Supplementary Materials

Experimental Setup

Fig. S1 shows the experimental setup used in this work for the demonstration of FeFET analog CAM cell and array for the decision tree and deep random forest.

Figure S1: The setup with the NI PXI system with PPMU and SMU modules (1), the switch matrix board (2), the 300mm wafer probe station (3) and the probe card connecting the wafer with the array structure (4).
**Multiple V\textsubscript{TH} States in FeFET**

Figure S2: Multiple states in FeFET [23]. (a) Switching dynamics in FeFET showing the memory window as a function of write pulse width at different write pulse amplitudes. FeFET is initialized with +4V, 1\textmu s write pulse before each measurement. The write pulse amplitudes changed from -1V to -3.8V with a step of -0.1V. Intermediate V\textsubscript{TH} states are observed. (b) I\textsubscript{D}-V\textsubscript{G} characteristics for four different states in 60 different FeFET devices. (c)/(d) The V\textsubscript{TH} distributions for 4/8 levels, respectively [23]. Different levels are written through a write pulse with different amplitudes. Tight V\textsubscript{TH} distribution is obtained given the present unoptimized FeFET devices.
Measurement On An 1×16 CAM Array: Other Cells In State 4

15 cells are in state 4, $V_{TH}=1.1V$

Figure S3: Measurement on a 1×16 analog CAM array. During testing, all $F_1$ transistors in the array are set to be high-$V_{TH}$ states, fixing the lower bounds. $F_0$ transistors in 15 cells are configured into the state $S_4$, i.e., $V_{TH}=1.1V$, and the rest target cell is configured to the four $V_{TH}$ states (black, red, blue, and green curves in each figure). Given this configuration, the target cell $V_{SL}$ is swept when the other cells are searched with 16 different $V_{SL}$ values from -0.3V to 1.2V in the step of 0.1V. Assume the current threshold is at $10^{-7}$A (red dashed line), then the matching range can be successfully realized when the other cells are searched with $V_{SL}$ below $V_{TH}$ (i.e., $\leq 1.0V$) and can be varied on the target cell $V_{SL}$ dimension.
Measurement On An 1×16 CAM Array: Other Cells In State 3

15 cells are in state 3, $V_{TH}=0.8V$

Figure S4: Measurement on a 1×16 analog CAM array. Similar to Fig S3, all the $F_1$ transistors in the array are set to high-$V_{TH}$ states. $F_0$ transistors in 15 cells are configured into the state $S_3$, i.e., $V_{TH}=0.8V$, and the target cell is configured to the four $V_{TH}$ states (black, red, blue, and green curves in each figure). After the cell configuration, the target cell $V_{SL}$ is swept when the other cells are searched with 16 different $V_{SL}$ values from -0.3V to 1.2V in the step of 0.1V. With the current threshold at $10^{-7}A$ (red dashed line), the matching range can be now reduced accordingly to $V_{SL} \leq 0.7V$ for other cells, and can be varied on the target cell $V_{SL}$ dimension.
Measurement On An 1×16 CAM Array: Other Cells In State 2

15 cells are in state 2, $V_{TH}=0.4V$

Figure S5: Measurement on a 1×16 analog CAM array. Similar to Fig S3, all the $F_1$ transistors in the array are set to high-$V_{TH}$ states. $F_0$ transistors in 15 cells are configured into the state $S_2$, i.e., $V_{TH}=0.4V$, and the target cell is configured to the four $V_{TH}$ states (black, red, blue, and green curves in each figure). After the cell configuration, the target cell $V_{SL}$ is swept when the other cells are searched with 16 different $V_{SL}$ values from -0.3V to 1.2V in steps of 0.1V. With the current threshold set at $10^{-7}$A (red dashed line), the matching range can now be reduced accordingly to $V_{SL} \leq 0.3V$ for other cells, and can be varied on the target cell $V_{SL}$ dimension.
Measurement on An 1×16 CAM Array: Other Cells In State 1

