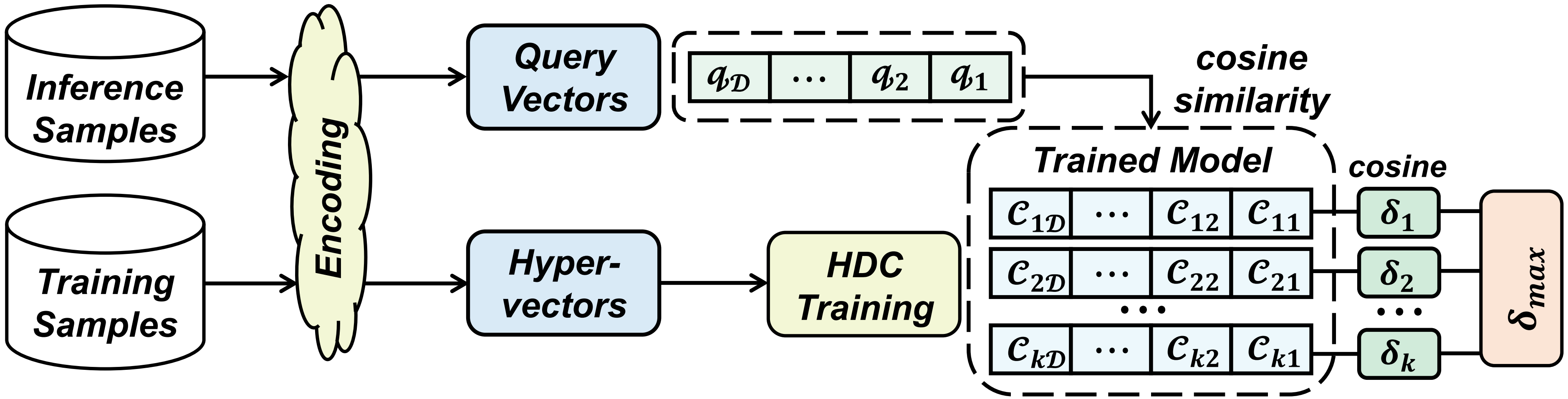
Overview

- Brief Intro of Hyperdimensional Computing
- A Major Challenge
- Motivation
- Workflow Overview
- Experimental Results

Hyperdimensional Computing (HDC)

- HDC: An effective solution for learning on edge
  - Inspired by the neuroscience observation that human cerebellum cortex efficiently and effortlessly processing memory, perception, and cognition information
- Compared to state-of-the-art learning algorithms:
  - In various applications, HDC provides comparable high-quality learning performance
  - Notably faster convergence and higher efficiency

Hyperdimensional Computing



Cerebellum

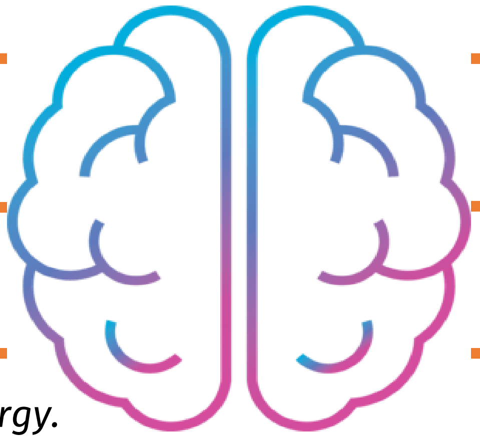
Cerebellum works with sparse high-dimensional representations.

Robustness

Brains can work with multiple noisy inputs.

Efficient

Brains work at around as low as 20W of energy.



High-dimensional

Basic elements are hypervectors.

Holographic Encoding

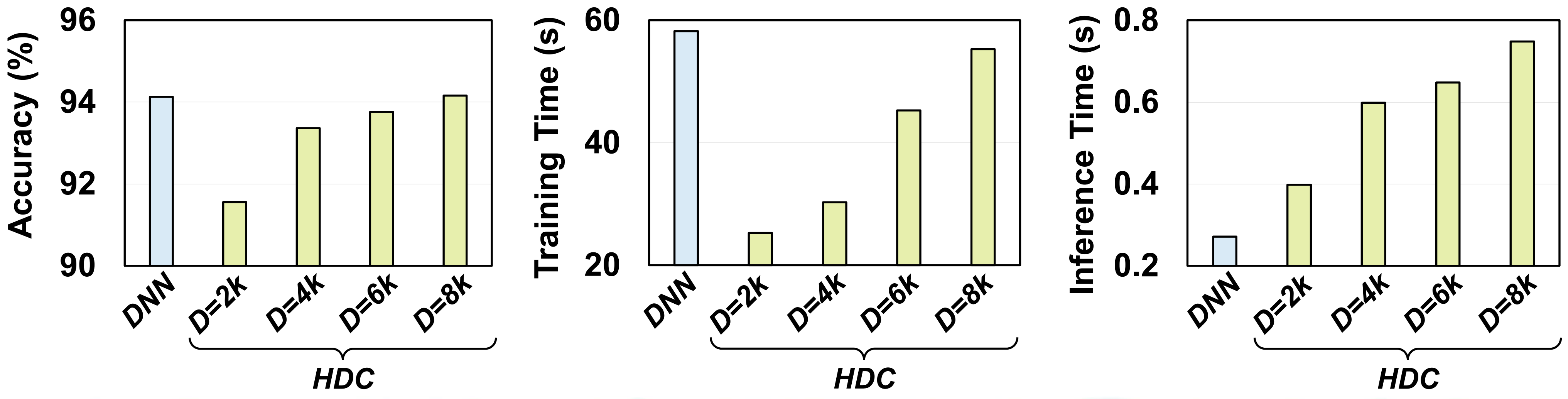
Information of every feature is found on all the dimensions of the hypervectors.

