

On-sensor Printed Machine Learning Classification via Bespoke ADC and Decision Tree Co-Design

Giorgos Armeniakos*, Paula L. Duarte[†], Priyanjana Pal[‡], Georgios Zervakis[‡],
Mehdi B. Tahoori[†], Dimitrios Soudris*

*National Technical University of Athens, GR, [†]University of Patras, GR, [‡]Karlsruhe Institute of Technology, DE
*{armeniakos, dsoudris}@microlab.ntua.gr, [†]{paula.duarte, priyanjana.pal, mehdi.tahoori}@kit.edu, [‡]zervakis@upatras.gr

Abstract—Printed electronics (PE) technology provides cost-effective hardware with unmet customization, due to their low non-recurring engineering and fabrication costs. PE exhibit features such as flexibility, stretchability, porosity, and conformality, which make them a prominent candidate for enabling ubiquitous computing. Still, the large feature sizes in PE limit the realization of complex printed circuits, such as machine learning classifiers, especially when processing sensor inputs is necessary, mainly due to the costly analog-to-digital converters (ADCs). To this end, we propose the design of fully customized ADCs and present, for the first time, a co-design framework for generating bespoke Decision Tree classifiers. Our comprehensive evaluation shows that our co-design enables self-powered operation of on-sensor printed classifiers in all benchmark cases.

Index Terms—Co-design, Decision Tree, Low-Power, Machine Learning, Printed Electronics

I. INTRODUCTION

Printed electronics technology, capturing substantial attention, holds promise for computing in various domains yet to embrace it. Examples include smart packaging, disposables (e.g., packaged foods), FMCG, in-situ monitoring, and healthcare products like smart bandages [1]. These domains impose rigorous prerequisites for ultra-low-cost fabrication, unattainable by silicon-based systems. Furthermore, aside from their high costs, silicon systems lack the stretchability, porosity, flexibility, and conformality inherent in printed electronics.

While printed circuits exhibit the potential to meet the demands of ultra-low cost (even sub-cent levels) due to their inexpensive additive manufacturing processes, the relatively large feature sizes, inherent in printed electronics, pose a challenge in realizing large-scale integration circuits for on-sensor processing. This becomes particularly relevant for Machine Learning (ML) classification circuits, as classification is the primary task for numerous applications in aforementioned domains [2, 3]. Such applications have to classify (analog) data collected from printed sensors to extract useful information.

Significant recent research has focused on addressing the intrinsic constraints of printed electronics, aiming to enable the realization of printed-battery-powered printed ML classifiers [2]–[7]. In [2], the authors leverage the substantial customization capabilities stemming from the low fabrication costs of printed circuits, and explored the concept of *bespoke* ML circuits that are tailored to a specific model. The term “bespoke” signifies fully customized circuit implementations, adapting to individual ML models and dataset, and can be

obtained by hardwiring all model parameters in the circuit design. This level of customization is infeasible in conventional silicon-based systems [2]. While bespoke design indeed yields significant area and power gains, further improvements are mandatory to enable printed-battery-powered ML classifiers. The works in [3]–[5] and [6] combined bespoke design with approximate [8] and stochastic [9] computing, respectively, targeting Multilayer Perceptrons and/or Support Vector Machines. However, in most cases, the obtained gains are either inadequate or come at the cost of a considerable accuracy degradation. Leveraging the fact that in printed applications the classification tasks and datasets are relatively simple [2], the authors in [2, 7] utilize Decision Trees due to their implementation simplicity. Combining Decision Trees with bespoke design and approximate computing, [7] delivered promising results. Despite the impressive results demonstrated in the aforementioned works, they all overlook the substantial—potentially dominant—power and area overheads associated with the mandatory analog-to-digital converters (ADCs) for processing sensor data. Thus, their efficiency remains unclear.

In this work, we address the limitation of the current state-of-the-art. Our focus goes beyond printed-battery operation; instead, we target the design of self-powered (i.e., using printed energy harvesters) printed classifiers, accounting also for the cost of the ADCs. To achieve this, we adopt Decision Trees as our classification algorithm and propose a model-circuit co-design framework. To maximize hardware efficiency, we implement bespoke Decision Tree classifiers based on the unary numeric system [10]. As we explain later, this approach may reduce a Decision Tree to a simple two-stage logic. Furthermore, we leverage the high customization in printed circuits and propose the design of bespoke ADCs tailored to our Decision Tree architecture by retaining only the bare minimum of ADC comparators. Finally, we propose and implement an ADC-aware Decision Tree training that identifies parameters that minimize the ADCs cost while still adhering to accuracy requirements. Our evaluation demonstrates that, compared to the state-of-the-art baseline, we reduce the area and power of printed Decision Trees by on average 8.6x and 12.2x, respectively, for less than 1% accuracy degradation.

