Research Report on

Hardware-efficient, quantization-robust in-memory bisector-based binary max-margin classifiers

Full search query: I want to find an implementation-focused, in-memory-compute–oriented study of binary linear maximum-margin classification that uses a decision boundary defined as the perpendicular bisector of the nearest opposite-class points, developing a hardware-efficient and quantization-robust training/inference method for linearly separable data and comparing it against classic baselines (hard/soft-margin SVM, perceptron, logistic regression) in terms of computational efficiency and robustness

Summary

No paper in this set implements a nearest-opposite-pair perpendicular-bisector linear classifier in an in-memory-compute (IMC) framework with quantization-robust training and full baseline comparisons; the closest algorithmic antecedent is the Schlesinger–Kozinec nearest-point separator for convex polytopes [3], while the most relevant IMC/robustness building blocks come from linear SVM/logistic/ADALINE implementations in SRAM/NAND/RRAM and precision theory for margin hyperplanes [5,7,8,10,1,9,6].

What matches the target and what does not

- Strong alignment (implementation building blocks):
    - IMC linear classifiers (SRAM/NAND) with demonstrated energy/throughput gains and, in one case, on-chip training for variation tolerance: DIMA SRAM SVM with on-chip SGD-based adaptation to PVT [5], multi-functional DIMA supporting SVM and others [7], foundational SRAM IMC classifier [8], and DIMA-F for NAND with non-ideality-aware models [10].
    - Precision/quantization guidance for margin hyperplanes, offering analytical bounds and co-assignment of input/weight precisions [6].
    - Device-centric lessons from RRAM crossbars implementing a binarized linear model (ADALINE), highlighting sneak-paths and variability impacts [1].

- Partial alignment (algorithmic proximity to the bisector rule):
    - Nearest-point separation between convex polytopes (SK algorithm) used to build piecewise linear "multiconlitron" classifiers; conceptually related but not a single nearest-pair bisector, and not IMC-oriented [3].

- Missing elements versus the stated goal:
    - No work defines w,b from the perpendicular bisector of the single nearest opposite-class training pair in an IMC setting [1,2,3,4,5,6,7,8,9,10].
    - No quantized-margin–aware pair selection and scaling that explicitly models $\hat{w}$, $\hat{b}$ and device non-idealities during training/mapping [1,2,3,4,5,6,7,8,9,10].
    - No unified comparisons across hard/soft-margin SVM, perceptron, and logistic regression that report both computational efficiency and robustness under quantization/IMC constraints [1,2,3,4,5,6,7,8,9,10].

Algorithmic antecedents relevant to a nearest-pair bisector

- SK nearest-point method between convex polytopes:
    - Computes separating hyperplanes by iteratively finding closest points between convex hulls; used to construct conlitron/multiconlitron classifiers for convexly or linearly separable data [3].
    - Relation to target: shares the nearest-opposing-elements geometry but operates on sets/hulls and yields piecewise linear models rather than a single hyperplane from the closest opposite-class data pair; no hardware/quantization focus [3].

IMC implementations of linear margin classifiers and their robustness techniques

- SRAM-based DIMA lineage:
    - In-memory linear classification in standard 6T SRAM, establishing MAC-in-memory execution [8]; extended to multi-algorithm support with 10× energy and 5.3× throughput gains in dot-product mode [7].
    - Variation-tolerant SVM with on-chip SGD adapts to PVT and analog periphery variations, enabling reduced bitline swing and 2.4× energy reduction over off-chip-trained operation; 42 pJ/decision at 32 M decisions/s (3.12 TOPS/W) [5].

- NAND-based IMC:
  - DIMA-F shows 8–23× energy and 9–15× throughput improvements for linear SVM and kNN vs. conventional external-ASIC processing; provides behavioral models capturing circuit non-idealities [10].

- Other linear IMC exemplars:
  - Logistic regression on an 8T SRAM array with mixed-signal accumulation, validated on MNIST [9].
  - Binarized ADALINE on OxRAM crossbar demonstrates mapping and reveals accuracy degradation from sneak-paths and device variability (67% experiment vs. 78% simulation) [1].

- Takeaway for a bisector-based IMC design:
  - The above establish practical pathways for mapping a single weight vector and bias to memory conductances, differential encoding for signed weights, and the utility of on-chip calibration/training to counter analog non-idealities [5,7,8,10,1,9].

Quantization and precision theory applicable to a bisector classifier

- Precision bounds for margin hyperplanes:
  - Provides analytical conditions on minimum precision for linear margin classifiers and a scheme to trade input versus weight precision to preserve accuracy while reducing resource cost; empirically superior to naive low-precision baselines on MNIST tasks [6].

- Implications:
  - Directly informs scaling/bit-allocation for w,b in an IMC bisector, and supports a quantized-margin objective when choosing the defining pair and the normalization to maximize label-preserving margins under $\mathbb{w}$, $\mathbb{b}$ and input dynamic range [6].

- Complementary modeling:
  - Behavioral non-ideality models for IMC peripheries (e.g., DIMA-F) can be used to calibrate $\mathbb{w}$, $\mathbb{b}$ distributions for robust mapping [10], while on-chip adaptation strategies demonstrate effectiveness against PVT in SRAM IMC [5].

Baseline accelerators and comparisons context

- Digital accelerators (non-IMC) for SVM:
  - Dual-mode online SVM trainer-classifier ASIC with queue-based selection and linear/kernel modes, achieving high operating frequency and low training latency [2].
  - Cluster-partitioning SVM accelerator with algorithm–architecture co-optimizations and substantial latency reductions for training/inference; PE array and dataflow optimized for efficiency [4].

- Comparative insights:
  - These works set strong digital baselines for training/inference latency but do not address analog non-idealities or quantization robustness inherent to IMC [2,4].
  - IMC papers generally compare energy/throughput within their paradigm, with limited cross-algorithm robustness benchmarking; logistic and ADALINE are demonstrated separately without unified comparisons to SVM under quantization [7,9,1,5,10].

Practical lessons and gaps to bridge

- What the corpus enables now:
  - Efficient IMC mapping of linear decision rules with proven energy/throughput benefits [7,8,10,5].
  - Precision-aware design of margin hyperplanes to mitigate low bit-depth effects [6].
  - Calibration/on-chip adaptation frameworks to counter device/circuit variation [5,10], and device-level pitfalls to avoid in RRAM-based realizations [1].

- What is still missing for the target study:
  - A training pipeline that:
    - Identifies the nearest opposite-class pair under a chosen (possibly diagonal/Mahalanobis) metric and constructs the bisector, and
    - Optimizes quantized margin by scaling/bit-slicing/differential encoding given calibrated $\mathbb{w}$, $\mathbb{b}$ and input range [— informed by 6,5,10].
  - A systematic, IMC-grounded benchmarking versus hard/soft-margin SVM, perceptron, and logistic regression that reports:
    - Training cost (nearest-pair search vs. iterative solvers), and
    - Robustness under controlled quantization/noise sweeps and device non-idealities on identical hardware substrates [— informed by 5,7,8,10,9].

- Opportunity:

- Combining SK-style nearest-element geometry as motivation [3] with the DIMA hardware stack [8,7,5,10] and precision theory [6] can yield a hardware-efficient, quantization-robust bisector classifier with minimal training overhead and strong IMC fit—addressing an open niche not covered by current literature.

Categories

Scope alignment and relevance to the target problem

- None of the retrieved papers implement the exact decision rule of "perpendicular bisector of the nearest opposite-class points" and a quantization-robust training/inference flow optimized for in-memory computing with systematic comparisons to SVM, perceptron, and logistic regression.

- The closest algorithmic lineage to a nearest-pair–defined hyperplane is the Schlesinger–Kozinec (SK) nearest-point method for separating convex polytopes, used in constructing piecewise linear "multiconlitron" classifiers; however, it is not in-memory-compute–focused and operates at the level of convex hulls rather than a single nearest opposite-class data pair [3].

- Multiple papers provide in-memory or near-memory implementations of linear margin classifiers (notably linear SVM) and discuss robustness/precision constraints in hardware, which are directly relevant to the implementation and quantization-robustness aspects of the goal [5,7,8,10,1,9,6]. Two papers present digital accelerators for SVM learning with significant architectural optimizations, but without in-memory compute emphasis [2,4].

Relevance overview (to the specified goal)

Reference	Uses nearest opposite-class pair bisector	In-memory compute focus	Linear max-margin focus	Quantization/variation robustness methods	Baseline comparisons (perceptron/logistic/SVM)
[3] SK algorithm, multiconlitron	Partial (nearest points between convex polytopes; not single-pair bisector)	No	Yes (separable case via convex polytopes)	No	Some algorithmic comparisons within piecewise-linear family (not perceptron/logistic/SVM)
[5] DIMA SRAM with on-chip training (SVM)	No	Yes	Yes (linear SVM)	Yes (on-chip adaptation to PVT; robust under reduced BL swing)	No (focuses on SVM)
[7] DIMA SRAM inference processor	No	Yes	Yes (supports SVM among others)	Partially implied by analog/mixed-signal design; no explicit quantization theory	No (multi-algorithm support rather than comparative study)
[8] In-memory ML classifier in 6T SRAM	No	Yes	Likely linear classifier support (used by [5,7])	Not specified in abstract; baseline work enabling [5,7]	No
[10] DIMA-F for NAND; SVM/kNN study	No	Yes	Yes (linear SVM studied)	Yes (behavioral models with circuit non-idealities/variations)	No
[1] Binarized-ADALINE on OxRAM crossbar	No	Yes	Linear classifier (ADALINE)	Notes device variability/sneak-path issues; shows accuracy drop	No
[9] Logistic regression on 8T SRAM	No	Yes	Linear (logistic)	Not emphasized	No
[2] Dual-mode SVM trainer/classifier ASIC	No	Accelerator (not in-memory)	Yes (linear and non-linear SVM)	Not the focus	Limited (accuracy/TPR/TNR; SVM modes)
[4] SVM learning accelerator (CP-SVM)	No	Accelerator (not in-memory)	Yes (linear and kernel SVM)	Architectural, not quantization	No
[6] Precision bounds for margin hyperplanes	No	Applicable to embedded/hardware	Yes (general margin hyperplanes)	Yes (analytical precision bounds; precision reduction scheme)	No

What each work contributes toward the target

- Nearest-pair bisector lineage (algorithmic, non-IMC): The SK nearest-point method constructs separating hyperplanes by iteratively finding nearest points between convex polytopes; this is conceptually related to bisector-based hyperplanes but not the single nearest opposite-class data-pair rule, and not IMC-oriented [3].

- In-memory SVM with robustness: Demonstrates on-chip training for a linear SVM in a deep in-memory SRAM architecture that adapts to PVT variations, improving energy by 2.4× over off-chip trained operation at reduced bitline swing; 42 pJ/decision at

32 M decisions/s (3.12 TOPS/W) [5]. DIMA platforms supporting SVM inference are shown in 6T SRAM [7,8], and extended to NAND flash with behavioral non-ideality modeling [10].

- In-memory linear classifiers beyond SVM: Binarized ADALINE on OxRAM crossbar shows end-to-end mapping and highlights hardware pitfalls (sneak-paths, variability) and their accuracy impact [1]. Logistic regression classification implemented on a custom 8T SRAM array using DAC-driven weights demonstrates feasibility for multi-class tasks [9].