Well-trackable Algebra

Well-defined and highly-parallel operations.

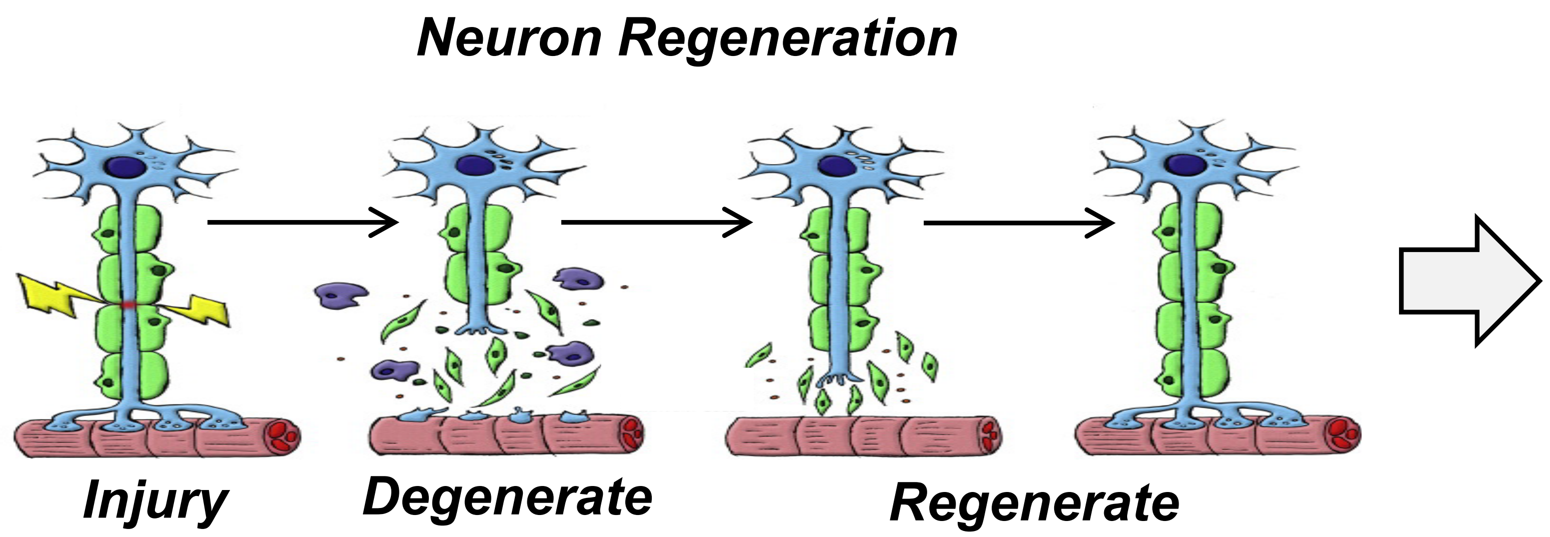
A Major Challenge

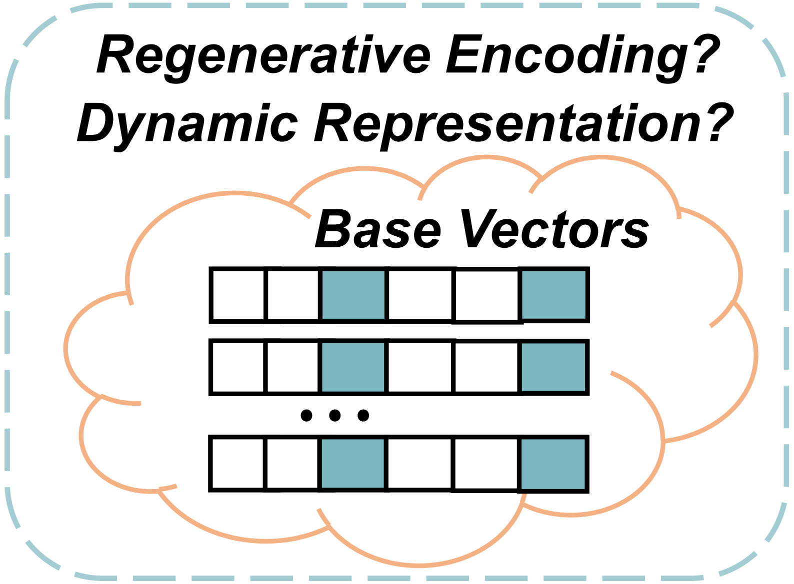
- *Static Encoding*
  - Existing encoding techniques lacks the capability to utilize and adapt to information learned during the learning process – *They are never updated during training.*
  - Very high dimensionality is required to achieve reasonable accuracy
  - Impedes the *feasibility* and *scalability* of HDC on resource-constrained computing platforms