15 cells are in state 1, $V_{TH}$=0V

Figure S6: Measurement on a 1×16 analog CAM array. Similar to Fig S3, all the $F_1$ transistors in the array are set to be high-$V_{TH}$ states. $F_0$ transistors in 15 cells are configured into the state $S_3$, i.e., $V_{TH}$=0V, and the target cell is configured to the four $V_{TH}$ states (black, red, blue, and green curves in each figure). After the cell configuration, the target cell $V_{SL}$ is swept when the other cells are searched with 16 different $V_{SL}$ values from -0.3V to 1.2V in the step of 0.1V. With the current threshold set at $10^{-7}$ A (red dashed line), the matching range can now be reduced accordingly to $V_{SL} \leq -0.1V$ for other cells, and can be varied on the target cell $V_{SL}$ dimension.
**Experimental Demonstration of an Example DRF**

An example DRF has been constructed and its operation is verified in experiment, as shown in Fig S7. The DRF consists of 3 layers, and for the ease of experimental demonstration, a single tree is considered for each layer (Fig S7(a)). The decision tree output vector is concatenated with part of input features and sent to the next layer. Fig S7(b)/(c) show the theoretical/experimental results of the decision tree outputs in each layer. The experimental results match the theoretical values, indicating the successful operation of the DRF.

![Diagram of a DRF](image)

**Figure S7:** Experimental demonstration of (a) an example 3-layer deep random forest. For the ease of experimental verification, a simple decision tree is used in each layer and the decision is concatenated with part of input features as input for the next layer. (b) The theoretical decision values for each layer. (c) The measured experimental match line currents.
SPICE Simulation Setup of Voltage Domain Sensing

Figure S8: SPICE simulation setup for voltage domain sensing. (a) Single CAM cell and (b) a CAM word with multiple cells connected on the same match line. A two-stage buffer circuit is adopted as the sense amplifier. The pMOSFET transistor is used to pre-charge the match line for the search operation.
The Decision Tree Implementation

Figure S9: The circuit schematic of a single decision tree implementation. (a) An ACAM array implementing a decision tree. (b) The sense amplifier circuit for sensing the ACAM array ML. Serially connected buffers are used to sense the $V_{ML}$. (c) The digital-analog-converter (DAC) circuit for applying the SL voltages.
Figure S10: The schematic of a single random forest consisting of n trees and the digital logic circuits post processing the match/mismatch results of k classes and performing the majority voting for the forest. The analog CAM arrays implementing the trees generate output vote vectors (i.e., ‘1’ for match and ‘0’ for mismatch) which are fed into their associated adders. Each adder records the votes for a class from all the trees whose leaf node is associated with the class, and the non-binary output vectors of all adders are combined to represent the forest output. Then the comparators binarize the vote vector by comparing the corresponding class vote count with a preset threshold value and generating a binary bit. Finally, after aggregating the majority results from the comparators, the output vote vector of the forest is generated and concatenated with the results from other forests and the input features.
Figure S11: The schematic of the adders used in the majority voting component in Fig. S10.
The Schematic of the Comparator

Figure S12: The schematic of the comparator used in the majority voting component in Fig. S10. It performs iterative comparisons from the most significant bit to the least significant bit.
The Schematic of DRF Implementation

Figure S13: The architecture of our proposed DRF accelerator and the associated circuitry for transmitting the voting information and control signals between different forests.
Figure S14: Transient waveform of the sense amplifier output in a single CAM cell. (a) Output voltage waveform at different $V_{SL}$ values for 8 different cell configurations. By configuring the $F_0$ FeFET into 1 of the 8 $V_{TH}$ states, the decision boundary threshold is shifted accordingly. In this work, a 10ns search time, i.e., the red dashed plane, is chosen to sense the output voltage vs. the $V_{SL}$. (b) The match line voltage and the final output voltage as a function of the $V_{SL}$ sensed at the search time of 10ns. A sharp $V_{SL}$ decision boundary can be realized with the sensing circuit.
Figure S15: Current sensing in an CAM array. Similar to the experimental measurement for the ACAM array shown in Fig. 2, an ACAM array with wordlength of 2 is simulated in SPICE and the match line current is measured when the array is configured into 1 of the 4 × 4 configurations. Similar to the experiment, the simulation confirms that the decision boundary on each \( V_{SL} \) dimension is independent of each other and together the \( V_{SL} \)'s of all the cells define the matching region in the \( V_{SL} \) space.
Figure S16: Voltage domain sensing of a CAM array with different numbers of columns. In the simulation, the worst sensing scenario is considered where only one cell is swept, its decision boundary is set to be close to 0.5V, all other cells connected on the same match line store the same state, and all other cells are searched with the same $V_{SL}$ (close to their decision boundary). With more cells connected to the match line, the leakage current becomes larger as contributed by other cells, thus advancing the discharge of the match line and the sense amplifier output. Therefore the search time, as indicated by the red dashed plane, needs to be adjusted for a CAM array with a larger number of columns.
Branch Split Threshold Precision Extension

