IV. Experimental Results

A. Experimental Setup

We describe the functionality of the proposed HDNA using a Python implementation. We compare the power consumption and execution time of the HDNA architectures running on traditional CPU cores. We used an Intel core i7 7600 processor with 16 GB memory (4-core, 2.8GHz) to test different designs. Power consumption is measured by Hioki 3334 power meter. To estimate the cost of digital design, we also use a standard cell-based flow to design dedicated hardware for HDNA. We describe the proposed designs using RTL System-Verilog. For the synthesis, we use *Synopsys Design Compiler* with the TSMC 45 nm technology library, the general purpose process with high V_{TH} cells. We measured the power consumption of HD designs using *Synopsys PrimeTime* at (1 V, 25°C, TT) corner.

To assess the efficiency of proposed design, we apply the application of HDNA over two popular DNA classification datasets, *Empirical* [11] and *Molecular Biology* [12] datasets. Both datasets are split into two parts: 80% per species for training and 20% for testing.

B. HDNA Accuracy

We compare the classification accuracy of of HDNA and the state-of-the-art classification techniques over Empirical and Molecular biology dataset [13], listed in Table I and Table II respectively. HDNA using Encoder I ad Encoder II can achieve at least 5.21% and 4.87% higher classification accuracy as compared to prior techniques. For molecular biology dataset, our evaluation shows that HDNA using Encoder I can achieve comparable accuracy as other classification techniques while Encoder II can provide 100% classification accuracy. This accuracy is 5.87% higher than other classification algorithms.

In addition, we compare the efficiency of HDNA designs with SVM and *K*-NN designs. We run all algorithms implemented in python code on CPU over Empirical and Molecular biology datasets. Table I and Table II show the average energy consumption and execution times of different designs when a query runs on CPU cores. All algorithms are written to provide the maximum parallelism. Comparing HDNA design with prior work shows that HDNA using Encoder I (Encoder II) can achieve at least 2.98× (4.32×) speedup and 3.26× (2.05×) energy efficiency improvement over empirical dataset. Similarly, over molecular biology dataset, Encoder I (Encoder II) provides at least 4.38× (5.44%) speedup and 4.34× (2.47×) energy efficiency improvement as compare to other classification techniques. As traditional cores have not been designed to work with long hypervectors, we expect HDNA provides much more efficiency when it implements on digital RTL design. The following sections show the efficiency of HDNA design over digital implementation.

TABLE I
Accuracy and Efficiency of SVM, Bayes and the proposed HDNA over Empirical dataset (Encoder I with $n = 12$).

	Classes	SVM	Bayes	Encoder I	Encoder II
Accuracy	*Cypraeidae*	94.3%	93.2%	100%	100%
	Drosophila	98.3%	96.5%	100%	100%
	Inga	89.8%	91.5%	100%	100%
	Bats	100.0%	100.0	98.2%	100%
	Fishes	95.5%	97.3%	100%	95.2%
	Birds	98.4%	94.3%	99.7%	100%
	Fungi	80.0%	70.0%	100%	100%
	Algae	100.0%	100.0%	100%	100%
	Average	**94.53%**	**92.85%**	**99.74%**	**99.40%**
Energy Consumption (mJ)		62.03	47.51	14.53	23.16
Execution Time (ms)		2.77	1.73	0.58	0.44

TABLE II
Accuracy and Efficiency of *K*-NN, KBANN and HDNA over molecular biology dataset (Encoder I with $n = 10$)

	Classes	*K*-NN	KBANN	Encoder I	Encoder II
Accuracy	*Exon/Intron*	94.3%	93.2%	100%	96.7%
	Intron/Exon	98.3%	96.5%	100%	91.5%
	Neither	89.8%	91.5%	100%	92.15%
	Average	**94.13%**	**93.7%**	**100%**	**93.4%**
Energy Consumption (mJ)		46.60	42.56	9.79	17.21
Execution Time (ms)		2.07	1.36	0.31	0.25

Acknowledgements

This work was partially supported by CRISP, one of six centers in JUMP, an SRC program sponsored by DARPA, and also NSF grants #1730158 and #1527034.

DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited.

References

[1] A. McKenna *et al.*, "The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data," *Genome research*, vol. 20, no. 9, pp. 1297–1303, 2010.

[2] H. Erlich, *PCR technology: principles and applications for DNA amplification*. Springer, 2015.

[3] R. C. Green *et al.*, "Acmg recommendations for reporting of incidental findings in clinical exome and genome sequencing," *Genetics in medicine: official journal of the American College of Medical Genetics*, vol. 15, no. 7, p. 565, 2013.

[4] M. Imani, T. Nassar, A. Rahimi, and T. Rosing, "Hdna: Energy-efficient dna sequencing using hyperdimensional computing," in *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on*, pp. 271–274, IEEE, 2018.

[5] P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cognitive Computation*, vol. 1, no. 2, pp. 139–159, 2009.

[6] P. Kanerva *et al.*, "Random indexing of text samples for latent semantic analysis," in *Proceedings of the 22nd annual conference of the cognitive science society*, vol. 1036, Citeseer, 2000.

[7] M. Imani *et al.*, "Voicehd: Hyperdimensional computing for efficient speech recognition," in *IEEE International Conference on Rebooting Computing (ICRC)*, IEEE, 2017.

[8] M. Imani, C. Huang, D. Kong, and T. Rosing, "Hierarchical hyperdimensional computing for energy efficient classification," in *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*, pp. 1–6, IEEE, 2018.

[9] M. Imani *et al.*, "Exploring hyperdimensional associative memory," in *High Performance Computer Architecture (HPCA), 2017 IEEE International Symposium on*, pp. 445–456, IEEE, 2017.

[10] M. Imani *et al.*, "Low-power sparse hyperdimensional encoder for language recognition," *IEEE Design & Test*, vol. 34, no. 6, pp. 94–101, 2017.

[11] "Empirical datasets:." http://dmb.iasi.cnr.it/supbarcodes.php.

[12] "Molecular Biology datasets." https://archive.ics.uci.edu/ml/datasets/Molecular+Biology+(Splice-junction+Gene+Sequences).

[13] E. Weitschek *et al.*, "Supervised dna barcodes species classification: analysis, comparisons and results," *BioData mining*, vol. 7, no. 1, p. 4, 2014.