DistHD: Motivation (1)

- What about human brains?
  - So much easier!
  - Neurons in human brains *dynamically change and regenerate* all the time and provide useful functionality when they learn new information.

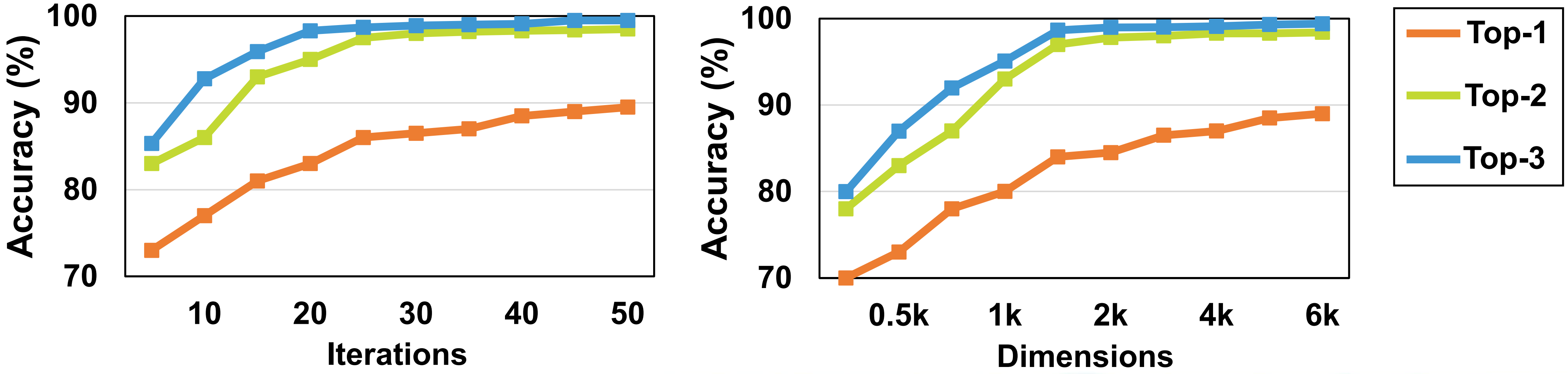




DistHD: Motivation (2)

- An interesting observation:
  - Define a top k-classification for a given data point as *correct* if the true label is one of the k most similar classes selected.
  - Accuracy (top-2) $\gg$ Accuracy (top-1)
  - Accuracy(top-3) – Accuracy(top-2) $\ll$ Accuracy(top-2) – Accuracy(top-1)

Top-1, Top-2, Top-3 Classification of State-of-the-art HDCs



DistHD: Workflow Overview

- Step 1: Top-2 Classification
  - Adaptive learning
  - Selecting top-2 labels
- Step 2: Dimension Regeneration
  - Identifying undesired dimensions
  - Regenerate undesired dimensions for enhanced learning quality

DistHD is open-sourced: https://github.com/jwang235/DistHD

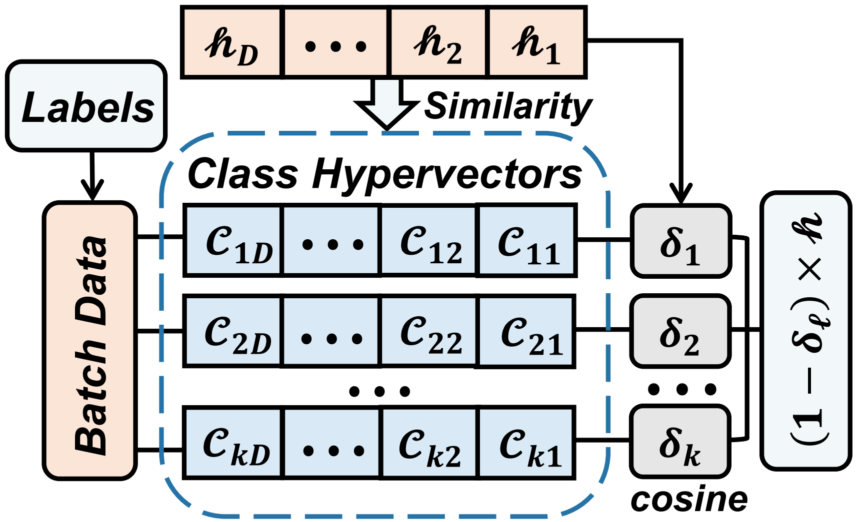
DistHD: Top-2 Classification

- Adaptive Learning*
  - Bundle encoded data points based on how much new information is added to class hypervectors
  - If an encoded point $\mathcal{H}$ has the highest cosine similarity with class $\mathcal{C}_i$ while it is actually in class $\mathcal{C}_j$, we update class hypervectors as:

$$\mathcal{C}_i \leftarrow \mathcal{C}_i + \eta \cdot [1 - \delta(\mathcal{H}_i, \mathcal{C}_i)] \times \mathcal{H}_i$$

$$\mathcal{C}_j \leftarrow \mathcal{C}_j - \eta \cdot \left[1 - \delta\left(\mathcal{H}_i, \mathcal{C}_j\right)\right] \times \mathcal{H}_i$$

  - A large δ_l $(1 - \delta_l \approx 0)$ indicates the input data point already exist in the model or marginally mismatched, a small δ_l $(1 - \delta_l \approx 1)$ indicate a noticeably new pattern.



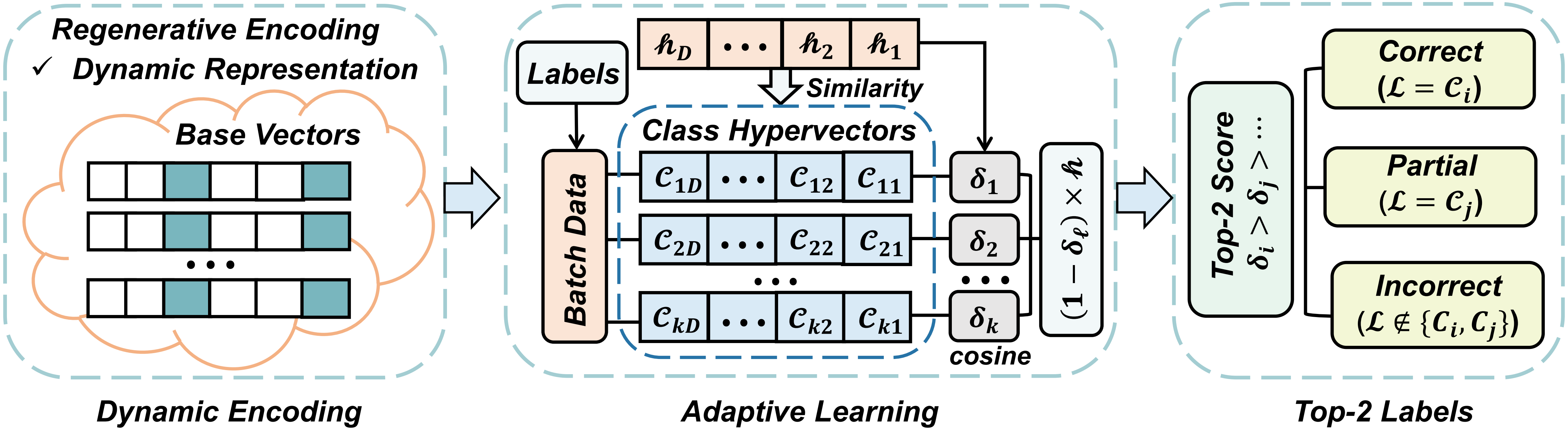
Adaptive Learning

- ✓ *Highly parallel matrix operations*
- ✓ *Capture uncommon patterns and reduce model saturation*

**Alejandro Hernández-Cano et al, "OnlineHD: Robust, Efficient, and Single-Pass Online Learning Using Hyperdimensional System", DATE, 2021.*