Figure S17: A possible approach to extend the precision of the branch split threshold using ferroelectric analog CAM cells with a limited precision. (a) The branch splits based on the threshold value $k$, which is represented as $N + M$ bits. The blue/red are the less-than/greater-than branches, respectively. (b) Realizing the $N + M$ bit split threshold precision utilizing only analog CAM cells with $N$ and $M$ bits precision. The key idea is to first locate the interval identified by $N$ MSB bits (the green interval), then the identified interval is used for $M$ LSB bits search. To identify the $N$ bit interval, two steps are used, where step 1 is to find the upper bound of the interval and step 2 is to find the lower bound of the interval. Then step 3 is to zoom in the interval for $M$ bits search.
Impact of Device-to-Device Variation on The System Accuracy

Figure S18: Impact of device-to-device variation on the accuracy of deep random forest. (a), (b), (c) are the simulated sense amplifier output of a single CAM cell given a FeFET $V_{TH}$ standard deviation of 1%, 3%, and 5% of the overall memory window, respectively. (d), (e), (f) are the histograms of the decision boundaries under 1%, 3%, and 5% variation. The variation in the decision boundary degrades as $V_{TH}$ variation increases. Studies of the deep random forest accuracy under different degrees of variation in the decision boundary for (g) the MNIST and (h) the SEMG dataset. For MNIST, the input is binary, which is highly robust to the decision boundary variation as long as correct distinction between the black and white pixels can be made. For the SEMG, accuracy degradation starts to emerge when the standard deviation of the decision boundary exceeds 7% of the overall memory window. This work exhibits a standard deviation of 4% of the overall memory window, as shown in Fig. S2.
Power Consumption Breakdown

The power model for a $128 \times 128$ FeFET ACAM array is presented in Fig. S19. Pre-charging the MLs before the search operation consumes approximately 1909 fJ, accounting for 41.2% of the total energy consumption. The SAs associated with the MLs consume around 1802 fJ during the search, which is 38.8% of the total energy consumption. A 3-bit DAC is used to charge the SL, and the DACs associated with the SLs consume 929 fJ on average for charging, which is around 20.0% of the total energy consumption. During the search, MLs are discharged to ground, resulting in the ML voltage reduction. This ML discharging is not considered as part of the energy consumption of current search cycle, since the charge from MLs to ground is not supplied by the voltage sources. The reduced ML voltage will be pre-charged in the next search operation.

Figure S19: Energy consumption breakdown of a $128 \times 128$ analog CAM array for a search operation, including MLs pre-charging (PC), the sense amplifiers (SA) associated with MLs and digital-analog-converters (DAC) for charging SLs.

To have a practical estimation of the energy in Fig. S19, parasitic capacitance have been extracted and included in the array. The total capacitance on the ML includes both all the transistor drain capacitance associated with the ML (which is captured by BSIM 4 and the predictive technology model (PTM) model card) and the interconnect capacitance on the ML. In our simulations, the parasitic (i.e., the interconnect) capacitance on the ML is extracted by the DESTINY tool [50]. The total parasitic capacitance associated with the
ML increases linearly as the number of cells per row grows. For a $128 \times 128$ CAM array, the parasitic capacitance on the ML is 6.4 fF. The discharging current of the ML during the search changes as the ML discharge follows an RC model, where R is the equivalent resistance from ML to ground, and C is the total capacitance associated with the ML. In the simulation, the ML discharge peak current reaches 65 $\mu$A immediately after search starts.
Compilation of Advanced Machine Learning Model Hardware

In addition to the DRF discussed in the main text, we have also implemented a random forest using the ferroelectric analog CAM. The performance of the implementation is evaluated with seizure detection on the EEG (CHB-MIT) and PET/CT datasets for pediatric patients. 3-bit precision is used to quantify the extracted features for the training of the random forest model. A 128×128 array is used to implement the trained RF models, accommodating all the features of PET/CT and EEG datasets. A compilation of various reported hardware for machine learning models are presented here. The intention is not to compare against different hardware implementations as direct comparison is unfair due to different models and technologies used.