Our novel contributions in this work are as follows:

- 1) This is the first work that considers and investigates the impact of ADCs in designing digital printed ML classifiers.
- 2) We propose the first ADC-aware co-design framework

dedicated to bespoke printed Decision Trees¹.

- 3) Our framework enables for the first time self-powered on-sensor digital ML classifiers in printed electronics.

II. BACKGROUND

A. Printed Electronics

Much akin to color printing, Printed Electronics (PE) technology follows an additive manufacturing process for electronic components, utilizing diverse printing techniques like roll-to-roll and jet printing [11]. Tailored to the printing materials and substrates used, printed circuits present multiple benefits such as non-toxicity, lightness, and flexibility. Also, due to the additive nature of PE, the production of printed circuits can be achieved at an exceptionally low cost. This advantage greatly simplifies the production of innovative products by enabling cost-effective manufacturing of customized, purpose-specific electronic devices in small quantities. Though, due to the low precision printing, PE exhibits low integration density (orders of magnitude lower than silicon VLSI), large feature sizes, and elevated device latencies. Nevertheless, in target domains, the performance and computation requirements are generally quite modest, e.g., sampling rate of a few Hz and few bits precision [12]. These requirements could be effectively met by printing technologies, while still adhering to acceptable area and energy constraints. In this work, we consider the inorganic Electrolyte-Gated FET (EGFET) technology [1] that features below 1V supply voltage and high mobility characteristics, unlike organic counterparts, making it a suitable match for low-power and energy-harvested applications.

B. Flash ADC

ADCs play a crucial role in modern electronics by converting analog signals, such as sensor data, into digital for further processing. We focus on Flash ADC because of its implementation simplicity, especially for very low precision requirements as in printed applications, and its minimal latency. Flash ADCs excel at delivering rapid data conversion, ensuring that data is promptly available even during short energy availability windows. Additionally, despite consuming high power during conversion, they mainly operate in a low-power state, thereby minimizing average power and energy consumption, aligning well with energy-harvesting approaches.

A typical 3-bit flash ADC block diagram is shown in Fig. 1a. A key component of a Flash ADC is a bank of comparators. The number of comparators in the array corresponds to the desired resolution of the ADC. For an N -bit ADC, $2^N - 1$ comparators are required. The input analog voltage (V_{in}) is connected to the non-inverting terminals of all the comparators. The reference voltage range is divided into 2^N segments to cover the entire range of the input analog voltage. Each comparator is connected to a reference voltage (V_{ref}) that corresponds to the midpoint of each segment. When an analog input voltage V_{in} is applied to the ADC, each comparator in the array compares V_{in} with its corresponding V_{ref} . If V_{in}

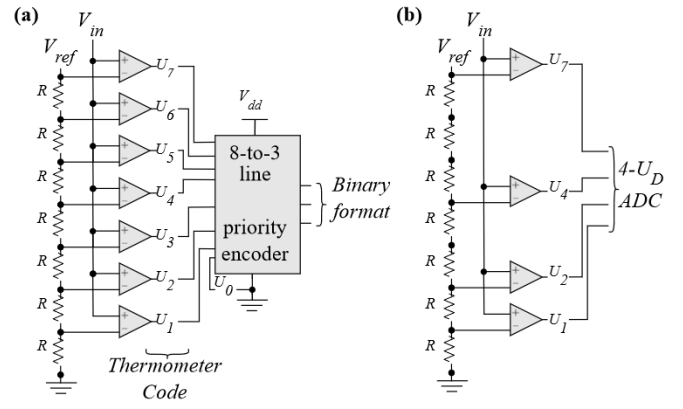


Fig. 1. Schematic of: a) conventional 3-bit Flash ADC and b) an example of an equivalent bespoke ADC with four unary digits of output.

is larger than V_{ref} , the comparator outputs a high signal (1); otherwise, it outputs a low signal (0). The comparators outputs form the digital representation of V_{in} in a thermometer code, which is then processed by a fast priority encoder to produce the corresponding binary value.

C. Unary Decision Tree Circuits

Unary computing traditionally operates on serial bitstreams using extremely simple logic [10]. State-of-the-art unary implementations typically rely on rate or temporal logic [10] and handle data in the form of serial bitstreams [13]. While these prior works present efficient unary computation units, they require expensive converters [14] and mandate multi-cycle operation, which is prohibitive in PE [2]. On the other hand, we exploit the unmatched customization capabilities of PE and design fully parallel, unary-based Decision Tree classifiers.

III. OUR PROPOSED CO-DESIGN FRAMEWORK

This section describes our co-design framework for generating printed Decision Trees. In brief, we first introduce the architecture of a fully-parallel Decision Tree based on the unary representation and highlight its area and power benefits over non-unary approaches. Then, we analyze our flash ADC design tailored for such architectures, and we describe our ADC-aware Decision Tree training that enables minimizing the ADCs hardware cost while maintaining high accuracy.