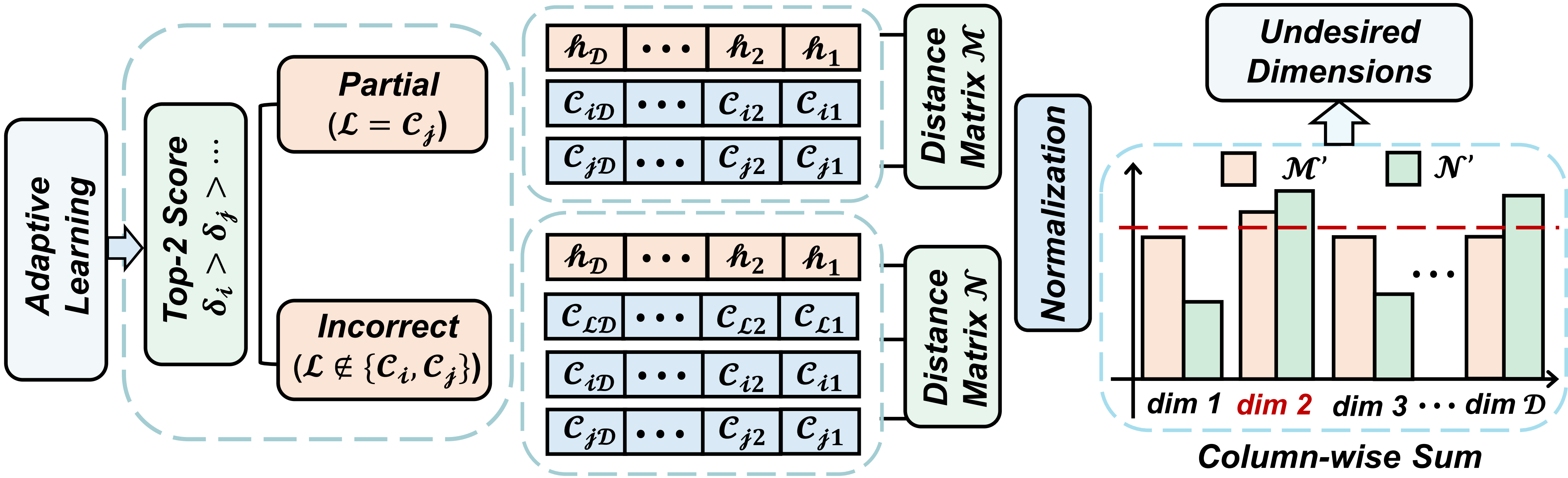
DistHD: Top-2 Labels

- Selecting top-2 labels
  - In each training iteration, we utilize the partially trained model to select the top two most similar classes for each training sample
  - Classify the result for each training sample as: *correct, partially correct, incorrect*.

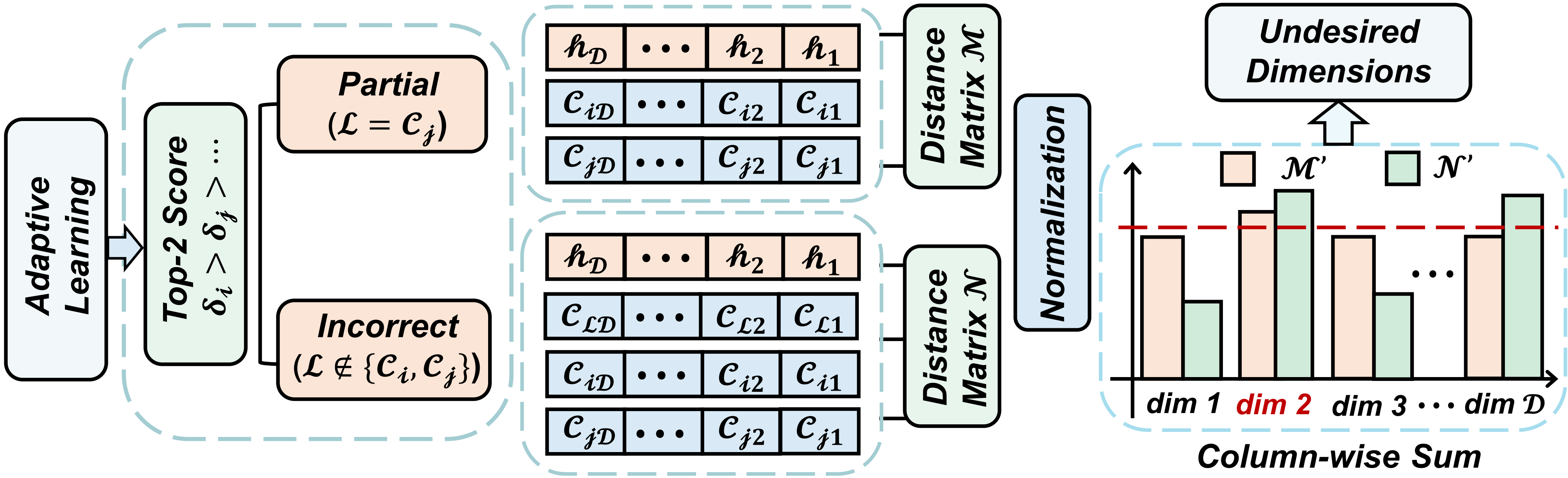


DistHD: Identifying Undesired Dimensions

- Dimensions that mislead the classification?
  - Our strategy is simple: *searching for dimensions that farthest away from the correct class hypervectors and closest to the incorrect class hypervectors.*



DistHD: Identifying Undesired Dimensions



- For samples classified as partial, we search for:

$$\arg\max_d \left|h_d - \mathcal{C}_{jd}\right| \cap \arg\min_d \left|h_d - \mathcal{C}_{id}\right| \; (1 \le d \le \mathcal{D}) \Rightarrow \mathcal{M}_t = \alpha \cdot \left|h - \mathcal{C}_j\right| - \beta \cdot \left|h - \mathcal{C}_i\right|$$