- Precision/quantization theory: Provides analytical bounds on precision for margin hyperplanes and a precision-reduction scheme balancing input/weight quantization, showing improved accuracy versus naive low-precision baselines at lower arithmetic cost [6].

- SVM accelerators (non-IMC): Digital architectures achieve large training/inference latency reductions via algorithm-architecture co-optimization (CP-SVM) and dual-mode training/classification pipelines [2,4]. These support linear and kernel SVM and present detailed hardware mapping and efficiency results.

Comparative technical matrix for in-memory/near-memory linear classifiers

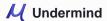| Aspect	[5] DIMA SRAM SVM	[7] DIMA SRAM multi-function	[8] 6T SRAM ML classifier	[10] DIMA-F (NAND)	[1] OxRAM binarized ADALINE	[9] 8T SRAM logistic
Memory substrate	6T SRAM	6T SRAM	6T SRAM	NAND flash	OxRAM crossbar	8T SRAM
Algorithm	Linear SVM (train+infer)	SVM, template matching, kNN, matched filter	ML classifier (enables [5,7])	Linear SVM, kNN (study context)	Binarized ADALINE	Logistic regression
Training locus	On-chip SGD-based	Inference processor	Not specified here	Off-chip analysis/simulation context	Ex-situ	Off-chip training; weights to DAC
Robustness approach	On-chip training adapts to PVT; robust at reduced bitline swing; 2.4× energy reduction vs off-chip trained at low BL swing [5]	Mixed-signal periphery; energy/throughput gains; robustness mechanisms not the core focus [7]	Foundational IMC classifier demo [8]	Behavioral models including non-idealities/variations; energy/throughput gains vs external ASIC [10]	Notes variability and sneak-paths as causes of accuracy drop [1]	Focus on architecture; robustness not central [9]
Signed-weight realization	Mixed-signal periphery; deep in-memory MAC [5]	Analog low-swing periphery, simultaneous multi-row access [7]	Mixed-signal periphery [8]	Analog periphery near NAND [10]	Binary conductance states; weight mapping technique [1]	DAC-driven classify wordline; sum on bitlines [9]
Reported efficiency	42 pJ/decision; 32 M decisions/s; 3.12 TOPS/W [5]	10× energy, 5.3× throughput gains (dot-product mode) vs conventional digital baselines [7]	Foundational; cited by [5,7] [8]	8–23× energy and 9–15× throughput improvements; EDP up to 345× vs external ASIC [10]	Experimental accuracy 67% (vs 78% in sim) due to variability/sneak-paths [1]	Operates at similar frequency in SRAM and classify modes with low area overhead [9]

Notes:

- [7] quantitatively reports simultaneous energy and throughput gains in dot-product mode, but does not provide absolute pJ/op like [5].

- [10] provides energy/throughput improvement factors vs a conventional NAND+external-ASIC baseline, with models incorporating circuit non-idealities.

Robustness and precision insights applicable to a bisector-based design

- Precision bounds for margin hyperplanes: Analytical conditions and a precision-reduction scheme show that careful co-assignment of input and weight precisions can outperform standard fixed-precision implementations at substantially lower arithmetic cost [6]. This is directly relevant to selecting scaling and bit allocation for a simple linear hyperplane such as a nearest-pair bisector.

- In-situ adaptation to non-idealities: On-chip SGD training in DIMA compensates for PVT and analog periphery variations, enabling lower bitline swings and energy reductions while maintaining accuracy [5]. A bisector-based classifier could benefit from lightweight on-chip calibration to estimate effective $\bar{w}$, $\bar{b}$, akin to [5], and behavioral modeling as in [10].

- Device/crossbar pitfalls: OxRAM crossbar implementation of a binarized linear classifier revealed accuracy deficits due to sneak-paths and programming variability [1]. For a nearest-pair bisector mapped to crossbar conductances, differential encoding and bit-slicing, along with device-aware scaling, would be necessary countermeasures; such techniques are consistent with the precision/robustness discussion in [6] and the mixed-signal design practices in [5,7,10].

SVM accelerators versus in-memory approaches (context for efficiency comparisons)

- In-memory compute emphasizes minimizing data movement and exploiting analog accumulation within memory arrays, with demonstrated energy/throughput gains for linear classifiers [5,7,8,10].

- Algorithm-architecture co-optimized SVM accelerators (not IMC) reduce computational latency via clustering/partitioning and kernel transformation, with detailed PE-array design space exploration and sparsity-aware scheduling [4], and provide dual-mode (linear/non-linear) end-to-end trainer-classifier pipelines with very high operating frequencies and short training

latencies [2]. These establish digital baselines for computational efficiency but do not address analog non-idealities or quantization robustness in memory arrays.

Gaps relative to the target and how the literature informs a future design

- Missing nearest-pair bisector in IMC: No paper realizes the perpendicular bisector of the nearest opposite-class points as the decision boundary in an IMC setting [1,2,3,4,5,6,7,8,9,10]. The SK algorithm [3] is algorithmically adjacent (nearest points between convex polytopes) but lacks the single-pair bisector constraint and any IMC/quantization treatment.
- Missing comprehensive baseline comparisons: None of the IMC papers compare a linear classifier against perceptron, logistic regression, and SVM with explicit computational efficiency and quantization robustness analyses; works tend to focus on a single algorithm or a small set without systematic robustness benchmarking [1,5,7,8,9,10].
- What can be transferred:
  - Energy/throughput techniques and calibration methods from DIMA SRAM/NAND [5,7,8,10].
  - Precision assignment and quantization bounds for margin hyperplanes [6].
  - Awareness of device non-idealities from OxRAM crossbar results [1].
  - Digital accelerator baselines for efficiency/latency framing [2,4].

Summary comparison table (alignment to desired elements)

Desired element	Evidence present in corpus	Key references
Decision boundary: perpendicular bisector of nearest opposite-class points	Not implemented; nearest-point separation between convex polytopes exists (non-IMC)	[3]
In-memory compute implementation	Yes, multiple linear classifiers including SVM, logistic, ADALINE	[5,7,8,10,1,9]
Quantization/device-variation robustness with theory or models	Yes (analytical precision bounds; behavioral non-ideality models; on-chip adaptation)	[6,10,5]
Hardware-efficient training/inference for linearly separable data	Yes for linear SVM (on-chip training; accelerators), logistic/ADALINE (IMC inference)	[5,2,4,9,1]
Comparative evaluation vs SVM, perceptron, logistic (accuracy, efficiency, robustness)	No single work provides the full triad comparison under quantization/IMC constraints	—

Collectively, the corpus provides the implementation and robustness "building blocks" (IMC SRAM/NAND platforms, on-chip adaptation, precision theory), but lacks a study that pairs the nearest opposite-class pair bisector with an IMC-centric, quantization-robust pipeline and comprehensive baseline comparisons. The most algorithmically proximate method is the SK nearest-point separation between convex polytopes [3], while the most implementation-relevant robustness insights stem from DIMA-based SVMs and precision analyses [5,6,10].

Timeline

Timeline of key milestones

- 2013 — Convex-hull separation and nearest-point hyperplanes
  - Introduces the Schlesinger–Kozinec (SK) nearest-point method to compute separating hyperplanes between convex polytopes and builds piecewise-linear "conlitron/multiconlitron" classifiers [3]. This anchors an algorithmic lineage for hyperplanes derived from nearest opposing sets, conceptually related to perpendicular bisectors, but it targets convex hulls (sets) rather than the single nearest opposite-class pair emphasized in the present goal.
- 2017 — First widely cited IMC linear classifier in standard 6T SRAM
  - Demonstrates in-memory computation of a linear classifier (including SVM-like operations) directly in a standard 6T SRAM array, establishing the deep in-memory architecture (DIMA) paradigm for mixed-signal dot products in memory periphery [8]. This is a foundational hardware milestone for linear margin classifiers in IMC.
- 2018 — Generalization and robustness emphasis in SRAM and NAND
  - Multi-functional DIMA with support for SVM, template matching, k-NN, and matched filter, highlighting energy/throughput gains and mixed-signal periphery design [7].
  - Variation-tolerant IMC classifier with on-chip SGD training for a linear SVM, demonstrating adaptation to PVT and device-level analog non-idealities; shows energy benefits from reduced BL swing enabled by on-chip training [5].
  - Extends DIMA concepts to NAND flash (DIMA-F) for SVM and k-NN tasks, quantifying large energy–throughput gains vs. conventional off-memory processing [10].
- 2019 — Precision analysis for margin hyperplanes and logistic-regression IMC
  -

Provides analytical bounds on minimum precision needed for margin hyperplane classifiers and a method to co-allocate input and weight precision to preserve accuracy under low bit-width, directly addressing quantization robustness for linear margin models [6].

- Presents an SRAM-based in-memory logistic regression classifier with mixed-signal implementation details and MNIST demonstration [9].

- 2020 — RRAM crossbar demonstration of binarized linear model
  - OxRAM crossbar implementation of a binarized ADALINE classifier; focuses on VMM mapping, weight programming, and highlights practical pitfalls (sneak paths, variability) degrading experimental accuracy [1].

- 2023–2024 — SVM-dedicated accelerators emphasizing end-to-end efficiency
  - ASIC architectures for SVM training/inference: dual-mode (linear/nonlinear) online trainer-classifier with specialized algorithmic/architectural optimizations [2], and a highly energy-efficient SVM learning accelerator with cluster-partitioning SVM and PE-array co-optimization [4]. These works target digital/ASIC acceleration of SVMs rather than mixed-signal IMC but mark the maturing of hardware for linear margin methods.

Thematic evolution and trends

- From enabling IMC primitives to robust linear classifiers
  - Early SRAM-based IMC established feasibility of in-situ MACs for linear models [8], rapidly followed by multi-functionality and mixed-signal periphery co-design [7]. Subsequent work emphasized robustness to analog variation with on-chip training [5] and extended the architecture beyond SRAM [10]. This trajectory shows a steady shift from proof-of-concept to robustness- and efficiency-aware classifier deployment in IMC.

- From inference-only to on-chip adaptation and precision theory
  - The field moved from off-chip training and on-chip inference [8,7,9,1] to on-chip training that compensates device and PVT variation [5]. In parallel, theory on precision requirements for margin hyperplanes emerged [6], providing principled guidelines for quantization—directly relevant to margin-preservation in mixed-signal or low-bit settings.

- Hardware specialization paths diverge
  - Two lines are visible:
    - Mixed-signal IMC in memories (SRAM, NAND, RRAM), prioritizing analog MAC efficiency, calibration, and tolerance to non-idealities for linear models [8,7,5,10,1,9].
    - Digital ASIC accelerators that specialize SVM training/inference via algorithm–architecture co-design, lowering latency and power without analog non-idealities [2,4].

  - This split suggests complementary opportunities: IMC excels for dense MAC with careful robustness design; digital accelerators excel for flexible, precise SVM optimization.

- Quantization and non-idealities as first-class constraints
  - Explicit treatment of precision and variation moved from empirical calibration to formal analysis [5,6]. Works report techniques like reduced BL swing with adaptive training [5], mixed-signal periphery linearization [7], and weight mapping/programming in RRAM with attention to sneak paths and variability [1]. However, model-level quantized-margin optimization strategies (e.g., scaling to maximize discrete margin under $\hat{w}$, $\hat{b}$) remain underexplored in these implementations.