A. Parallel Unary Decision Trees

In unary coding (or else thermometer code), an N -bit binary number is represented by a code of length $2^N - 1$. The count of '1's in the unary code corresponds to the value being represented. Unary coding can express many types of numbers, such as integers, fixed-point, etc. For instance:

$$\begin{aligned} 0011111_U &= 101_2 = 5_{10} \\ .111_U &= .11_2 = .75_{10} \end{aligned} \quad (1)$$

A parallel unary format is generally not preferred due to its increased size to represent a number (N vs $2^N - 1$ bits) [15]. However, as we show hereafter, this does not hold in our case and thus, we investigate and propose the

¹Available at <https://github.com/garmeniakos/Ax-Printed-ML-Classifiers>.

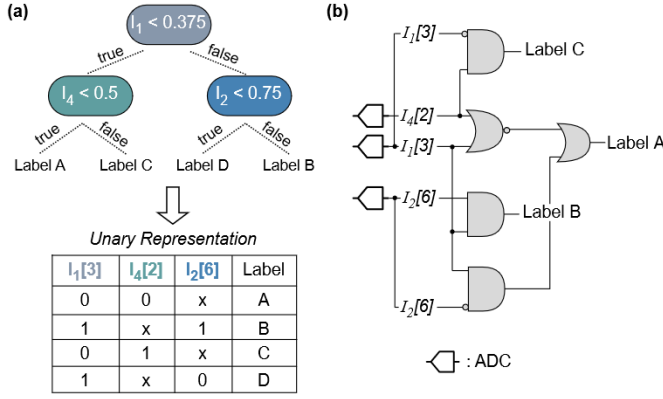


Fig. 2. Illustration of how a conventional Decision Tree (a) is translated into unary format, represented by a set of unary digits. This example assumes Q0.4 formatted values, e.g., $0.75 \rightarrow 6$. (b) depicts the simplified schematic.

implementation of parallel unary printed Decision Trees. As aforementioned, the unmatched customization in printed circuits enables hardcoding the model's parameters. As a result, a comparison $I \geq C$ where I is an input and C is a model parameter, becomes $I \geq .1011_2$ assuming that $.1011_2$ is the trained value of C . In unary format C can be written as $.000011111111111_2$. Therefore, although parallel unary representation doesn't typically make sense in conventional architectures due to the exponentially increased number of bits to be compared, in bespoke Decision Trees the comparator is essentially reduced to simply checking a bit from the input:

$$I \geq .1011_2 \xrightarrow{\text{Unary}} I \geq .000011111111111_2 \equiv I[11] \quad (2)$$

In other words, if the 11th bit of I is '1' then the initial inequality $I \geq C$ is true. In general, if the most significant '1' of C is at bit position k , then $I \geq C \equiv I[k]$. In addition, the inequality is directly computed (I is in parallel format), eliminating the need to wait for bit k to arrive from a serial input. Similar relations are derived for all comparisons: $I > C \equiv I[k+1]$, $I < C \equiv \neg(I[k])$, and $I \leq C \equiv \neg(I[k+1])$. Note that in unary format if $I[k] = 1$, then $I[j] = 1 \forall j \leq k$.

The analysis above highlights that in a bespoke Decision Tree, if the inputs are provided in a parallel unary format, all Tree comparators can be removed. This is perfectly in line with the limited hardware resources of printed circuits. It is crucial to note that a serial (e.g., temporal) unary representation would not only enforce a printed-unfriendly multi-cycle operation but would also introduce a significant hardware overhead such as control circuitry and potentially numerous registers.

Fig. 2 presents an example architecture of a Decision Tree when its inputs are available in a parallel unary representation. Instead of traditional comparisons, bespoke unary DTs can now be viewed as a simple logic over a set of unary digits corresponding to the trained parameters (e.g., $I_1[3]$, $I_4[2]$, $I_2[6]$). The truth table for each label of the unary Decision Tree example is also depicted. As shown, only a few gates are sufficient. Each label is obtained through a simple two-level logic (e.g., AND-OR) and each input signal in this two-level

logic denotes a node in the Decision Tree.

B. Bespoke ADCs

As shown in the previous section, ensuring the availability of inputs in a parallel unary representation minimizes the hardware overheads of printed Decision Tree classifiers. As shown in Fig. 1a, the flash ADC calculates an intermediate result, which is the thermometer code of the input signal (each U_i denotes if V_{in} is larger than V_{ref}). As a result, by simply removing the encoder, we not only decrease the ADC's hardware requirements but also achieve our objective: transforming the sensor input into a parallel unary representation before supplying it to the Decision Tree classifier. This also further justifies our flash ADC consideration.