| Reference | Machine Learning Model | Architecture | Technology | Bit cell size | Dataset | Energy / Classification | Classification time |
|-----------|------------------------|--------------|------------|---------------|---------|-------------------------|---------------------|
| 51        | RF                     | Intel X5560  | CMOS 45nm  | N/A           | URL Reputation | 20.4mJ                  | 107.5µs             |
| 51        | RF                     | NVIDIA Tesla M2050 | CMOS 45nm  | N/A           | URL Reputation | 11mJ                  | 49µs                |
| 51        | RF                     | Xilinx Virtex-6 | CMOS 40nm  | N/A           | URL Reputation | 0.351mJ               | 31.9µs              |
| 53        | Vocabulary tree        | Digital      | CMOS 65nm  | N/A           | COIL-100 | 460µJ                  | 16.7ms              |
| 55        | Vocabulary tree        | Digital      | CMOS 65nm  | N/A           | N/A      | 186.7µJ                | 33.3ms              |
| 56        | AdaBoost               | In-memory    | CMOS 180nm | 4.33µm²       | MNIST   | 0.6nJ                  | 20ns                |
| 57        | SVM                    | In-memory    | CMOS 65nm  | 1.94µm²       | MIT CBCL | 963pJ                  | 107.5ns             |
| 59        | SVM                    | In-memory    | CMOS 65nm  | 2.56µm²       | MIT CBCL | 42pJ                   | 31.2ns              |
| 60        | RF                     | In-memory    | CMOS 65nm  | 1.94µm²       | MIT CBCL | 19.4nJ                 | 2.7µs               |
| 64        | RF                     | In-Memory    | RRAM + CMOS 16nm | 0.52µm²   | IRIS | 0.17nJ                  | 48ns                |
| This work | RF                     | In-Memory    | FeFET + CMOS 28nm | 0.06µm²   | EEG | 4.64pJ                  | 1.04ns              |

Table S1: Performance summary of advanced machine learning model implementations. The random forest hardware based on ferroelectric analog CAM is compact and energy-efficient.
REFERENCES AND NOTES

1. A. Keshavarzi, W. van den Hoek, Edge intelligence—On the challenging road to a trillion smart connected IoT devices. *IEEE Des. Test* **36**, 41–64 (2019).

2. A. Keshavarzi, K. Ni, W. van den Hoek, S. Datta, A. Raychowdhury Ferroelectronics for edge intelligence. *IEEE Micro* **40**, 33–48 (2020).

3. Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, J. Zhang Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proc. IEEE* **107**, 1738–1762 (2019).

4. X. Xu, Y. Ding, S. X. Hu, M. Niemier, J. Cong, Y. Hu, Y. Shi Scaling for edge inference of deep neural networks. *Nat. Electron.* **1**, 216–222 (2018).

5. X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, X. Chen Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **22**, 869–904 (2020).

6. D. Xu, T. Li, Y. Li, X. Su, S. Tarkoma, T. Jiang, J. Crowcroft, P. Hui, Edge intelligence: Architectures, challenges, and applications. arXiv:2003.12172 [cs.NI] (2020).

7. Z. Feng, S. George, J. Harkes, P. Pillai, R. Klatzky, M. Satyanarayanan, Edge-based discovery of training data for machine learning, in *2018 IEEE/ACM Symposium on Edge Computing (SEC)* (IEEE, 2018), pp. 145–158.

8. F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning. arXiv:1702.08608 [stat.ML] (2017).

9. S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, T. D. Kelley, D. Braines, M. Sensoy, C. J. Willis, P. Gurram, Interpretability of deep learning models: A survey of results, in *Proceedings of the 2017 IEEE Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (Smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)* (IEEE, 2017), pp. 1–6.
10. A. Vellido, The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput. Appl.* **32**, 18069–18083 (2020).

11. A. B. Arrieta, N. D. Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).

12. Z.-H. Zhou, J. Feng, Deep forest. *Natl. Sci. Rev.* **6**, 74–86 (2019).

13. M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **15**, 3133–3181 (2014).

14. G. Pedretti, C. E. Graves, S. Serebryakov, R. Mao, X. Sheng, M. Foltin, C. Li, J. P. Strachan, Tree-based machine learning performed in-memory with memristive analog cam. *Nat. Commun.* **12**, 5806 (2021).