- Limited adoption of nearest-opposite-pair bisector decision rules
  - None of the IMC-focused works construct the classifier explicitly as the perpendicular bisector of the nearest opposite-class pair. SK-style nearest-set separation [3] is algorithmically akin but operates on convex hulls and for piecewise-linear models, not the single-pair bisector intended here. SVM-focused works tackle classical hard/soft-margin optimization [5,2,4] rather than the two-point bisector variant.

Collaborations and influential threads

- Shanbhag–Verma–collaborators (Princeton/UIUC and co-authors)
  - Core DIMA lineage: seminal IMC classifier in 6T SRAM [8], multi-functional DIMA [7], on-chip training for variation tolerance [5], and NAND extension [10]. This group also contributed precision theory for margin hyperplanes [6].
  - Impact: Established the architectural blueprint for mixed-signal IMC classifiers; tied algorithmic choices to circuit non-idealities; advanced understanding of precision allocation for margin preservation.

- SVM accelerator community (independent groups)
  - Dual-mode SVM trainer/classifier ASIC with queue-based selection and efficient kernel support [2]; CP-SVM learning accelerator with extensive PE-array and dataflow optimizations [4].
  - Impact: Demonstrated that tailored SVM algorithms can reach high energy efficiency and low latency in digital ASICs, complementing IMC efforts with scalable training throughput.

- Other device-centric demonstrations
  - OxRAM binarized ADALINE [1] provides a concrete RRAM crossbar pipeline, surfacing practical programming and array-level issues that motivate robustness-aware training/mapping.

- Algorithmic antecedents
  - SK nearest-point algorithm for convex polytopes [3] represents a conceptual bridge from geometric separation via nearest opposing elements to piecewise-linear classifiers; however, it has not been coupled to IMC hardware or to quantization-aware bisector training.

Comparative evaluation practices over time

- Baselines and metrics
  - IMC works typically report accuracy, energy, and throughput; when algorithmic baselines are compared, they emphasize SVM vs. template matching/k-NN within the same hardware [7] or SVM variants within accelerator designs [2,4]. Explicit, side-by-side comparisons against perceptron and logistic regression under quantization/variation are sparse. Logistic-regression IMC demonstrations exist [9] but are not co-evaluated with SVM in a unified robustness framework.
  - Robustness reporting evolved from qualitative tolerance under reduced BL swing [5] to more principled precision–accuracy trade-offs [6], yet evaluations rarely include adversarial quantization sweeps or w/b-calibrated decision-boundary stability analyses across models.

Gaps relative to the target topic and implications for future work

- Missing nearest opposite-class pair bisector in IMC
  - No surveyed work implements a binary linear classifier whose boundary is explicitly the perpendicular bisector of the nearest opposite-class data points, nor a Mahalanobis-generalized bisector implemented for IMC [1,2,3,4,5,6,7,8,9,10]. SK-based nearest-set methods [3] are the closest conceptual relatives but differ substantially from the single-pair construction.

- Underdeveloped quantized-margin–aware training for IMC linear models
  - While [5,6] address variation tolerance and precision bounds for general margin hyperplanes, there is no reported pipeline that:
    - Selects the defining pair to maximize quantized margin under anticipated w, b and input dynamic range,
    - Co-optimizes differential encoding, bit-slicing, and scaling to guarantee label preservation for separable training data,
    - Benchmarks robustness vs. hard/soft-margin SVM, perceptron, and logistic regression on the same IMC substrate.

- Opportunity suggested by existing threads
  - The DIMA lineage provides the hardware and calibration infrastructure [8,7,5,10], and [6] supplies analytical tools for precision assignment in margin classifiers. Combining these with a nearest-pair bisector training rule and explicit quantized-margin optimization would fill a clear niche.
  - Digital SVM accelerators [2,4] set comparative baselines for training latency/energy; an IMC bisector approach could claim near-zero on-chip training cost and single-shot model instantiation, enabling compelling efficiency comparisons.

- Likely near-term trends
  - Integration of margin-aware precision scaling into IMC toolflows (informed by [6]) and more rigorous robustness metrics that incorporate calibrated w/b.
  - Exploration of simplified margin-maximizing surrogates that are hardware friendly (e.g., pairwise bisectors with diagonal Mahalanobis scaling) to bridge the gap between pure SVM and hardware efficiency.
  - Cross-technology demonstrations (SRAM, RRAM) of quantization-robust linear classifiers with decision rules explicitly designed for IMC constraints, accompanied by standardized comparisons to SVM/perceptron/logistic baselines.

Foundational Work

Which papers form the foundational references on this topic?

The below table shows the resources that are most often cited by the relevant papers on this topic. This is measured by the **reference rate**, which is the fraction of relevant papers that cite a resource. Use this table to determine the most important core papers to be familiar with if you want to deeply understand this topic. Some of these core papers may not be directly relevant to the topic, but provide important context.

Ref.	Reference Rate	Title	Cited By These Relevant Papers
[8]	0.18	In-Memory Computation of a Machine-Learning Classifier in a Standard 6T SRAM Array	[5, 7, 16, 19, 20]

[89]	0.16	An energy-efficient VLSI architecture for pattern recognition via deep embedding of computation in SRAM	[5, 7, 8, 20]
[129]	0.13	Support-Vector Networks	[5, 6, 21]
[7]	0.13	A Multi-Functional In-Memory Inference Processor Using a Standard 6T SRAM Array	[5, 16, 19]
[36]	0.12	Minimum precision requirements for the SVM-SGD learning algorithm	[5, 6, 21]
[141]	0.12	Deep Learning with Limited Numerical Precision	[6, 7, 21]
[142]	0.11	Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks	[5, 7, 21]
[124]	0.10	LIBSVM: A library for support vector machines	[6, 12, 21, 31]
[46]	0.10	A 19.4 nJ/decision 364K decisions/s in-memory random forest classifier in 6T SRAM array	[5, 7]
[39]	0.09	A 481pJ/decision 3.4M decision/s Multifunctional Deep In-memory Inference Processor using Standard 6T SRAM Array	[7, 8]
[13]	0.08	A machine-learning classifier implemented in a standard 6T SRAM array	[8, 16, 32]
[131]	0.07	BinaryNet: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1	[6, 21]
[143]	0.07	Bitwise Neural Networks	[6, 21]
[144]	0.07	XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks	[6, 21]
[145]	0.07	Finite-precision analysis of the pipelined strength-reduced adaptive filter	[6, 21]
[146]	0.07	A 4 + 2T SRAM for Searching and In-Memory Computing With 0.3-V V_{DDmin}	[16, 19]
[147]	0.07	Error Adaptive Classifier Boosting (EACB): Leveraging Data-Driven Training Towards Hardware Resilience for Signal Inference	[8, 20]
[148]	0.06	Deep learning with COTS HPC systems	[6, 21]
[149]	0.06	Low-energy Formulations of Support Vector Machine Kernel Functions for Biomedical Sensor Applications	[6, 21]
[150]	0.06	A roundoff error analysis of the LMS adaptive algorithm	[6, 21]

Adjacent Work

Which papers cite the same foundational papers as relevant papers?

Use this table to discover related papers on adjacent topics, to gain a broader understanding of the field and help generate ideas for useful new research directions.

Ref.	Adjacency score	Title	References These Foundational Papers
[78]	0.19	Application Driven Memory - Circuits and Architecture	[7, 13, 64, 96, 103, 134]
[16]	0.17	An Energy Efficient In-Memory Computing Machine Learning Classifier Scheme	[7, 8, 13, 64, 96, 103]
[19]	0.12	Energy-efficient and reliable in-memory classifier for machine-learning applications	[7, 8, 64, 96]
[97]	0.05	Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM)	[64, 129]
[82]	0.05	An On-Chip Trainable and Scalable In-Memory ANN Architecture for AI/ML Applications	[5, 7, 8]
[164]	0.05	Incremental and parallel proximal SVM algorithm tailored on the Jetson Nano for the ImageNet challenge	[64]
[165]	0.05	Predicting thermophysical properties of dialkylimidazolium ionic liquids from sigma profiles	[64]
[166]	0.05	Comparison between fuzzy kernel k-medoids using radial basis function kernel and polynomial kernel function in hepatitis classification	[64]

[167]	0.05	Estimation of particulate matter (PM2.5, PM10) concentration and its variation over urban sites in Bangladesh	[64]
[168]	0.05	Critical exponents for total positivity, individual kernel encoders, and the Jain-Karlin-Schoenberg kernel.	[64]
[169]	0.05	Assessing the impact of parameters tuning in ensemble based breast Cancer classification	[64]
[170]	0.05	A model to predict ammonia emission using a modified genetic artificial neural network: Analyzing cement mixed with fly ash from a coal-fired power plant	[64]
[171]	0.05	N-GlyDE: a two-stage N-linked glycosylation site prediction incorporating gapped dipeptides and pattern-based encoding	[64]
[172]	0.05	Automatic Learning Algorithms for Local Support Vector Machines	[64]
[173]	0.05	Data scarcity, robustness and extreme multi-label classification	[64]
[174]	0.05	Gene selection via BPSO and Backward generation for cancer classification	[64]
[175]	0.05	Automatic Hyper-parameters Tuning for Local Support Vector Machines	[64]
[176]	0.05	An Efficient Alternating Newton Method for Learning Factorization Machines	[64]
[177]	0.05	Galvanic Skin Response Data Classification for Emotion Detection	[64]
[178]	0.05	Recurrent Neural Network for Partial Discharge Diagnosis in Gas-Insulated Switchgear	[64]

References

[1] Methodology for Realizing VMM with Binary RRAM Arrays: Experimental Demonstration of Binarized-ADALINE using OxRAM Crossbar
Sandeep Kaur Kingra, ..., and Manan Suri. 2020 IEEE International Symposium on Circuits and Systems (ISCAS), 2020. 7 citations.
54% Topic Match

Demonstrates a hardware-mapped binarized linear classifier (ADALINE) on OxRAM crossbar.
Implements ex-situ trained binary weights mapped/programmed into an 8×8 OxRAM array to perform VMM for inference, with experimental and simulated accuracy reported.
Does not use the perpendicular-bisector-of-nearest-opposite-points classifier or max-margin/bisector construction; focuses on binarized weights, RRAM programming, sneak-path/device variability, and reports no comparisons to SVM/perceptron/logistic or explicit quantization-robust margin analysis.

[2] Design and Implementation of Dual-Mode Support Vector Machine (SVM) Trainer and Classifier Chip Architecture for Human Disease Detection Applications
Xin-Yu Shih, ..., and Ming-Xian Cai. IEEE Transactions on Circuits and Systems I: Regular Papers, 2023. 2 citations.
44% Topic Match

Proposes a dual-mode SVM trainer/classifier chip (linear and non-linear).
Implements on-chip online SVM training and inference with hardware techniques (QFS, DLNK, LLU) in TSMC 40nm, low area/power, high throughput.
Relevance notes: focuses on SVM QP/SMO-style training (not nearest-opposite-pair bisector), lacks in-memory analog crossbar, device quantization/noise modeling, or explicit comparisons to perceptron/logistic/regarding quantization-robustness.

[3] Construction of Multiconlitron Using SK Algorithm
Leng Qiangku. Journal of Frontiers of Computer Science and Technology, 2013. 0 citations.
42% Topic Match

Proposes a nearest-point–based method (SK algorithm) to compute separating hyperplanes.
Uses the Schlesinger–Kozinec algorithm to find the hyperplane between two convex polytopes (nearest opposite-class points) for separable data.
Relevant? Partially: addresses nearest-point bisector between convex hulls (not explicitly single-point nearest opposite-class pair), lacks in-memory compute, hardware/quantization discussion, and no comparison against SVM/perceptron/logistic or device-robust quantization strategies.