To further improve the hardware-efficiency of our ADCs, we also implement them in a fully customized manner. The simplified Decision Tree design described above (e.g., Fig. 2) relies on a parallel unary representation of its inputs. Though, as indicated by (2), each comparison requires only a specific input bit. Consequently, the remaining unary digits, if not needed for another comparison, can be discarded and don't have to be generated. An illustrative example of a bespoke ADC is presented in Fig. 1b. In this example, we assume that the respective input is involved in multiple comparisons and is compared against four different parameters. Hence, only four unary digits need to be calculated. Without any loss of generality, this example assumes that the 1st, 2nd, 4th, and 7th unary digits are required for the comparisons. Fig. 1b presents the architecture of the corresponding "4- U_D " (four output unary digits) ADC. To generate the specific ADC, we only need to retain the corresponding four comparators, and we can eliminate the remaining three comparators and the encoder. In general, bespoke ADC design entails retaining only the resistors and the bare-minimum comparators required.

The number of comparators, as well as the specific ones to be retained, is determined by the trained Decision Tree parameters. Considering a conventional 4-bit ADC, Fig. 3 presents how the area and power characteristics of our bespoke ADCs scale w.r.t. the number and position of their output unary digits. In Fig. 3, the number of output digits ranges from 1 to 15. For example, a 2- U_D ADC denotes a bespoke ADC with only two outputs, while a 15- U_D ADC indicates the retention of all comparators. To showcase the power behavior, a few representative examples are presented w.r.t. the specific outputs that are retained. The area and power of the conventional 4-bit ADC are 11mm² and 0.83mW, respectively. To obtain the area and power values, we designed the ADCs in Cadence Virtuoso using the EGFET Process Design Kit (PDK) [1] and conducted SPICE simulations. The simulations are conducted with a voltage supply of 1V. As shown in Fig. 3, the area of a bespoke ADC solely depends on the number of output unary digits of the ADC. Specifically, the area scales linearly along with the number of the comparators. On the other hand, power consumption also depends on which outputs are selected. For example, in a 4- U_D bespoke ADC, power consumption ranges from 47uW up to 205uW, i.e., $4.4\times$ increase. It is observed

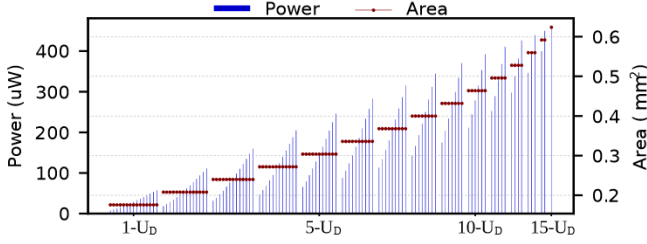


Fig. 3. The Area and power of (4-bit) bespoke ADCs w.r.t. their output unary digits. The first values correspond to 1-output ADCs, while the last one to 15-output ADC. Different points denote different output digits. Selected output digits are in sequential order, i.e., “ U_1 - U_2 ” 2- U_D ADC is followed by “ U_2 - U_3 ” 2- U_D ADC and so on, only to showcase the behavior of power.

that the power is substantially decreased when lower-order outputs are selected. This can be attributed to the lower V_{ref} of the lower-order comparators. Fig. 3 shows a linear increase in the comparators power consumption as we move towards higher-order outputs. The key takeaways from this analysis are twofold. First, bespoke ADCs provide significant hardware improvements over conventional ADCs. Second, the area and power consumption of a bespoke ADC are highly influenced by its outputs. By minimizing the number of outputs for each ADC, we can minimize its area, and by carefully selecting the specific outputs, we can further boost its power efficiency.

C. ADC-aware Decision Tree Training

As illustrated in Fig. 2b, assuming parallel unary inputs, the logic of the Decision Tree is inherently simplified. Hence, since the hardware-efficiency of this two-level logic is mostly determined by the Decision Tree depth hyperparameter, our training methodology focuses on optimizing the cost of the ADCs. As demonstrated in Section III-B, the hardware efficiency of the ADCs, and consequently of the entire classifier, is determined by the trained parameters of the Decision Tree (i.e., output unary digits of each ADC). Hence, carefully selecting the Decision Tree parameters in an ADC-aware manner, while still achieving high classification accuracy, is essential for enabling power-autonomous printed Decision Trees.

To achieve this, we propose and implement an ADC-aware Decision Tree training approach. Our primary objective is to minimize the number of comparators induced by the ADCs. This is accomplished by minimizing the number of unique inputs involved among the total comparisons (i.e., minimizing the number of ADCs) and, for each remaining input, by minimizing the number of different parameters, it is compared against, in all the comparisons involved in. Our secondary objective is to select more power-efficient values/parameters for each comparison (i.e., optimize the order of the output digits in the ADC). Since feasibility is the foremost requirement for printed ML circuits, prioritizing it over strict accuracy constraints is a typical procedure [3]. Consequently, we also explore the trade-off between some accuracy degradation and the potential for additional hardware gains.