15. G. Pedretti, S. Serebryakov, J. P. Strachan, C. E. Graves, A general tree-based machine learning accelerator with memristive analog CAM, in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, 2022), pp. 220–224.

16. G. Pedretti, J. Moon, P. Bruel, S. Serebryakov, R. M. Roth, L. Buonanno, A. Gajjar, T. Ziegler, C. Xu, M. Foltin, P. Faraboschi, J. Ignowski, C. E. Graves, X-time: An in-memory engine for accelerating machine learning on tabular data with CAMs. arXiv:2304.01285 [cs.LG] (2023).

17. K. Pagiamtzis, A. Sheikholeslami, Content-addressable memory (CAM) circuits and architectures: A tutorial and survey. *IEEE J. Solid-State Circuits* **41**, 712–727 (2006).

18. R. Karam, R. Puri, S. Ghosh, S. Bhunia Emerging trends in design and applications of memory-based computing and content-addressable memories. *Proc. IEEE* **103**, 1311–1330 (2015).

19. M. Imani, A. Rahimi, D. Kong, T. Rosing, J. M. Rabaey, Exploring hyperdimensional associative memory, in *2017 IEEE International Symposium on High Performance Computer Architecture (HPCA)* (IEEE, 2017), pp. 445–456.
20. K. Ni, X. Yin, A. F. Laguna, S. Joshi, S. Dünkel, M. Trentzsch, J. Müller, S. Beyer, M. Niemier, X. S. Hu, S. Datta Ferroelectric ternary content-addressable memory for one-shot learning. *Nat. Electron.* **2**, 521–529 (2019).

21. X. Yin, C. Li, Q. Huang, L. Zhang, M. Niemier, X. S. Hu, C. Zhuo, K. Ni Fecam: A universal compact digital and analog content addressable memory using ferroelectric. *IEEE Trans. Electron Devices* **67**, 2785–2792 (2020).

22. C. Li, C. E. Graves, X. Sheng, D. Miller, M. Foltin, G. Pedretti, J. P. Strachan Analog content-addressable memories with memristors. *Nat. Commun.* **11**, 1638 (2020).

23. C. Li, C. E. Graves, X. Sheng, D. Miller, M. Foltin, G. Pedretti, J. P. Strachan, A scalable design of multi-bit ferroelectric content addressable memory for data-centric computing, in 2020 *IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2020), pp. 29–33.

24. A. Kazemi, M. M. Sharifi, A. F. Laguna, F. Müller, R. Rajaei, R. Olivo, T. Kämpfe, M. Niemier, X. Sharon Hu, In-memory nearest neighbor search with FeFET multi-bit content-addressable memories. arXiv:2011.07095 [cs.ET] (2020).

25. Y. Xiao, Y. Xu, Z. Jiang, S. Deng, Z. Zhao, A. Mallick, L. Sun, R. Joshi, X. Li, N. Shukla, V. Narayanan, K. Ni, On the write schemes and efficiency of FeFET 1T NOR array for embedded nonvolatile memory and beyond, in 2022 *International Electron Devices Meeting (IEDM)* (IEEE, 2022), pp. 13.6.1–13.6.4.

26. Z. Jiang, Z. Zhao, S. Deng, Y. Xiao, Y. Xu, H. Mulaosmanovic, S. Duenkel, S. Beyer, S. Meninger, M. Mohamed, R. Joshi, X. Gong, S. Kurinec, V. Narayanan, K. Ni On the feasibility of 1T ferroelectric fet memory array. *IEEE Trans. Electron Devices* **69**, 6722–6730 (2022).

27. K. Ni, X. Li, J. A. Smith, M. Jerry, S. Datta Write disturb in ferroelectric FETs and its implication for 1T-FeFET and memory arrays. *IEEE Electron Device Lett.* **39**, 1656–1659 (2018).

28. U. Schroeder, S. Slesazeck, H. Mulaosmanovic, T. Mikolajick, Nonvolatile field-effect transistors using ferroelectric-doped HfO$_2$ films, in *Ferroelectric-Gate Field Effect Transistor*
Memories. Topics in Applied Physics, vol 131, B. E. Park, H. Ishiwa, M. Okuyama, S. Sakai, S. M. Yoon, Eds. (Springer, 2020).