[4] A 73.8k-Inference/mJ SVM Learning Accelerator for Brain Pattern Recognition
Tzu-Wei Tong, ..., and Chia-Hsiang Yang. IEEE Journal of Solid-State Circuits, 2024. 0 citations.
35% Topic Match

Demonstrates an energy-efficient SVM learning accelerator for brain-pattern recognition.
Achieves this via a cluster-partitioning SVM (CP-SVM), kernel transformation, sparsity-aware skipping, PE-array co-optimization, and data scheduling to cut training/inference latency and hardware cost.
Relevance notes: focuses on SVM/QP-style methods and accelerator architecture (40 nm CMOS) but does not use the perpendicular-bisector nearest-opposite-point classifier, nor emphasize in-memory compute, quantization-robust bisector selection, or direct comparisons to perceptron/logistic regression under device noise.

[5] A Variation-Tolerant In-Memory Machine Learning Classifier via On-Chip Training
Sujan Kumar Gonugondla, ..., and Naresh R Shanbhag. IEEE Journal of Solid-State Circuits, 2018. 85 citations.
30% Topic Match

Demonstrates an in-memory, on-chip trained binary classifier robust to device variation.
Implements SGD-based on-chip training in a 6T SRAM deep in-memory architecture (DIMA) to learn chip-specific SVM weights, reducing PVT sensitivity.
Relevant partly: focuses on in-memory SVM and quantization/variation robustness and compares on-/off-chip training energy, but does NOT use the perpendicular-bisector-nearest-opposite-point classifier nor explicit nearest-pair bisector selection or direct comparisons to perceptron/logistic regression.

[6] Minimum Precision Requirements of General Margin Hyperplane Classifiers
Charbel Sakr, ..., and Naresh R Shanbhag. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2019. 2 citations.
27% Topic Match

Analyzes minimum precision needs for margin hyperplane classifiers.
Derives analytical bounds and a precision-reduction scheme; validates via simulations (e.g., MNIST 2-vs-4) showing accuracy/resource gains.
Focuses on quantization/precision trade-offs for linear margin classifiers (SVM-like); lacks nearest-opposite-point bisector, in-memory compute mapping, hardware device non-idealities, or explicit comparisons to perceptron/logistic/SVM runtime/robustness under analog noise.

[7] A Multi-Functional In-Memory Inference Processor Using a Standard 6T SRAM Array
Mingu Kang, ..., and Naresh R Shanbhag. IEEE Journal of Solid-State Circuits, 2018. 170 citations.
25% Topic Match

Demonstrates an in-memory inference IC accelerating linear classifiers and template matching.
Implements dot-product–centric modes (SVM, template matching, k-NN, matched filter) in a 6T SRAM deep in-memory architecture to boost energy and throughput.
Relevant for hardware-efficient inference (single-dot-product mapping, bias via periphery) but does not focus on the nearest-opposite-point bisector classifier, quantization-robust training, or direct comparisons vs. perceptron/logistic/SVM robustness under device noise; training/quantization-aware pair selection appears absent.

[8] In-Memory Computation of a Machine-Learning Classifier in a Standard 6T SRAM Array
Jintao Zhang, ..., and N. Verma. IEEE Journal of Solid-State Circuits, 2017. 375 citations.
23% Topic Match

No summary or abstract available

[9] An In-Memory Architecture for Machine Learning Classifier Using Logistic Regression
Prasanna Kumar Saragada, ..., and B. P. Das. 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), 2019. 1 citations.
22% Topic Match

Demonstrates an in-memory logistic-regression classifier implemented in 8T SRAM.
Implements binary-input logistic regression by mapping weights to DACs, using SRAM columns to compute dot-products; prototyped 128×128 SRAM in 65 nm and evaluated on downsampled/binarized MNIST.
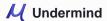Relevant? Partially: it's in-memory and hardware-focused for a linear binary/multi-class classifier and discusses DAC/analog inference, but it uses logistic loss (not bisector-of-nearest-opposite-pair), lacks quantization robust margin analysis, and does not compare to SVM/perceptron or address bisector-based max-margin methods.

[10] Title Energy-Efficient Deep In-memory Architecture for NAND Flash Memories
Sujan Kumar Gonugondla, ..., and Naresh R Shanbhag. Journal Not Provided, 2018. 0 citations.
20% Topic Match

Demonstrates an in-memory NAND-flash architecture for ML inference (SVM, k-NN).
Implements analog periphery processing in NAND flash (DIMA-F) with behavioral models, showing energy/throughput gains for linear SVM and k-NN face tasks.
Relevance notes: targets inference (not training) on NAND flash, models circuit non-idealities and variations, compares energy/throughput vs conventional ASIC but lacks nearest-opposite-pair bisector classifier, quantization-robust bisector training, and explicit comparisons to perceptron/logistic/SVM robustness under weight/bias quantization.

[11] A Hardware-Efficient ADMM-Based SVM Training Algorithm for Edge Computing
Shuo-An Huang and Chia-Hsiang Yang. ArXiv, 2019. 8 citations.
18% Topic Match

Abstract: This work demonstrates a hardware-efficient support vector machine (SVM) training algorithm via the alternative direction method of multipliers (ADMM) optimizer. Low-rank approximation is exploited to reduce the dimension of the kernel matrix by employing the Nystr\"{o}m method. Verified in four datasets, the proposed ADMM-based training algorithm with rank approximation reduces 32$\times$ of matrix dimension with only 2% drop in inference accuracy. Compared to the conventional sequential minimal optimization (SMO) algorithm, the ADMM-based training algorithm is able to achieve a 9.8$\times$10^7 shorter latency for training 2048 samples. Hardware design techniques, including pre-computation and memory sharing, are proposed to reduce the computational...

[12] A Machine Learning Accelerator In-Memory for Energy Harvesting
Salonik Resch, ..., and Ulya R. Karpuzcu. ArXiv, 2019. 1 citations.
16% Topic Match

Abstract: There is increasing demand to bring machine learning capabilities to low power devices. By integrating the computational power of machine learning with the deployment capabilities of low power devices, a number of new applications become possible. In some applications, such devices will not even have a battery, and must rely solely on energy harvesting techniques. This puts extreme constraints on the hardware, which must be energy efficient and capable of tolerating interruptions due to power outages. Here, as a representative example, we propose an in-memory support vector machine learning accelerator utilizing non-volatile spintronic memory. The combination of processing-in-memory and non-volatility...

[13] A machine-learning classifier implemented in a standard 6T SRAM array
Jintao Zhang, ..., and N. Verma. 2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits), 2016. 130 citations.
15% Topic Match

No summary or abstract available

[14] An Efficient FPGA-Based Hardware Accelerator for Convex Optimization-Based SVM Classifier for Machine Learning on Embedded Platforms
Srikanth Ramadurgam and Darshika G. Perera. Electronics, 2021. 20 citations.
14% Topic Match

Abstract: Machine learning is becoming the cornerstones of smart and autonomous systems. Machine learning algorithms can be categorized into supervised learning (classification) and unsupervised learning (clustering). Among many classification algorithms, the Support Vector Machine (SVM) classifier is one of the most commonly used machine learning algorithms. By incorporating convex optimization techniques into the SVM classifier, we can further enhance the accuracy and classification process of the SVM by finding the optimal solution. Many machine learning algorithms, including SVM classification, are compute-intensive and data-intensive, requiring significant processing power. Furthermore, many machine learning algorithms have found their way into portable and embedded devices,...

[15] Training Fixed-Point Classifiers for On-Chip Low-Power Implementation
H. Albalawi, ..., and Xin Li. ACM Transactions on Design Automation of Electronic Systems (TODAES), 2017. 5 citations.
13% Topic Match

No summary or abstract available

[16] An Energy Efficient In-Memory Computing Machine Learning Classifier Scheme
Shixiong Jiang, ..., and R. Sridhar. 2019 32nd International Conference on VLSI Design and 2019 18th International Conference on Embedded Systems (VLSID), 2019. 0 citations.
12% Topic Match

Abstract: Large-scale machine learning (ML) algorithms require extensive memory interactions. Managing or preventing data movement can significantly increase the speed and efficiency of many ML tasks. Towards this end, we devise an energy efficient in-memory computing kernel for a ML linear classifier and a prototype is designed. Compared with another in-memory computing kernel for ML applications [1], we achieve a power savings of over 6.4 times than a conventional discrete system while improving reliability by 54.67%. We employ a split-data-aware technique to manage process, voltage and temperature variations. We utilize a trimodal architecture with hierarchical tree structure to further decrease power...

[17] Cross-Entropy Loss Leads To Poor Margins
Kamil Nar, ..., and K. Ramchandran. Unknown journal, 2018. 12 citations.
11% Topic Match

Abstract: Neural networks could misclassify inputs that are slightly different from their training data, which indicates a small margin between their decision boundaries and the training dataset. In this work, we study the binary classification of linearly separable datasets and show that linear classifiers could also have decision boundaries that lie close to their training dataset if cross-entropy loss is used for training. In particular, we show that if the features of the training dataset lie in a low-dimensional affine subspace and the cross-entropy loss is minimized by using a gradient method, the margin between the training points and the decision...

[18] Deep In-Memory Architectures for Machine Learning–Accuracy Versus Efficiency Trade-Offs
Mingu Kang, ..., and Naresh R Shanbhag. IEEE Transactions on Circuits and Systems I: Regular Papers, 2020. 15 citations.
10% Topic Match

Abstract: In-memory architectures, in particular, the <italic>deep in-memory architecture</italic> (DIMA) has emerged as an attractive alternative to the traditional von Neumann (digital) architecture for realizing energy and latency-efficient machine learning systems in silicon. Multiple DIMA integrated circuit (IC) prototypes have demonstrated energy-delay product (EDP) gains of up to <inline-formula> <tex-math notation="LaTeX">$100\times $ </tex-math></inline-formula> over a digital architecture. These EDP gains were achieved <italic>minimal</italic> or sometimes <italic>no loss</italic> in decision-making accuracy which is surprising given its intrinsic analog mixed-signal nature. This paper establishes models and methods to understand the fundamental energy-delay and accuracy trade-offs underlying DIMA by: 1) presenting silicon-validated energy, delay,...

[19] Energy-efficient and reliable in-memory classifier for machine-learning applications
James Clay, ..., and R. Sridhar. IET Comput. Digit. Tech., 2019. 0 citations.
9% Topic Match

Abstract: Large-scale machine-learning (ML) algorithms require extensive memory interactions. Managing or reducing data movement can significantly increase the speed and efficiency of many ML tasks. Towards this end, the authors devise an energy efficient in-memory computing (IMC) kernel for linear classification and design an initial prototype. The authors achieve a power savings of over 6.4 times than a conventional discrete system while improving reliability by 54.67%. The authors employ a split-data-aware technique to manage process, voltage, and temperature variations and to achieve fair trade-offs between energy efficiency, area requirements, and accuracy. The authors utilise a trimodal architecture with a hierarchical tree...