Our ADC-aware training essentially trains a Decision Tree using the *Gini index* [16] cost function. Gini is used to evaluate a split in the dataset. The latter involves one input feature

Algorithm 1 ADC-aware Decision Tree Training Pseudocode

Input: 1) Dataset, 2) Tree Depth, 3) Gini Threshold τ

Output: Trained Decision Tree and Classification accuracy

```

1:  $\mathcal{DT} = \emptyset$  #selected split nodes
2: for  $0 \leq \text{node} < \text{Total nodes}$  do
3:   for  $\forall I_i \in \text{Input Features}$  and  $\forall C$  value in dataset for  $I_i$  do
4:     calculate  $\text{Gini}(I_i, C)$ 
5:    $G = \text{minimum Gini score}$ 
6:    $\mathcal{S} = \{(I_i, C) \mid \text{Gini}(I_i, C) \leq G + \tau, \forall(I_i, C)\}$ 
7:    $\mathcal{S}_Z = \{(I_i, C) \in \mathcal{S} \mid (I_i, C) \in \mathcal{DT}\}$ 
8:    $\mathcal{S}_M = \{(I_i, C) \in \mathcal{S} \mid \exists(I_i, C') \in \mathcal{DT}, C \neq C'\}$ 
9:    $\mathcal{S}_H = \{(I_i, C) \in \mathcal{S} \mid \nexists(I_i, C') \in \mathcal{DT}, \forall C'\}$ 
10:  if  $\mathcal{S}_Z \neq \emptyset$  then
11:     $g_m = \text{calculate minimum Gini score } \forall(I_i, C) \in \mathcal{S}_Z$ 
12:     $\text{split} = \text{random}(\{(I_i, C) \in \mathcal{S}_Z \mid \text{Gini}(I_i, C) = g_m\})$ 
13:  else
14:    if  $\mathcal{S}_M \neq \emptyset$  then  $\mathcal{Z} = \mathcal{S}_M$  else  $\mathcal{Z} = \mathcal{S}_H$ 
15:     $c_m = \min(\{C \mid \forall(I_i, C) \in \mathcal{Z}\})$ 
16:     $\mathcal{U} = \{(I_i, C) \in \mathcal{Z} \mid C = c_m\}$ 
17:     $g_m = \text{calculate minimum Gini score } \forall(I_i, C) \in \mathcal{U}$ 
18:     $\text{split} = \text{random}(\{(I_i, C) \in \mathcal{U} \mid \text{Gini}(I_i, C) = g_m\})$ 
19:   $\mathcal{DT} = \mathcal{DT} \cup \{\text{split}\}$ 

```

(e.g., I_i in Fig. 2a) and a trainable parameter C that will be compared with (i.e., output unary digit of the ADC of I_i). We modify typical Gini-based Decision Tree training to incorporate ADC-awareness as follows. Initially, our algorithm seeks a split node and evaluates the Gini score for all possible combinations between input features and their corresponding values in the training dataset. At this point, ADC-unaware training would randomly select one combination among those with the best (minimum) Gini score. Assuming G is the best Gini score computed, we form a set of candidate split pairs $\mathcal{S} = \{(I_i, C) \mid \text{Gini}(I_i, C) \leq G + \tau\}$, with τ being a training hyperparameter. Next, we group the pairs $(I_i, C) \in \mathcal{S}$ into three sets based on the hardware induced from their selection.

- 1) \mathcal{S}_Z (zero-cost): if (I_i, C) has been previously selected at a split node, it won't require additional hardware, only additional wiring.
- 2) \mathcal{S}_M (medium-cost): if (I_i, C') with $C' \neq C$ has been previously selected at a split node, then the same ADC for I_i is reused, but a new comparator is added to that ADC because a different output digit is required. Since the area of our bespoke ADCs is linear with the number of comparators, all these pairs induce the same overhead, as each of them will add one comparator to one ADC.
- 3) \mathcal{S}_H (high-cost): if no pair (I_i, C') has been previously selected at a split node, i.e., I_i is selected for the first time. These pairs induce the highest (and same) area overhead because a new ADC with one comparator is required.

Among these sets, we select the first non-empty one based on the order listed above. If \mathcal{S}_M or \mathcal{S}_H is chosen, we then identify the (I_i, C) pair in that set that results in the lowest power overhead. This can be accomplished by selecting the (I_i, C) pair with the minimum C value since it will necessitate the lowest-order output digit and, consequently, the induced comparator will have the lowest power consumption (see Section III-B). If multiple pairs feature the same minimum C value, or if \mathcal{S}_Z is chosen, we select the pair with the best Gini score, or random one if the Gini scores are equal. Our

TABLE I
EVALUATION OF THE BASELINE BESPOKE DECISION TREES.