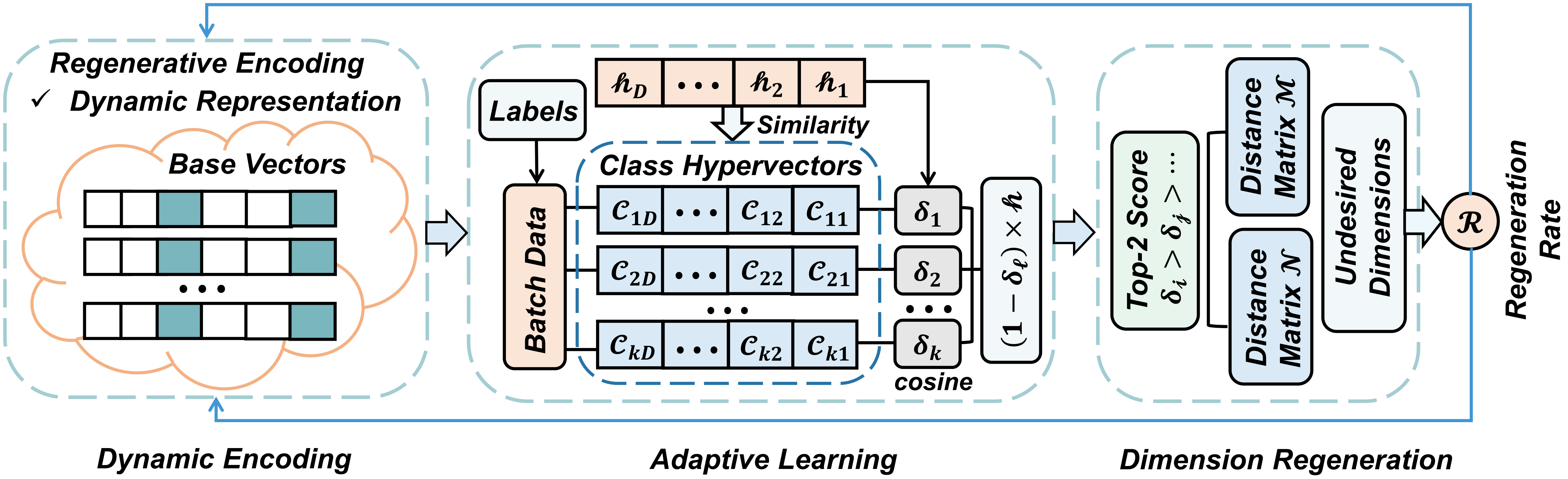
- Similarly, for samples classified as incorrect,

$$\mathcal{N}_t = \alpha \cdot \left|h - \mathcal{C}_\mathcal{L}\right| - \beta \cdot \left|h - \mathcal{C}_i\right| - \theta \cdot \left|h - \mathcal{C}_j\right|$$

$$\mathcal{M}' = sum(\mathcal{M}_t), \mathcal{N}' = sum(\mathcal{N}_t)$$

DistHD: Dimension Regeneration

- Once we find the dimensions mislead the classification (i.e., hurt the accuracy), we replace them with another randomly generated hypervector.



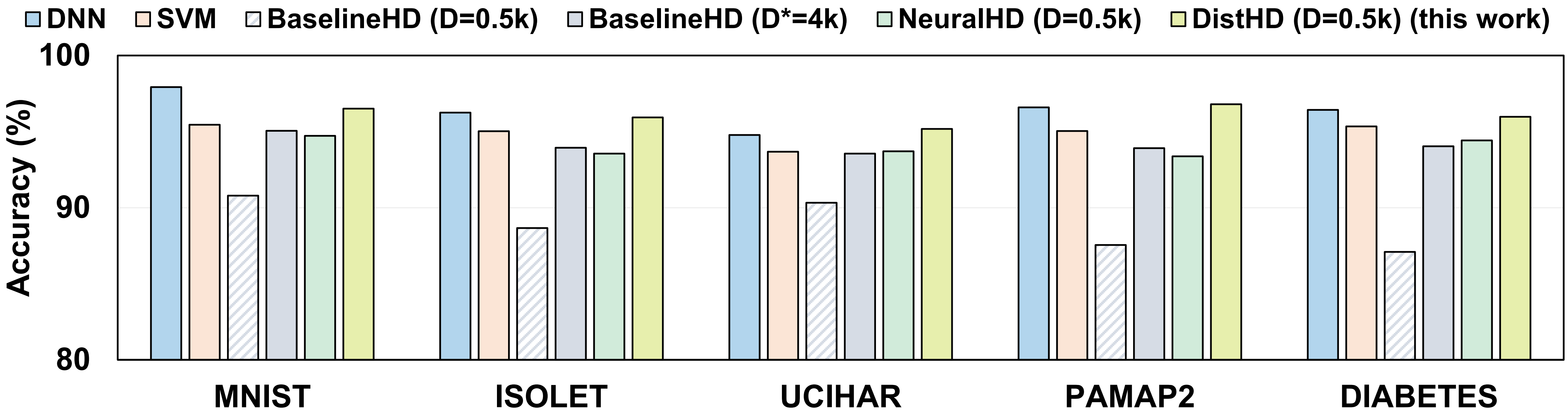
Evaluation: Setup

- We compare DistHD with the following learning algorithms:
  - SOTA DNN with comparable inference latency
  - SVM
  - SOTA HDC without dynamic encoding (BaselineHD)
  - NeuralHD* (the first dynamic HDC framework aiming at improving learning efficiency by reducing dimensionalities)