[20] Scaling Up In-Memory-Computing Classifiers via Boosted Feature Subsets in Banked Architectures
Yinqi Tang, ..., and N. Verma. IEEE Transactions on Circuits and Systems II: Express Briefs, 2019. 10 citations.
9% Topic Match

Abstract: In-memory computing is an emerging approach for overcoming memory-accessing bottlenecks, by eliminating the costs of explicitly moving data from point of storage to point of computation outside the array. However, computation increases the dynamic range of signals, such that performing it via the existing structure of dense memory substantially squeezes the signal-to-noise ratio (SNR). In this brief, we explore how computations can be scaled up, to jointly optimize energy/latency/bandwidth gains with SNR requirements. We employ algorithmic techniques to decompose computations so that they can be mapped to multiple parallel memory banks operating at chosen optimal points. Specifically focusing on in-memory...

[21] Analytical guarantees for reduced precision fixed-point margin hyperplane classifiers

Charbel Sakr. Unknown journal, 2017. 0 citations.

7% Topic Match

Abstract: Margin hyperplane classifiers such as support vector machines are strong predictive models having gained considerable success in various classification tasks. Their conceptual simplicity makes them suitable candidates for the design of embedded machine learning systems. Their accuracy and resource utilization can effectively be traded off each other through precision. We analytically capture this trade-off by means of bounds on the precision requirements of general margin hyperplane classifiers. In addition, we propose a principled precision reduction scheme based on the trade-off between input and weight precisions. Our analysis is supported by simulation results illustrating the gains of our approach in terms...

[22] A Low-Power Analog Integrated Implementation of the Support Vector Machine Algorithm with On-Chip Learning Tested on a Bearing Fault Application

Vassilis Alimisis, ..., and P. Sotiriadis. Sensors (Basel, Switzerland), 2023. 25 citations.

6% Topic Match

Abstract: A novel analog integrated implementation of a hardware-friendly support vector machine algorithm that can be a part of a classification system is presented in this work. The utilized architecture is capable of on-chip learning, making the overall circuit completely autonomous at the cost of power and area efficiency. Nonetheless, using subthreshold region techniques and a low power supply voltage (at only 0.6 V), the overall power consumption is 72 μW. The classifier consists of two main components, the learning and the classification blocks, both of which are based on the mathematical equations of the hardware-friendly algorithm. Based on a real-world...

[23] Design optimization for ADMM-Based SVM Training Processor for Edge Computing

Shuo-An Huang, ..., and Chia-Hsiang Yang. 2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2021. 0 citations.

6% Topic Match

Abstract: This paper presents an optimized support vector machine (SVM) training processor employing the alternative direction method of multipliers (ADMM) optimizer. Low-rank approximation is exploited to reduce the dimension of the kernel matrix by employing the Nyström method. Verified in four datasets, the proposed ADMM-based training processor with rank approximation reduces $32\times$ of matrix dimension with only 2% drop in inference accuracy. Compared to the conventional sequential minimal optimization (SMO) algorithm, the ADMM-based training algorithm is able to achieve a 9.8×10^{7} shorter latency for training 2048 samples. Hardware optimization techniques, including pre-computation and memory sharing, are proposed to reduce the...

[24] CASH-RAM: Enabling In-Memory Computations for Edge Inference Using Charge Accumulation and Sharing in Standard 8T-SRAM Arrays

Amogh Agrawal, ..., and K. Roy. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2020. 15 citations.

5% Topic Match

Abstract: Machine Learning (ML) workloads being memory- and compute-intensive, consume large amounts of power running on conventional computing systems, restricting their implementations to large-scale data centers. Transferring large amounts of data from the edge devices to the data centers is not only energy expensive, but sometimes undesirable in security-critical applications. Thus, there is a need for building domain-specific hardware primitives for energy-efficient ML processing at the edge. One such approach - in-memory computing, eliminates frequent and unnecessary data-transfers between the memory and the compute units, by directly computing the data where it is stored. However, the analog nature of computations introduces...

[25] Computer-aided design of machine learning algorithm: Training fixed-point classifier for on-chip low-power implementation

H. Albalawi, ..., and Xin Li. 2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC), 2014. 1 citations.

5% Topic Match

No summary or abstract available

[26] Improved Design for Hardware Implementation of Graph-Based Large Margin Classifiers for Embedded Edge Computing

J. Arias-Garcia, ..., and A.P. Braga. IEEE Transactions on Neural Networks and Learning Systems, 2022. 2 citations.

4% Topic Match

Abstract: The number of connected embedded edge computing Internet of Things (IoT) devices has been increasing over the years, contributing to the significant growth of available data in different scenarios. Thereby, machine learning algorithms arise to enable task automation and process optimization based on those data. However, due to some learning methods' computational complexity implementing geometric classifiers, it is a challenge to map these on embedded systems or devices with limited resources in size, processing, memory, and power, to accomplish the desired requirements. This hampers the applicability of these methods to complex industrial embedded edge applications. This work evaluates strategies to...

[27] Sub-threshold Computational Circuits for High-order Data-driven Analysis of Physiological Signals

Shuayb Zarar, ..., and N. Verma. Unknown journal, 2011. 1 citations.

4% Topic Match

Abstract: Data driven methods, based on supervised machine learning, provide a powerful framework for creating and applying high order models for physiological signal analysis [5]. This work explores methods to dramatically reduce the energy of the kernel computations involved by exploiting the parallelism that they potentially offer. The throughput requirements for real time signal analysis can thus be met at substantially reduced energy levels through low voltage operation. The framework used is the support vector machine (SVM), which enables efficient construction of high order models by first training on signal features. Detection of targeted states then involves computing signal features and classifying these in real time....

[28] MOUSE: Inference In Non-volatile Memory for Energy Harvesting Applications

Salonik Resch, ..., and Ulya R. Karpuzcu. 2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), 2020. 34 citations.

4% Topic Match

Abstract: There is increasing demand to bring machine learning capabilities to low power devices. By integrating the computational power of machine learning with the deployment capabilities of low power devices, a number of new applications become possible. In some applications, such devices will not even have a battery, and must rely solely on energy harvesting techniques. This puts extreme constraints on the hardware, which must be energy efficient and capable of tolerating interruptions due to power outages. Here, we propose an in-memory machine learning accelerator utilizing non-volatile spintronic memory. The combination of processing-in-memory and non-volatility provides a key advantage in that...

[29] Efficient Implementation of SVM-Based Speech/Music Classifier by Utilizing Temporal Locality

Chungsoo Lim and Joon Hyuk Chang Journal of the Institute of Electronics Engineers of Korea, 2012. 0 citations.

3% Topic Match

Abstract: Support vector machines (SVMs) are well known for their pattern recognition capability, but proper care should be taken to alleviate their inherent implementation cost resulting from high computational intensity and memory requirement, especially in embedded systems where only limited resources are available. Since the memory requirement determined by the dimensionality and the number of support vectors is generally too high for a cache in embedded systems to accomodate, frequent accesses to the main memory occur inevitably whenever the cache is not able to provide requested data to the processor. These frequent accesses to the main memory result in overall performance...
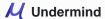
[30] Machine Learning Training on a Real Processing-in-Memory System

Juan Gómez-Luna, ..., and Onur Mutlu. 2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2022. 16 citations.

3% Topic Match

Abstract: Machine learning (ML) algorithms [1]–[6] have become ubiquitous in many fields of science and technology due to their ability to learn from and improve with experience with minimal human intervention. These algorithms train by updating their model parameters in an iterative manner to improve the overall prediction accuracy. However, training machine learning algorithms is a computation-ally intensive process, which requires large amounts of training data. Accessing training data in current processor-centric systems (e.g., CPU, GPU) implies costly data movement between memory and processors, which results in high energy consumption and a large percentage of the total execution cycles. This data...

[31] Perpendicular Bisector Constraint on Artificial Neural Network

Yiqi Chen, ..., and Yufeng Ren. 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), 2017. 0 citations.

3% Topic Match

No summary or abstract available

[32] Deep In-Memory Architectures in SRAM: An Analog Approach to Approximate Computing

Mingu Kang, ..., and Naresh R Shanbhag. Proceedings of the IEEE, 2020. 29 citations.

2% Topic Match

Abstract: This article provides an overview of recently proposed deep in-memory architectures (DIMAs) in SRAM for energy- and latency-efficient hardware realization of machine learning (ML) algorithms. DIMA tackles the data movement problem in von Neumann architectures head-on by deeply embedding mixed-signal computations into a conventional memory array. In doing so, it trades off its computational signal-to-noise ratio (compute SNR) with energy and latency, and therefore, it represents an analog form of approximate computing. DIMA exploits the inherent error immunity of ML algorithms and SNR budgeting methods to operate its analog circuitry in a low-swing/low-compute SNR regime, thereby achieving $> 100\times$...

[33] In-Memory Principal Component Analysis by Crosspoint Array of Resistive Switching Memory: A new hardware approach for energy-efficient data analysis in edge computing

P. Mannocci, ..., and Daniele Ielmin. IEEE Nanotechnology Magazine, 2022. 8 citations.

2% Topic Match

Abstract: In-Memory Computing (IMC) is one of the most promising candidates for data-intensive computing accelerators of machine learning (ML). A key ML algorithm for dimensionality reduction and classification is principal component analysis (PCA), which heavily relies on matrix-vector multiplications (MVM) for which classic von Neumann architectures are not optimized. Here, we provide the experimental demonstration of a new IMC-based PCA algorithm based on power iteration and deflation executed in a 4-kbit array of resistive switching random-access memory (RRAM). The classification accuracy of the Wisconsin Breast Cancer data set reaches 95.43%, close to floating-point implementation. Our simulations indicate a 250× improvement in...

[34] A Robust Minimax Approach to Classification

Gert R. G. Lanckriet, ..., and Michael I. Jordan. J. Mach. Learn. Res., 2002. 517 citations.

2% Topic Match

Abstract: When constructing a classifier, the probability of correct classification of future data points should be maximized. We consider a binary classification problem where the mean and covariance matrix of each class are assumed to be known. No further assumptions are made with respect to the classconditional distributions. Misclassification probabilities are then controlled in a worst-case setting: that is, under all possible choices of class-conditional densities with given mean and covariance matrix, we minimize the worst-case (maximum) probability of misclassification of future data points. For a linear decision boundary, this desideratum is translated in a very direct way into a (convex)...

[35] DIMCA: An Area-Efficient Digital In-Memory Computing Macro Featuring Approximate Arithmetic Hardware in 28 nm

Chuan-Tung Lin, ..., and Mingoo Seok. IEEE Journal of Solid-State Circuits, 2024. 9 citations.

2% Topic Match

Abstract: Recent SRAM-based in-memory computing (IMC) hardware demonstrates high energy efficiency and throughput for matrix–vector multiplication (MVM), the dominant kernel for deep neural networks (DNNs). Earlier IMC macros have employed analog-mixed-signal (AMS) arithmetic hardware. However, those so-called AIMCs suffer from process, voltage, and temperature (PVT) variations. Digital IMC (DIMC) macros, on the other hand, exhibit better robustness against PVT variations, but they tend to require more silicon area. This article proposes novel DIMC hardware featuring approximate arithmetic (DIMCA) to improve area efficiency without hurting compute density (CD). We also propose an approximation-aware training model and a customized number format to compensate...

[36] Minimum precision requirements for the SVM-SGD learning algorithm

Charbel Sakr, ..., and Naresh R Shanbhag. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017. 18 citations.

2% Topic Match

No summary or abstract available

[37] VLSI Extreme Learning Machine: A Design Space Exploration

Enyi Yao and A. Basu. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2016. 34 citations.