Dataset	Acc (%)	#Comp.	#Inputs	Area (mm ²)		Power (mW)	
				ADCs	Total	ADCs	Total
Whitewine	52.8	207	11	17.3	261.3	5.4	14.6
Cardio	90.6	85	19	22.3	114.4	9.1	12.5
Arrhythmia	62.7	39	21	23.5	79.9	10.0	12.0
Balance-Scale	77.7	15	4	12.9	30.6	2.2	2.9
Vertebral-3C	86.0	7	5	13.6	16.8	2.5	2.8
Seeds	90.5	23	5	13.6	27.3	2.5	3.2
Vertebral-2C	87.1	7	5	13.6	16.4	2.5	2.8
Pendigits	95.0	215	16	20.4	268.7	7.7	17.2

ADC-aware training assumes user-defined fixed depth and τ hyperparameters. Our algorithm identifies the most hardware-efficient split at each node. While our approach is inherently greedy, it introduces hardware-awareness into the traditionally employed, also greedy, Gini-based training. Still, alternative heuristic approaches could also be used. Algorithm 1 provides an abstract overview of our proposed training methodology.

The hyperparameter τ (if higher than 0) may lead to some accuracy degradation, as a pair with a Gini score higher than the best score might be selected. However, τ increases the size of the set \mathcal{S} , thereby increasing the chances of finding a more hardware-efficient (I_i, C). $\tau = 0$ will not affect the accuracy.

IV. RESULTS AND ANALYSIS

In this section, we evaluate our printed Decision Tree classifiers, first examining hardware gains from our bespoke ADC design and ADC-aware Decision Tree training. Evaluation is based on 8 datasets listed in Table I. These datasets are selected for two primary reasons: i) to facilitate direct comparisons with the state-of-the-art [2, 7], and ii) because these datasets utilize sensor inputs suitable for printed applications [2, 6]. The datasets are obtained from the the UCI ML repository [17]. Normalized inputs in the range $[0, 1]$ are used for training/testing with a random 70%/30% split. Synopsys Design Compiler and PrimeTime analyze the digital part of circuits, all operated at 20Hz, a common frequency aligned with typical performance of target PE applications [1, 3].

As our evaluation baseline, we consider the fully parallel bespoke Decision Trees designed as described in [2]. The minimum tree depth (up to 8) that achieves the maximum accuracy is used for each model and the input precision is set to 4 bits, since this is the value delivered close to floating-point accuracy for all datasets. Table I summarizes the accuracy and hardware overheads of the baseline Decision Trees [2]. Specifically, Table I reports the total area and total power requirements for each Decision Tree classifier as well as the respective values for the ADCs. As shown, the average total area and power consumption are 102mm² and 8.5mW, respectively. Notably, all the classifiers exhibit power demands that exceed the capabilities ($> 2\text{mW}$) of printed energy harvesters [18]. Consequently, none of these circuits can be self-powered. Finally, it's worth noting that, on average, for the Decision Trees in Table I, 40% of the total area and 74% of the total power consumption is attributed to the ADCs.

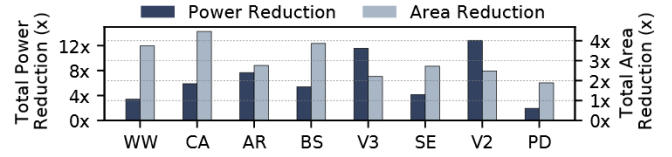


Fig. 4. Total area and power reduction (x) compared to the baseline designs [2] (i.e., vs Table I). For our printed Decision Trees only our proposed bespoke ADCs and parallel unary architecture are considered.

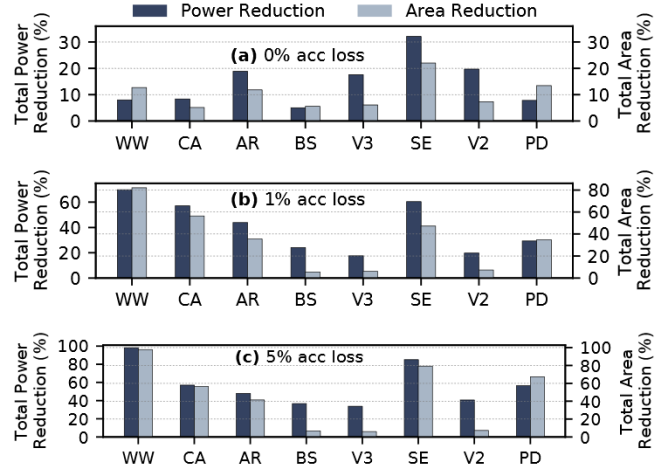


Fig. 5. Evaluation of the additional hardware gains delivered by our ADC-aware training. Total area and power reductions (%) of our printed DTs w/ and w/o our ADC-aware training are compared (i.e., vs Fig. 4). Three accuracy loss constraints are considered: a) 0% (i.e., no accuracy loss), b) 1%, c) 5%.