**Zhuowen Zou et al, "Scalable edge-based hyperdimensional learning system with brain-inspired neural adaptation", The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), 2021*

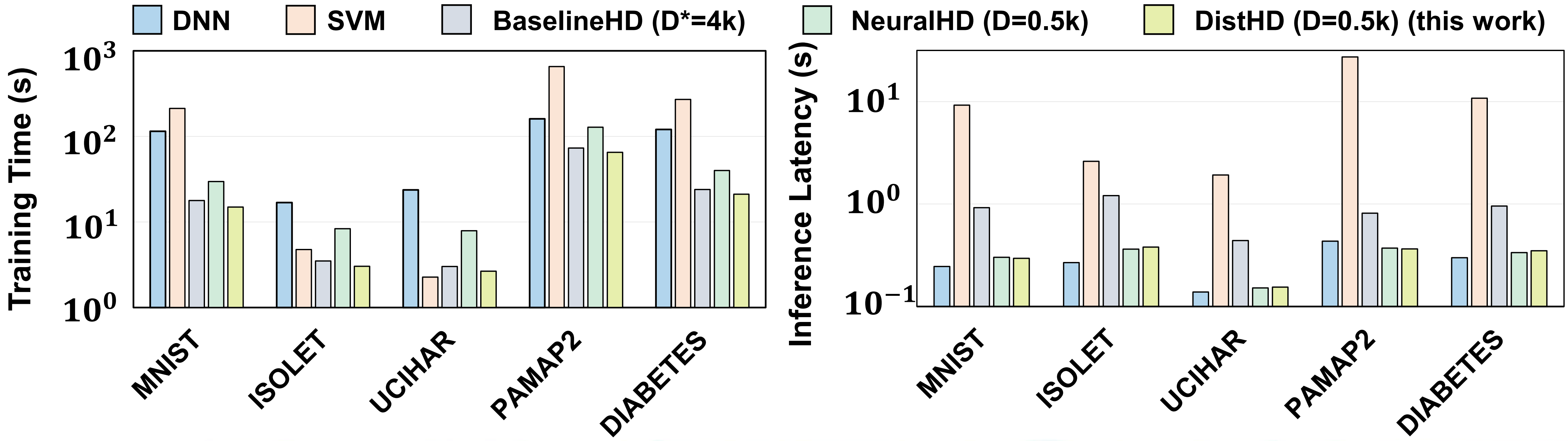
Evaluation: Accuracy

- ✓ Comparable accuracy to SOTA DNN, 1.17% higher accuracy than SVM
- ✓ 6.96 % and 1.82% higher accuracy than BaselineHD (*D=0.5k*) and BaselineHD (*D*=4k*), respectively
- ✓ 1.89% higher accuracy than NeuralHD (*D=0.5k*)