2% Topic Match

Abstract: In this paper, we describe a compact low-power high-performance hardware implementation of extreme learning machine for machine learning applications. Mismatches in current mirrors are used to perform the vector-matrix multiplication that forms the first stage of this classifier and is the most computationally intensive. Both regression and classification (on UCI data sets) are demonstrated and a design space tradeoff between speed, power, and accuracy is explored. Our results indicate that for a wide set of problems, $\bar{A}/T$ in the range of 15-25 mV gives optimal results. An input weight matrix rotation method to extend the input dimension and hidden...

[38] EmbML Tool: Supporting the use of Supervised Learning Algorithms in Low-Cost Embedded Systems

Lucas Costa da Silva, ..., and Gustavo E. A. P. A. Batista. Unknown journal, 2019. 0 citations.

1% Topic Match

Abstract: Machine Learning (ML) is becoming a ubiquitous technology employed in many real-world applications. In some applications, sensors measure the environment while ML algorithms are responsible for interpreting the data. These systems often face three main restrictions: power consumption, cost, and lack of infrastructure. Therefore, we need highly-efficient classifiers suitable to execute in unresourceful hardware. However, this scenario conflicts to the state-of-practice of ML, in which classifiers are frequently implemented in high-level interpreted languages, make unrestricted use of floating-point operations and assume plenty of resources. In this paper, we present a software tool named EmbML that implements a pipeline to develop...

[39] A 481pJ/decision 3.4M decision/s Multifunctional Deep In-memory Inference Processor using Standard 6T SRAM Array

Mingu Kang, ..., and Naresh R Shanbhag. ArXiv, 2016. 27 citations.

1% Topic Match

Abstract: This paper describes a multi-functional deep in-memory processor for inference applications. Deep in-memory processing is achieved by embedding pitch-matched low-SNR analog processing into a standard 6T 16KB SRAM array in 65 nm CMOS. Four applications are demonstrated. The prototype achieves up to 5.6X (9.7X estimated for multi-bank scenario) energy savings with negligible (<1%) accuracy degradation in all four applications as compared to the conventional architecture.

[40] A robust minimax approach to classification

Gert R. G. Lanckriet, ..., and Michael I. Jordan. Journal of Machine Learning Research, 2003. 17 citations.

1% Topic Match

Abstract: When constructing a classifier, the probability of correct classification of future data points should be maximized. We consider a binary classification problem where the mean and covariance matrix of each class are assumed to be known. No further assumptions are made with respect to the class-conditional distributions. Misclassification probabilities are then controlled in a worst-case setting: that is, under all possible choices of class-conditional densities with given mean and covariance matrix, we minimize the worst-case (maximum) probability of misclassification of future data points. For a linear decision boundary, this desideratum is translated in a very direct way into a (convex)...

[41] Deep learning acceleration based on in-memory computing

E. Eleftheriou, ..., and A. Sebastian. IBM J. Res. Dev., 2019. 29 citations.

1% Topic Match

Abstract: Performing computations on conventional von Neumann computing systems results in a significant amount of data being moved back and forth between the physically separated memory and processing units. This costs time and energy, and constitutes an inherent performance bottleneck. In-memory computing is a novel non-von Neumann approach, where certain computational tasks are performed in the memory itself. This is enabled by the physical attributes and state dynamics of memory devices, in particular, resistance-based nonvolatile memory technology. Several computational tasks such as logical operations, arithmetic operations, and even certain machine learning tasks can be implemented in such a computational memory unit....

[42] A current-mode linearly classifier based on CMOS

Heng Jin. Journal of Yunnan University, 2014. 0 citations.

1% Topic Match

Abstract: In this paper,a current-mode linearly classifier which can classify linearly non-separable data is proposed.The proposed classifier circuit is composed of trapezoidal activation function circuits and linearly weighted circuits.Linearly weighted circuit is achieved by fully balanced differential transconductor and trapezoidal activation function circuit is mainly comprised of threshold circuits.In order to achieve the classification of the linearly non-separable dataset,the weighting coefficients can be obtained based on Fisher's linear discriminant analysis by MATLAB software.The proposed circuit is simulated and analyzed by PSPICE.The results show that the proposed circuit recognizes the characteristics of simple structures with high accuracy and low power consumption.The proposed...

[43] Effects of Reduced Precision on Floating-Point SVM Classification Accuracy

Prof. Mitsuhisa Sato, ..., and W. Gansterer. Journal Not Provided, Unknown year. 2 citations.

1% Topic Match

Abstract: There is growing interest in performing ever more complex classification tasks on mobile and embedded devices in real-time, which results in the need for efficient implementations of the respective algorithms. Support vector machines (SVMs) represent a powerful class of nonlinear classifiers, and reducing the working precision represents a promising approach to achieving efficient implementations of the SVM classification phase. However, the relationship between SVM classification accuracy and the arithmetic precision used is not yet sufficiently understood. We investigate this relationship in floating-point arithmetic and illustrate that often a large reduction in the working precision...

[44] Computational Complexity of Linear Large Margin Classification With Ramp Loss

Søren Frejstrup Maibing and C. Igel. Unknown journal, 2015. 5 citations.

1% Topic Match

Abstract: Minimizing the binary classification error with a linear model leads to an NP-hard problem. In practice, surrogate loss functions are used, in particular loss functions leading to large margin classification such as the hinge loss and the ramp loss. The intuitive large margin concept is theoretically supported by generalization bounds linking the expected classification error to the empirical margin error and the complexity of the considered hypotheses class. This article addresses the fundamental question about the computational complexity of determining whether there is a hypotheses class with a hypothesis such that the upper bound on the generalization error is below...

[45] Online Learning of Approximate Maximum Margin Classifiers with Biases

Kosuke Ishibashi, ..., and Masayuki Takeda. Journal Not Provided, 2007. 0 citations.

1% Topic Match

Abstract: We consider online learning of linear classifiers which approximately maximize the 2-norm margin. Given a linearly separable sequence of instances, typical online learning algorithms such as Perceptron and its variants, map them into an augmented space with an extra dimension, so that those instances are separated by a linear classifier without a constant bias term. However, this mapping might decrease the margin over the instances. In this paper, we propose a modified version of Li and Long's ROMMA that avoids such the mapping and we show that our modified algorithm achieves higher margin than previous online learning algorithms.

[46] A 19.4 nJ/decision 364K decisions/s in-memory random forest classifier in 6T SRAM array

Mingu Kang, ..., and Naresh R Shanbhag. ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference, 2017. 29 citations.

1% Topic Match

No summary or abstract available

[47] Special Session - Non-Volatile Memories: Challenges and Opportunities for Embedded System Architectures with Focus on Machine Learning Applications

Jörg Henkel, ..., and Hsiang-Yun Cheng. 2023 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems (CASES), 2023. 1 citations.

1% Topic Match

Abstract: This paper explores the challenges and opportunities of integrating non-volatile memories (NVMs) into embedded systems for machine learning. NVMs offer advantages such as increased memory density, lower power consumption, non-volatility, and compute-in-memory capabilities. The paper focuses on integrating NVMs into embedded systems, particularly in intermittent computing, where systems operate during periods of available energy. NVM technologies bring persistence closer to the CPU core, enabling efficient designs for energy-constrained scenarios. Next, computation in resistive NVMs is explored, highlighting its potential for accelerating machine learning algorithms. However, challenges related to reliability and device non-idealities need to be addressed. The paper also discusses...

[48] A PAC-Bayesian margin bound for linear classifiers

R. Herbrich and T. Graepel. IEEE Trans. Inf. Theory, 2002. 6 citations.

0% Topic Match

Abstract: We present a bound on the generalization error of linear classifiers in terms of a refined margin quantity on the training sample. The result is obtained in a probably approximately correct (PAC)-Bayesian framework and is based on geometrical arguments in the space of linear classifiers. The new bound constitutes an exponential improvement of the so far tightest margin bound, which was developed in the luckiness framework, and scales logarithmically in the inverse margin. Even in the case of less training examples than input dimensions sufficiently large margins lead to nontrivial bound values and-for maximum margins-to a vanishing complexity term. In...

[49] Evaluating Machine LearningWorkloads on Memory-Centric Computing Systems

Juan Gómez-Luna, ..., and O. Mutlu. 2023 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2023. 33 citations.

0% Topic Match

Abstract: Training machine learning (ML) algorithms is a computationally intensive process, which is frequently memory-bound due to repeatedly accessing large training datasets. As a result, processor-centric systems (CPU, GPU) waste large amounts of energy and execution cycles due to the data movement between memory units and processing units. Memory-centric computing systems, i.e., systems with processing-in-memory (PIM) capabilities, can alleviate this data movement bottleneck. Our goal is to understand the potential of general-purpose PIM architectures to accelerate ML training. To do so, we (1) implement several classic ML algorithms (namely, linear regression, logistic regression, decision tree, K-Means clustering) on a real-world generalpurpose...

[50] SRAM-Based Computing-in-Memory Macro With Fully Parallel One-Step Multibit Computation

E. Choi, ..., and Minkyu Je. IEEE Solid-State Circuits Letters, 2022. 16 citations.

0% Topic Match

Abstract: In this letter, we present a multibit static random-access memory computing-in-memory (CIM) macro with enhanced energy efficiency for edge devices tasking machine learning (ML) deep neural networks (DNNs). The proposed CIM macro computes matrix-vector multiplications (MVM) in an efficient "one-step" method reducing the energy consumption and control complexity. Furthermore, the proposed method computes not only the multiplications of a single weight but also the multibit weight with bit-shifting in the charge domain without the use of additional CMOS switches, thereby achieving very high energy efficiency. Measurement results in a 65-nm CMOS prototype chip show that it achieves the highest throughput...

[51] Classification using margin pursuit

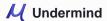Matthew J. Holland. Unknown journal, 2018. 6 citations.

0% Topic Match

Abstract: In this work, we study a new approach to optimizing the margin distribution realized by binary classifiers, in which the learner searches the hypothesis space in such a way that a pre-set margin level ends up being a distribution-robust estimator of the margin location. This procedure is easily implemented using gradient descent, and admits finite-sample bounds on the excess risk under unbounded inputs, yielding competitive rates under mild assumptions. Empirical tests on real-world benchmark data reinforce the basic principles highlighted by the theory.

[52] A 19.4-nJ/Decision, 364-K Decisions/s, In-Memory Random Forest Multi-Class Inference Accelerator

Mingu Kang, ..., and Naresh R Shanbhag. IEEE Journal of Solid-State Circuits, 2018. 48 citations.

0% Topic Match

Abstract: This paper presents an integrated circuit (IC) realization of a random forest (RF) machine learning classifier in a 65-nm CMOS. Algorithm, architecture, and circuits are co-optimized to achieve aggressive energy and delay benefits by taking advantage of the inherent error resiliency derived from the ensemble nature

of an RF classifier. Deterministic sub-sampling (DSS) and regularized decision trees reduce interconnect complexity, and avoid irregular memory access patterns and computations, thereby reducing the energy-delay product (EDP). The prototype IC also employs low-swing analog in-memory computations embedded in a standard 6T SRAM to enable massively parallel tree node comparisons, thereby minimizing the memory...