Fig. 4 depicts the area and power gains achieved by solely considering our bespoke ADCs along with the parallel unary Decision Tree design, i.e., the same ADC-unaware trained model used in [2]. In Fig. 4, values are reported w.r.t. the baseline [2] (see Table I). As shown, given that the overall area and power consumption of the printed Decision trees are predominantly governed by the ADCs, employing our bespoke ADC design delivers substantial hardware gains. Furthermore, as explained in Section III, using our ADCs streamlines the classifier's implementation, resulting in a simplified two-stage logic and additional hardware savings over the baseline [2]. Specifically, compared to [2], the achieved area and power reduction are 3.0x and 6.6x on average, respectively.

Next, in Fig. 5 we investigate the impact of our ADC-aware training. The area and power gains in Fig. 5 are reported w.r.t. the area and power of the designs of Fig. 4 (i.e., over the simplified Decision Trees with our bespoke ADCs). For this analysis, we consider three accuracy loss thresholds: 0%, 1%, and 5%. Specifically, we conduct a brute-force exploration of hyperparameters, including τ values ranging from 0 to 0.03 in increments of 0.005 and depth values from 2 to 8 with a step of 1. The rather simple classification tasks in printed electronics [2] and the independence of different trainings (i.e., they can run in parallel), enable this exploration to be rapidly conducted. The average execution time is only 6 min on an Intel Xeon 6138 server with 256GB RAM. As demonstrated in Fig. 5a, for *no accuracy loss*, our ADC-aware training leads to an average reduction of 11% in area and 15% in power when

TABLE II
EVALUATION OF OUR DECISION TREES FOR UP TO 1% ACCURACY LOSS.

Dataset	Proposed		Reduction vs [2]		Reduction vs [7]	
	Area ¹ (mm ²)	Power ¹ (mW)				
WhiteWine	11.99	1.26	21.8x	11.3x	10.5x	4.3x
Cardio	10.13	0.88	11.3x	14.1x	4.4x	2.4x
Arrhythmia	16.24	0.85	4.9x	14.1x	1.5x	1.3x
Balance-Scale	4.92	0.35	6.2x	8.2x	5.8x	3.6x
Vertebral-3C	2.71	0.17	6.2x	16.2x	3.4x	2.7x
Seeds	3.26	0.27	8.4x	11.9x	1.2x	1.1x
Vertebral-2C	2.22	0.15	7.4x	18.5x	²	²
Pendigits	89.00	6.12	3.0x	2.8x	4.2x	2.6x
Average	17.56	1.26	8.6x	12.2x	4.4x	2.6x

¹Total area and total power, including ADCs. ²Not evaluated in [7].

compared to using the conventional ADC-unaware training. Similarly, for only 5% accuracy loss the average area and power reduction increase to 45% and 57%, respectively.

Finally, Table II evaluates the effectiveness of our co-design framework in generating self-powered printed Decision Trees, considering up to 1% accuracy loss. In Table II, we also compare our Decision Trees against the baseline exact [2] (i.e., Table I) and the approximate ones of [7] (with up to 1% accuracy loss). Note that for [2] conventional 4-bit ADCs are used, whereas for [7], since precision scaling is applied, the smallest suitable conventional ADC for each input is used. To the best of our knowledge, these are the only works that have investigated printed Decision Trees. As shown, our Decision Trees achieve on average 8.6x and 12.2x lower area and power, respectively, compared to [2]. Similarly, compared to [7], the corresponding gains are 4.4x and 2.6x, respectively. Note that in some cases (i.e., BS, V3, and PD), [7] features higher area and power consumption compared to our baseline [2] due to the use of deeper trees in [7] to compensate for the accuracy loss caused by their applied approximation. Concluding, as demonstrated by Table II all our classifiers except for Pendigits feature power consumption well below 2mW, even when accounting the significant cost of ADCs. Pendigits also adhered to the 2mW power constraint but at a 10% accuracy loss. This analysis suggests that our co-design framework can be used to effectively target printed applications with similar complexity to the datasets in Table II, even after considering the cost of sensors, which is negligible compared to the hardware overheads of printed classifiers. For instance, relevant sensors reviewed in [1] for such printed applications consume only 5μW. Hence, for the examined datasets, the power increase due to sensors is less than 0.11mW. As a result, for less than 1% accuracy loss, our framework can efficiently produce printed classifiers, even with 207 comparators and 11 inputs, that demand less than 2mW of power, being thus suitable for printed energy harvester operation [18].