29. M. Trentzsch, S. Flachowsky, R. Richter, J. Paul, B. Reimer, D. Utess, S. Jansen, H. Mulaosmanovic, S. Müller, S. Slesazeck, J. Ocker, M. Noack, J. Müller, P. Polakowski, J. Schreiter, S. Beyer, T. Mikolajick, B. Rice, A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs, in 2016 IEEE International Electron Devices Meeting (IEDM) (IEEE, 2016), pp. 11–15.

30. M. Lederer, T. Kämpfe, R. Olivo, D. Lehninger, C. Mart, S. Kirbach, T. Ali, P. Polakowski, L. Roy, K. Seidel Local crystallographic phase detection and texture mapping in ferroelectric Zr doped HfO₂ films by transmission-EBSD. Appl. Phys. Lett. 115, 222902 (2019).

31. X. Yin, F. Müller, Q. Huang, C. Li, M. Imani, Z. Yang, J. Cai, M. Lederer, R. Olivo, N. Laleni, S. Deng, Z. Zhao, Z. Shi, Y. Shi, C. Zhuo, T. Kämpfe, K. Ni An ultracompact single-ferroelectric field-effect transistor binary and multibit associative search engine. Adv. Intell. Syst. 5, 2200428 (2023).

32. M. Lederer, A. Reck, K. Mertens, R. Olivo, P. Bagul, A. Kia, B. Volkmann, T. Kämpfe, K. Seidel, L. M. Eng Impact of the SiO₂ interface layer on the crystallographic texture of ferroelectric hafnium oxide. Appl. Phys. Lett. 118, 012901 (2021).

33. H. Mulaosmanovic, J. Ocker, S. Müller, U. Schroeder, J. Müller, P. Polakowski, S. Flachowsky, R. van Bentum, T. Mikolajick, S. Slesazeck, Switching kinetics in nanoscale hafnium oxide based ferroelectric field-effect transistors. ACS Appl. Mater. Interfaces 9, 3792–3798 (2017).

34. H. Mulaosmanovic, S. Dunkel, M. Trentzsch, S. Beyer, E. T. Breyer, T. Mikolajick, S. Slesazeck Investigation of accumulative switching in ferroelectric FETs: Enabling universal modeling of the switching behavior. IEEE Trans. Electron Devices 67, 5804–5809 (2020).

35. H. Bae, S. G. Nam, T. Moon, Y. Lee, S. Jo, D.-H. Choe, S. Kim, K.-H. Lee, J. Heo, Sub-ns polarization switching in 25nm Fe FinFET toward post CPU and spatial-energetic mapping of
traps for enhanced endurance, in *2020 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2020), pp. 31–33.

36. M. Jerry, P.-Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, S. Datta, Ferroelectric FET analog synapse for acceleration of deep neural network training, in *2017 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2017), pp. 6.2.1–6.2.4.

37. X. Sun, P. Wang, K. Ni, S. Datta, S. Yu, Exploiting hybrid precision for training and inference: A 2T-1FeFET based analog synaptic weight cell, in *2018 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2018), pp. 3.1.1–3.1.4.

38. M. Halter, L. Bégon-Lours, V. Bragaglia, M. Sousa, B. J. Offrein, S. Abel, M. Luisier, J. Fompeyrine Back-end, CMOS-compatible ferroelectric field-effect transistor for synaptic weights. *ACS Appl. Mater. Interfaces* 12, 17725–17732 (2020).

39. K. Ni, M. Jerry, J. A. Smith, S. Datta, A circuit compatible accurate compact model for ferroelectric-FETs, in *2018 IEEE Symposium on VLSI Technology* (IEEE, 2018), pp. 131–132.

40. O. Guehairia, A. Ouamane, F. Dornaika, A. Taleb-Ahmed Feature fusion via deep random forest for facial age estimation. *Neural Netw.* 130, 238–252 (2020).

41. S. A. Roseline, S. Geetha, S., Kadry, Y. Nam, Intelligent vision-based malware detection and classification using deep random forest paradigm. *IEEE Access* 8, 206303–206324 (2020).

42. X. Cao, R. Li, Y. Ge, B. Wu, L. Jiao, Densely connected deep random forest for hyperspectral imagery classification. *Int. J. Remote Sens.* 40, 3606–3622 (2019).

43. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324 (1998).

44. C. Sapsanis, G. Georgoulas, A. Tzes, D. Lymberopoulos, Improving EMG based classification of basic hand movements using EMD, in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2013), pp. 5754–5757.
45. H. Lee, R. Grosse, R. Ranganath, A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML ‘09, 609–616 (Association for Computing Machinery, 2009).