[53] Optimal multivariate classification by linear thresholding
Baro Hyun, ..., and A. Girard. 2012 American Control Conference (ACC), 2012. 1 citations.
0% Topic Match

No summary or abstract available

[54] Polyhedral Conic Classifiers for Computer Vision Applications and Open Set Recognition
Hakan Cevikalp and Halil Saglamlar. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021. 15 citations.
0% Topic Match

Abstract: This paper introduces a family of quasi-linear discriminants that outperform current large-margin methods in sliding window visual object detection and open set recognition tasks. In these applications, the classification problems are both numerically imbalanced – positive (object class) training and test windows are much rarer than negative (non-class) ones – and geometrically asymmetric – the positive samples typically form compact, visually-coherent groups while negatives are much more diverse, including anything at all that is not a well-centered sample from the target class. For such tasks, there is a need for discriminants whose decision regions focus on tightly circumscribing the positive...

[55] An In-Memory Computing SRAM Macro for Memory-Augmented Neural Network
Sunghoon Kim, ..., and Dongsuk Jeon. IEEE Transactions on Circuits and Systems II: Express Briefs, 2022. 10 citations.
0% Topic Match

Abstract: In-Memory Computing (IMC) has been widely studied to mitigate data transfer bottlenecks in von Neumann architectures. Recently proposed IMC circuit topologies dramatically reduce data transfer requirements by performing various operations such as Multiply-Accumulate (MAC) inside the memory. In this brief, we present an SRAM macro designed for accelerating Memory-Augmented Neural Network (MANN). We first propose algorithmic optimizations for a few-shot learning algorithm employing MANN for efficient hardware implementation. Then, we present an SRAM macro that efficiently accelerates the algorithm by realizing key operations such as L1 distance calculation and Winner-Take-All (WTA) operation through mixed-signal computation circuits. Fabricated in 40nm LP...

[56] Rounding Methods for Discrete Linear Classification (Extended Version)
Y. Chevaleyre, ..., and Jean-Daniel Zucker. Unknown journal, 2013. 1 citations.
0% Topic Match

Abstract: Learning discrete linear classifiers is known as a difficult challenge. In this paper, this learning task is cast as combinatorial optimization problem: given a training sample formed by positive and negative feature vectors in the Euclidean space, the goal is to find a discrete linear function that minimizes the cumulative hinge loss of the sample. Since this problem is NP-hard, we examine two simple rounding algorithms that discretize the fractional solution of the problem. Generalization bounds are derived for several classes of binary-weighted linear functions, by analyzing the Rademacher complexity of these classes and by establishing approximation bounds for our...

[57] Rounding Methods for Discrete Linear Classification
Y. Chevaleyre, ..., and Jean-Daniel Zucker. Unknown journal, 2013. 25 citations.
0% Topic Match

Abstract: Learning discrete linear classifiers is known as a difficult challenge. In this paper, this learning task is cast as combinatorial optimization problem: given a training sample formed by positive and negative feature vectors in the Euclidean space, the goal is to find a discrete linear function that minimizes the cumulative hinge loss of the sample. Since this problem is NP-hard, we examine two simple rounding algorithms that discretize the fractional solution of the problem. Generalization bounds are derived for several classes of binary-weighted linear functions, by analyzing the Rademacher complexity of these classes and by establishing approximation bounds for our...

[58] Robust Classification with Interval Data
L. Ghaoui and Gert R. G. Lanckriet. Journal Not Provided, 2003. 74 citations.
0% Topic Match

Abstract: We consider a binary, linear classification problem in which the data points are assumed to be unknown, but bounded within given hyper-rectangles, i.e., the covariates are bounded within intervals explicitly given for each data point separately. We address the problem of designing a robust classifier in this setting by minimizing the worst-case value of a given loss function, over all possible choices of the data in these multi-dimensional intervals. We examine in detail the application of this methodology to three specific loss functions, arising in support vector machines, in logistic regression and in minimax probability machines. We show that in...

[59] An Interpretable SVM Classifier by Discrete-Weight Optimization
Fethi Jarray and Sabri Boughorbel. Journal Not Provided, 2017. 2 citations.
0% Topic Match

Abstract: The main problem investigated in this paper is to learn "interpretable" linear classifiers from data. Interpretable models are captured using "discrete" linear functions. The learning problem is formulated as minimizing the cumulative zero-one loss of a discrete hyperplane, penalized by the standard L2 regularizer. This learning task is cast as a MILP problem, and solved using convex relaxation and rounding. Experiments on both synthetic and real-world datasets corroborate the interest of this approach. We benchmarked the proposed method against two classifiers: iDILSVM a discrete version of SVM based a hinge-loss and iithe traditional linear L1-norm SVM. Our algorithm outperforms DILSVM...

[60] Classification on proximity data with LP-machines
T. Graepel, ..., and R. C. Williamson. Unknown journal, 1999. 111 citations.
0% Topic Match

Abstract: We provide a new linear program to deal with classification of data in the case of data given in terms of pairwise proximities. This allows to avoid the problems inherent in using feature spaces with indefinite metric in support vector machines, since the notion of a margin is purely needed in input space where the classification actually occurs. Moreover in our approach we can enforce sparsity in the proximity representation by sacrificing training error. This turns out to be favorable for proximity data. Similar to /spl nu/-SV methods, the only parameter needed in the algorithm is the (asymptotical) number of...

[61] Deep In-Memory Architectures for Machine Learning-Accuracy
Member Ieee Mingu Kang, ..., and F. I. Naresh R. Shanbhag. Journal Not Provided, Unknown year. 0 citations.
0% Topic Match

Abstract: —In-memory architectures, in particular, the deep in-memory architecture (DIMA) has emerged as an attractive alternative to the traditional von Neumann (digital) architecture for realizing energy and latency-ef cient machine learning systems in silicon. Multiple DIMA integrated circuit (IC) prototypes have demonstrated energy-delay product (EDP) gains of up to 100 × over a digital architecture. These EDP gains were achieved minimal or sometimes no loss in decision-making accuracy which is surprising given its intrinsic analog mixed-signal nature. This paper establishes models and methods to understand the fundamental energy-delay and accuracy trade-offs underlying DIMA by: 1) presenting silicon-validated energy, delay, and accuracy models;...

[62] An RRAM-based implementation of a template matching circuit for low-power analogue classification
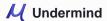Patrick Foster, ..., and Spyros Stathopoulos Themis Prodromakis. Frontiers in Electronics, 2023. 0 citations.
0% Topic Match

Abstract: Recent advances in machine learning and neuro-inspired systems enabled the increased interest in efficient pattern recognition at the edge. A wide variety of applications, such as near-sensor classification, require fast and low-power approaches for pattern matching through the use of associative memories and their more well-known implementation, Content Addressable Memories (CAMs). Towards addressing the need for low-power classification, this work showcases an RRAM-based analogue CAM (ACAM) intended for template matching applications, providing a low-power reconfigurable classification engine for the extreme edge. The circuit uses a low component count at 6T2R2M, comparable with the most compact existing cells of this type....

[63] An In-Memory VLSI Architecture for Convolutional Neural Networks
Mingu Kang, ..., and Naresh R Shanbhag. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2018. 48 citations.
0% Topic Match

Abstract: This paper presents an energy-efficient and high throughput architecture for convolutional neural networks (CNN). Architectural and circuit techniques are proposed to address the dominant energy and delay costs associated with data movement in CNNs. The proposed architecture employs a deep in-memory architecture, to embed energy-efficient low swing mixed-signal computations in the periphery of the SRAM bitcell array. An efficient data access pattern and a mixed-signal multiplier are proposed to exploit data reuse opportunities in convolution. Silicon-validated energy, delay, and behavioral models of the proposed architecture are developed and employed to perform large-scale system simulations. System-level simulations using these models show...

[64] Linear and Kernel Classification: When to Use Which?

Hsin-Yuan Huang and Chih-Jen Lin. Unknown journal, 2016. 43 citations.

0% Topic Match

Abstract: Kernel methods are known to be a state-of-the-art classification technique. Nevertheless, the training and prediction cost is expensive for large data. On the other hand, linear classifiers can easily scale up, but are inferior to kernel classifiers in terms of predictability. Recent research has shown that for some data sets (e.g., document data), linear is as good as kernel classifiers. In such cases, the training of a kernel classifier is a waste of both time and memory. In this work, we investigate the important issue of efficiently and automatically deciding whether kernel classifiers perform strictly better than linear for a...

[65] Tightened L 0 -relaxation Penalties for Classification

Noam Goldberg and Jonathan Eckstein. Journal Not Provided, 2009. 1 citations.

0% Topic Match

Abstract: We thank Rob Schapire for helpful discussions. The first author would also like to thank Kristin Bennett for her comments during earlier stages of this work, and Martin Milanic for his comments regarding a related feature selection problem. ABSTRACT In optimization-based classification model selection, for example when using linear programming formulations, a standard approach is to penalize the L 1 norm of some linear functional in order to select sparse models. Instead, we propose a novel integer linear program for sparse classifier selection, generalizing the minimum disagreement hyperplane problem whose complexity has been investigated in computational learning theory. Specifically, our...

[66] Design of digital classifier circuits with nearest neighbour prior sample selection

W. S. Lacerda and A. Braga. Fifth International Conference on Hybrid Intelligent Systems (HIS'05), 2005. 0 citations.

0% Topic Match

No summary or abstract available

[67] 33.2 A Fully Integrated Analog ReRAM Based 78.4TOPS/W Compute-In-Memory Chip with Fully Parallel MAC Computing

Qi Liu, ..., and Huaqiang Wu. 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020. 207 citations.

0% Topic Match

Abstract: Non-volatile memory (NVM) based computing-in-memory (CIM) shows significant advantages in handling deep learning tasks for artificial intelligence (AI) applications. To overcome the decreasing cost effectiveness of transistor scaling and the intrinsic inefficiency of data-shuttling in the von-Neumann architecture, CIM is proposed to realize high-speed and low-power system with parallel multiplication accumulation (MAC) computing [1] [2]. However, current demonstrations are mainly based on single macro and present limited computing parallelism. Realizing a fully-integrated CIM chip with a complete neural network model is still missing. The major challenges lie in: (1) The IR drop and transient errors when carrying out MAC operations...

[68] An In-Memory-Computing Charge-Domain Ternary CNN Classifier

Xiangxing Yang, ..., and Nan Sun. IEEE Journal of Solid-State Circuits, 2023. 4 citations.

0% Topic Match

Abstract: The article presents a charge-domain computing ternary neural network (TNN) classifier with a complete four-layer neural network (NN) on a chip. The proposed ternary network provides 1.5-b resolution (0/+1/ 1) for weights and activations, leading to 3.9× fewer operations (OPs) per inference than binary neural network (BNN) for the same Modified National Institute of Standards and Technology (MNIST) accuracy. The 1.5-b multiply-and-accumulate (MAC) is implemented by <inline-formula> <tex-math notation="LaTeX">$V_{\text {CM}}$ </tex-math></inline-formula>-based capacitor switching scheme, which inherently benefits from the reduced signal swing on the capacitive digital-to-analog converter (CDAC). Also, the <inline-formula> <tex-math notation="LaTeX">$V_{\text {CM}}$ </tex-math></inline-formula>-based MAC introduces sparsity during training, resulting...

[69] Introduction to Binary Support Vector Machines

C-Myb Snf, ..., and Snf2 C-Myb. Journal Not Provided, 2004. 0 citations.