V. CONCLUSION

Printed electronics offer a solution to extend computing into application domains previously untouchable by silicon-based

technologies. In this work, we propose a co-design framework for printed Decision Trees that incorporates fully parallel unary architectures, bespoke ADCs tailored for such architectures, and an ADC-aware training approach aimed at minimizing the overall hardware cost of on-sensor processing. Our circuits achieve substantial area and power savings, paving the way towards digital, self-powered, on-sensor printed ML classifiers.

ACKNOWLEDGMENT

This work is partially supported by EU Horizon research and innovation programme, under project CONVOLVE, grant agreement No 101070374, by the European Research Council (ERC), and co-funded by the H.F.R.I call “Basic research Financing (Horizontal support of all Sciences)” under the National Recovery and Resilience Plan “Greece 2.0” (H.F.R.I. Project Number: 17048)

REFERENCES

- [1] N. Bleier, M. Mubarak, F. Rasheed, J. Aghassi-Hagmann, M. B. Tahoori, and R. Kumar, “Printed microprocessors,” in *Annu. Int. Symp. Computer Architecture (ISCA)*, jun 2020, pp. 213–226.
- [2] M. H. Mubarak *et al.*, “Printed machine learning classifiers,” in *Annu. Int. Symp. Microarchitecture (MICRO)*, 2020, pp. 73–87.
- [3] G. Armeniakos, G. Zervakis, D. Soudris, M. B. Tahoori, and J. Henkel, “Co-design of approximate multilayer perceptron for ultra-resource constrained printed circuits,” *IEEE Trans. Comput.*, pp. 1–8, 2023.
- [4] G. Armeniakos, G. Zervakis, D. Soudris, M. B. Tahoori, and J. Henkel, “Model-to-circuit cross-approximation for printed machine learning classifiers,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 2023.
- [5] A. Kokkinis *et al.*, “Hardware-aware automated neural minimization for printed multilayer perceptrons,” in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2023, pp. 1–2.
- [6] D. D. Weller *et al.*, “Printed stochastic computing neural networks,” in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2021, pp. 914–919.
- [7] K. Balaskas *et al.*, “Approximate decision trees for machine learning classification on tiny printed circuits,” in *Int. Symp. Quality Electronic Design*, 2022, pp. 1–6.
- [8] M. Shafique, R. Hafiz, S. Rehman, W. El-Harouni, and J. Henkel, “Cross-layer approximate computing: From logic to architectures,” in *Design Automation Conference (DAC)*, 2016, pp. 1–6.
- [9] A. Ren *et al.*, “Sc-dcnn: Highly-scalable deep convolutional neural network using stochastic computing,” in *Int. Conf. Architectural Support for Programming Languages and Operating Systems*, 2017, p. 405–418.
- [10] D. Wu, J. Li, R. Yin, H. Hsiao, Y. Kim, and J. S. Miguel, “Ugemm: Unary computing architecture for gemm applications,” in *Int. Symp. Computer Architecture (ISCA)*, 2020, pp. 377–390.
- [11] Z. Cui, *Printed electronics: materials, technologies and applications*. John Wiley & Sons, 2016.
- [12] J. Henkel *et al.*, “Approximate computing and the efficient machine learning expedition,” in *Int. Conf. Computer Aided Design*, 2022.
- [13] G. Tzimpragos, A. Madhavan, D. Vasudevan, D. Strukov, and T. Sherwood, “Boosted race trees for low energy classification,” in *Int. Conf. Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS ’19, 2019, p. 215–228.
- [14] M. H. Najafi, D. J. Lilja, M. D. Riedel, and K. Bazargan, “Low-cost sorting network circuits using unary processing,” *IEEE Trans. VLSI Syst.*, vol. 26, no. 8, pp. 1471–1480, 2018.
- [15] A. Khataei, G. Singh, and K. Bazargan, “Approximate hybrid binary-unary computing with applications in bert language model and image processing,” in *Int. Symp. Field Programmable Gate Arrays (FPGA)*, 2023, p. 165–175.
- [16] M. Jaworski, P. Duda, and L. Rutkowski, “New splitting criteria for decision trees in stationary data streams,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2516–2529, 2018.
- [17] K. N. Markelle Kelly, Rachel Longjohn, “The uci machine learning repository.” [Online]. Available: <https://archive.ics.uci.edu>
- [18] A. Ahmed *et al.*, “Additively manufactured nano-mechanical energy harvesting systems: advancements, potential applications, challenges and future perspectives,” *Nano Convergence*, vol. 8, no. 37, 2021.