46. S. Beyer, S. Dünkel, M. Trentzsch, J. Müller, A. Hellmich, D. Utess, J. Paul, D. Kleimaier, J. Pellerin, S. Müller, J. Ocker, A. Benoist, H. Zhou, M. Mennenga, M. Schuster, F. Tassan, M. Noack, A. Pourkeramati, F. Müller, M. Lederer, T. Ali, R. Hoffmann, T. Kämpfe, K. Seidel, H. Mulaosmanovic, E. T. Breyer, T. Mikolajick, S. Slesazeck, FeFET: A versatile CMOS compatible device with game-changing potential, in *2020 IEEE International Memory Workshop (IMW)* (IEEE, 2020), pp. 1–4.

47. A. H. Shoeb, “Application of machine learning to epileptic seizure onset detection and treatment,” thesis, Massachusetts Institute of Technology, Cambridge, MA (2009).

48. Q. Zhang, Y. Liao, X. Wang, T. Zhang, J. Feng, J. Deng, K. Shi, L. Chen, L. Feng, M. Ma, L. Xue, H. Hou, X. Dou, C. Yu, L. Ren, Y. Ding, Y. Chen, S. Wu, Z. Chen, H. Zhang, C. Zhuo, M. Tian A deep learning framework for $^{18}$F-FDG pet imaging diagnosis in pediatric patients with temporal lobe epilepsy. *Eur. J. Nucl. Med. Mol. Imaging* **48**, 2476–2485 (2021).

49. Predictive technology model (ptm) (2012). http://ptm.asu.edu/.

50. M. Poremba, S. Mittal, D. Li, J. S. Vetter, Y. Xie, Destiny: A tool for modeling emerging 3D NVM and EDRAM caches, in *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (IEEE, 2015).

51. B. Van Essen, C. Macaraeg, M. Gokhale, R. Preger, Accelerating a random forest classifier: Multi-core, GP-GPU, or FPGA?, in *2012 IEEE 20th International Symposium on Field-Programmable Custom Computing Machines* (IEEE, 2012).

52. J. Ma, J. K. Saul, S. Savage, G. M. Voelker. Identifying suspicious urls: An application of large-scale online learning, in *Proceedings of the 26th Annual International Conference on Machine Learning* (Association for Computing Machinery, 2009), pp. 681–688.
53. K. J. Lee, G. Kim, J. Park, H. J. Yoo, A vocabulary forest object matching processor with 2.07 M-vector/s throughput and 13.3 nJ/vector per-vector energy for full-HD 60 fps video object recognition. *IEEE J. Solid-State Circuits* **50**, 1059–1069 (2015).

54. S. A. Nene, S. K. Nayar, H. Murase, Columbia object image library (coil-100) (1996).

55. T.-W. Chen, Y.-C. Su, K.-Y. Huang, Y.-M. Tsai, S.-Y. Chien, L.-G. Chen, Visual vocabulary processor based on binary tree architecture for real-time object recognition in full-HD resolution. *IEEE Trans. Very Large Scale Integr.* **20**, 2329–2332 (2012).

56. J. Zhang, Z. Wang, N. Verma, In-memory computation of a machine-learning classifier in a standard 6T SRAM array. *IEEE J. Solid-State Circuits* **52**, 915–924 (2017).

57. M. Kang, A multi-functional in-memory inference processor using a standard 6t sram array. *IEEE J. Solid-State Circuits* **53**, 642–655 (2018).

58. Center for Biological & Computational Learning (CBCL) at MIT, http://poggio-lab.mit.edu/codedatasets [accessed 4 June 2021].

59. S. K. Gonugondla, M. Kang, N. Shanbhag, A 42pJ/decision 3.12 TOPS/W robust in-memory machine learning classifier with on-chip training, in *2018 IEEE International Solid-State Circuits Conference-(ISSCC)* (IEEE, 2018), pp. 490–492.

60. M. Kang, S. K. Gonugondla, S. Lim, N. R. Shanbhag, A 19.4-nj/decision, 364-k decisions/s, in-memory random forest multi-class inference accelerator. *IEEE J. Solid-State Circuits* **53**, 2126–2135 (2018).

61. R. A. Fisher, The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936).