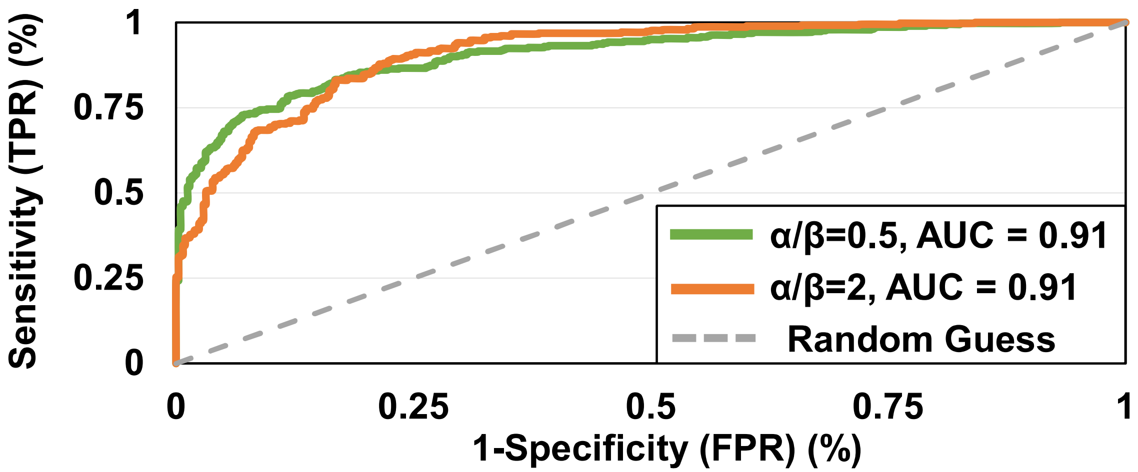
Evaluation: Efficiency

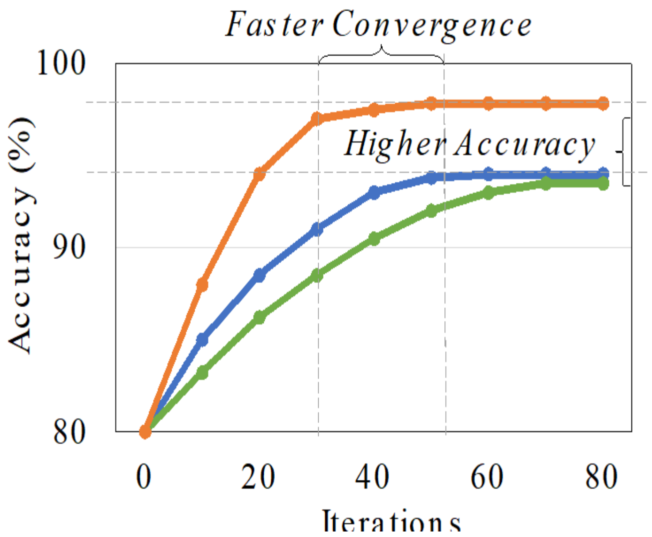
- Training:
  - ✓ 5.97× faster than SOTA DNN
  - ✓ 1.15× faster than BaselineHD, 2.32× faster than NeuralHD
- Inference:
  - ✓ 8.09× faster than BaselineHD

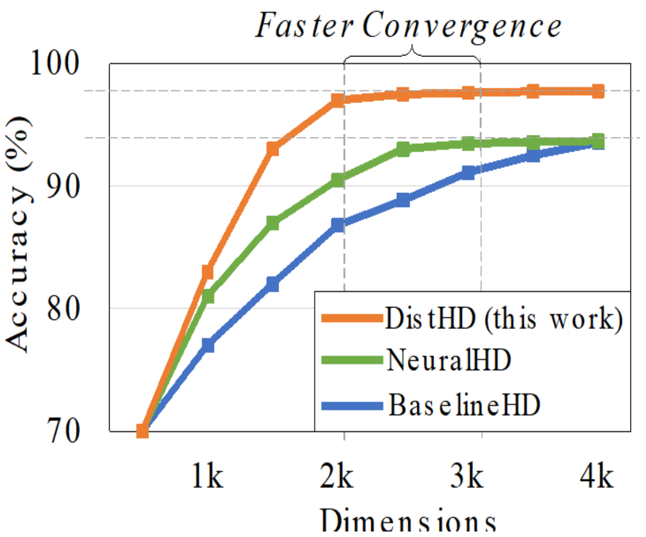


Other Evaluation Results

- ROC Curve with Different Hyperparameters
- Faster Convergence Speed of DistHD Compared to other HDC algorithms
- Higher Robustness Against Noise







Hardware Error			***1.0%***	***2.0%***	***5.0%***	***10.0%***	***15.0%***
DNN			3.9%	10.7%	17.8%	32.1%	41.2%
DistHD	***1 bit***	***0.5k***	1.1%	1.7%	3.6%	5.4%	7.2%
		1k	0.7%	1.3%	2.8%	4.2%	6.4%
		2k	0.4%	0.7%	1.3%	3.7%	5.9%
		4k	0.0%	0.0%	1.0%	3.1%	4.1%
	2 bits	***0.5k***	1.9%	2.3%	4.5%	7.9%	10.4%
		1k	1.2%	1.7%	3.7%	6.8%	9.9%
		2k	0.5%	1.1%	2.5%	5.9%	8.7%
		4k	0.0%	0.5%	1.6%	4.8%	8.0%
	4 bits	***0.5k***	2.3%	4.7%	8.4%	13.1%	17.3%
		1k	1.6%	3.2%	6.9%	12.7%	15.9%
		2k	0.9%	2.1%	4.7%	10.2%	13.7%
		4k	0.2%	1.0%	2.9%	7.4%	11.7%
	8 bits	***0.5k***	3.6%	7.9%	13.7%	18.3%	22.9%
		1k	2.7%	6.1%	10.8%	15.7%	20.1%
		2k	1.9%	4.9%	8.1%	14.1%	19.8%
		4k	1.4%	3.6%	5.1%	12.8%	17.6%

Conclusion

- **Our work:**
  - An accurate, efficient, and robust HDC learning framework
  - With a powerful dynamic encoding module, DistHD identifies and regenerates dimensions that mislead the classification and reduce the learning accuracy
- **Future work:**
  - Better dynamic encoding /regenerating technique
  - Customized hardware acceleration on resource-constrained platforms
  - Broader Applications & More complicated dataset

Thank you for your attention.

DistHD is open-sourced: https://github.com/jwang235/DistHD

Paper Available at: https://sites.uci.edu/junyaowang/