0% Topic Match

Abstract: , linear SVMs solve an optimization problem that seeks a maximum margin classifier (i.e. a hyperplane with the maximum margin width that separates training instances of two classes). This classifier is defined by a subset of training data points called support vectors. Then unseen data instances (i.e. samples, which were not used for training) are classified based on which side of the hyperplane they fall into. Mathematically, the separation hyperplane is defined by the equation where and b come from the solution of the optimization problem. The decision function is then defined by . m n x x R ...

[70] Incremental-Precision Based Feature Computation and Multi-Level Classification for Low-Energy Internet-of-Things

Sandhya Koteshwara and K. Parhi. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2018. 14 citations.

0% Topic Match

Abstract: This paper presents a novel technique to reduce energy consumption of a machine learning classifier based on incremental-precision feature computation and classification. Specifically, the algorithm starts with features computed using the lowest possible precision. Depending on the classification accuracy, the features of the previous level are combined with features of the incremental-precision to compute the features in higher-precision. This process is continued till a desired accuracy is obtained. A certain threshold that allows many samples to be classified using a low-precision classifier can reduce energy consumption, but increases misclassification error. To implement hardware which provides the required updates in precision,...

[71] SEFR: A Fast Linear-Time Classifier for Ultra-Low Power Devices

Hamidreza Keshavarz, ..., and Reza Rawassizadeh. ArXiv, 2020. 15 citations.

0% Topic Match

Abstract: One of the fundamental challenges for running machine learning algorithms on battery-powered devices is the time and energy needed for computation, as these devices have constraints on resources. There are energy-efficient classifier algorithms, but their accuracy is often sacrificed for resource efficiency. Here, we propose an ultra-low power binary classifier, SEFR, with linear time complexity, both in the training and the testing phases. The SEFR method runs by creating a hyperplane to separate two classes. The weights of this hyperplane are calculated using normalization, and then the bias is computed based on the weights. SEFR is comparable to state-of-the-art classifiers...

[72] Practical Near-Data Processing for In-Memory Analytics Frameworks

Mingyu Gao, ..., and Christos Kozyrakis. 2015 International Conference on Parallel Architecture and Compilation (PACT), 2015. 262 citations.

0% Topic Match

No summary or abstract available

[73] An Energy-Efficient and High Throughput in-Memory Computing Bit-Cell With Excellent Robustness Under Process Variations for Binary Neural Network

Gobinda Saha, ..., and M. Ahosan Ul Karim. IEEE Access, 2020. 7 citations.

0% Topic Match

Abstract: In-memory computing (IMC) is a promising approach for energy cost reduction due to data movement between memory and processor for running data-intensive deep learning applications on the computing systems. Together with Binary Neural Network (BNN), IMC provides a viable solution for running deep neural networks at the edge devices with stringent memory and energy constraints. In this paper, we propose a novel 10T bit-cell with a back-end-of-line (BEOL) metal-oxide-metal (MOM) capacitor laid on pitch for in-memory computing. Our IMC bit-cell, when arranged in a memory array, performs binary convolution (XNOR followed by Bit-count operations) and binary activation generation operations. We...

[74] Effect of OTS Selector Reliabilities on NVM Crossbar-based Neuromorphic Training

Wen Ma, ..., and Martin Lueker-Boden. 2022 IEEE International Reliability Physics Symposium (IRPS), 2022. 1 citations.

0% Topic Match

Abstract: This paper studies the use of 1-selector-1-resistor (1S1R) devices as the cross-point components in a crossbar array for deep neural network (DNN) training. Two training algorithms—minibatch gradient descent (MBGD) and a low-rank approximation to MBGD—are evaluated with compact models of ovonic threshold switching (OTS) selectors considering several reliability issues, and resistive random-access memory (ReRAM), a type of emerging non-volatile memory (NVM) device. We explore various reliability concerns of the OTS selectors including the low on-state conductance, threshold voltage variation, and internal voltage offset, etc. and propose corresponding techniques to overcome the reliability challenges.

[75] Scalable-effort classifiers for energy-efficient machine learning

Swagath Venkataramani, ..., and M. Shoaib. Proceedings of the 52nd Annual Design Automation Conference, 2015. 81 citations.

0% Topic Match

No summary or abstract available

[76] A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection

Preeti Mishra, ..., and Emmanuel S. PIlli. IEEE Communications Surveys & Tutorials, 2019. 519 citations.

0% Topic Match

Abstract: Intrusion detection is one of the important security problems in todays cyber world. A significant number of techniques have been developed which are based on machine learning approaches. However, they are not very successful in identifying all types of intrusions. In this paper, a detailed investigation and analysis of various machine learning techniques have been carried out for finding the cause of problems associated with various machine learning techniques in detecting intrusive activities. Attack classification and mapping of the attack features is provided corresponding to each attack. Issues which are related to detecting low-frequency attacks using network attack dataset are...

[77] A Linearly Convergent Linear-Time First-Order Algorithm for Support Vector Classification with a Core Set Result

Piyush Kumar and E. A. Yildirim. INFORMS J. Comput., 2011. 14 citations.

0% Topic Match

Abstract: We present a simple first-order approximation algorithm for the support vector classification problem. Given a pair of linearly separable data sets and e (0,1), the proposed algorithm computes a separating hyperplane whose margin is within a factor of (1-e) of that of the maximum-margin separating hyperplane. We discuss how our algorithm can be extended to nonlinearly separable and inseparable data sets. The running time of our algorithm is linear in the number of data points and in 1/e. In particular, the number of support vectors computed by the algorithm is bounded above by O(¶) for all sufficiently small e...

[78] Application Driven Memory - Circuits and Architecture

Shixiong Jiang. Unknown journal, 2018. 0 citations.

0% Topic Match

Abstract: .... xviii Chapter

[79] Ee 127/227at: Optimization Models in Engineering Support Vector Machines 1 Support Vector Machines Revisited 1.1 (strictly) Separable Case: the Linear Hard-margin Classifier

Arturo Fernandez. Journal Not Provided, 2014. 0 citations.

0% Topic Match

Abstract: Consider a binary classification prediction problem. Training data are given and we denote them observations as T = {X,y} = {xi, yi}i=1, where xi R, yi {+1, 1} (accordingly X Rmxn). First, we will consider the case where the two classes are linearly separable. That is, by an n-dimensional decision boundary which is the result of an n + 1-dimensional hyperplane). Furthermore, since there can exist multiple hyperplanes that split the classes, we would like to find the one with the maximal margin. The motivation for this being that the decision boundary will be prone to variations in...

[80] Random Projections for Linear Support Vector Machines

Saurabh Paul, ..., and P. Drineas. ACM Trans. Knowl. Discov. Data, 2012. 51 citations.

0% Topic Match

Abstract: Let X be a data matrix of rank Áwhose rows represent n points in d-dimensional space. The linear support vector machine constructs a hyperplane separator that maximizes the 1-norm soft margin. We develop a new oblivious dimension reduction technique that is precomputed and can be applied to any input matrix X. We prove that, with high probability, the margin and minimum enclosing ball in the feature space are preserved to within μrelative error, ensuring comparable generalization as in the original space in the case of classification. For regression, we show that the margin is preserved to within μrelative error with...

[81] A Hybrid Edge Classifier: Combining TinyML-Optimised CNN with RRAM-CMOS ACAM for Energy-Efficient Inference

Kieran Woodward, ..., and T. Prodromakis. ArXiv, 2025. 0 citations.

0% Topic Match

Abstract: In recent years, the development of smart edge computing systems to process information locally is on the rise. Many near-sensor machine learning (ML) approaches have been implemented to introduce accurate and energy efficient template matching operations in resource-constrained edge sensing systems, such as wearables. To introduce novel solutions that can be viable for extreme edge cases, hybrid solutions combining conventional and emerging technologies have started to be proposed. Deep Neural Networks (DNN) optimised for edge application alongside new approaches of computing (both device and architecture -wise) could be a strong candidate in implementing edge ML solutions that aim at competitive...

[82] An On-Chip Trainable and Scalable In-Memory ANN Architecture for AI/ML Applications

Abhash Kumar, ..., and B. Gupta. Circuits, Systems, and Signal Processing, 2022. 2 citations.

0% Topic Match

No summary or abstract available

[83] A Soft Margin Method for Multiconlitron Design

Leng Qiang. Pattern Recognition and Artificial Intelligence, 2013. 0 citations.

0% Topic Match

Abstract: Multiconlitron is a general framework for constructing piecewise linear classifiers. For the convexly separable and the commonly separable datasets,it can correctly separate them by using support conlitron algorithm(SCA) and multiconlition algorithm(SMA),respectively. On this basis,a soft margin method for multiconlitron design is proposed. Firstly,the training samples are mapped from the input space to a high dimensional feature space,and one class of those samples is clustered into some groups by K-means algorithm. Then,the conlitron is constructed between each group and another class of samples,and the integrated model,multiconlitron,is obtained. The proposed method can overcome the inapplicability of the original model to...

[84] Achieving Low Power Classification with Classifier Ensemble

Fanglei Hu, ..., and Hailong Jiao. 2018 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2018. 1 citations.

0% Topic Match

Abstract: Machine learning algorithms, such as Support Vector Machine (SVM) and Artificial Neural Networks, have been widely applied in many aspects of daily life. Low power/energy integrated circuit implementation of machine learning algorithms with high accuracy is however still a great challenge, which is critical for portable and wearable devices. In this paper, classifier ensemble is investigated for achieving low power classification. The classifier ensemble algorithm Adaboost is employed to combine multiple SVM classifiers with linear kernel to achieve high classification accuracy while reducing the hardware complexity. The proposed classifier ensemble is evaluated on the MNIST dataset by using a 45-nm...

[85] Perceptron-like large margin classifiers

Petroula Tsampouka. Unknown journal, 2005. 4 citations.

0% Topic Match

Abstract: We consider perceptron-like algorithms with margin in which the standard classification condition is modified to require a specific value of the margin in the augmented space. The new algorithms are shown to converge in a finite number of steps and used to approximately locate the optimal weight vector in the augmented space following a procedure analogous to Bolzano's bisection method. We demonstrate that as the data are embedded in the augmented space at a larger distance from the origin the maximum margin in that space approaches the maximum geometric one in the original space. Thus, our algorithmic procedure could be...

[86] Linear Classification of Badly Conditioned Data

Helene Dörksen and V. Lohweg. 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 2018. 0 citations.

0% Topic Match

Abstract: We present a method for the fast and robust linear classification of badly conditioned data. In our considerations, badly conditioned data are such data which are numerically difficult to handle. Due to, e.g. a large number of features or a large number of objects representing classes as well as noise, outliers or incompleteness, the common software computation of the discriminating linear combination of features between classes fails or is extremely time consuming. The theoretical foundations of our approach are based on the single feature ranking, which allows fast calculation of the approximative initial classification boundary. For the increasing of classification...

[87] Negative Capacitance FET 8T SRAM Computing in-Memory based Logic Design for Energy Efficient AI Edge Devices

Venu Birudu, ..., and Ramesh Vaddi. 2024 IEEE International Symposium on Circuits and Systems (ISCAS), 2024. 2 citations.

0% Topic Match

Abstract: Recent hardware developments in artificial intelligence (AI) edge devices expect architectures to support multiply and accumulation operations while preserving high inference accuracy and energy efficiency. This work proposes a compute in-memory (CiM) cell design with steep slope Negative capacitance field effect transistors (NCFET) for energy efficient computing architectures. The NCFET based 8T SRAM cell has been designed and analyzed for performance metrics such as noise margins and energy consumption during read/write modes for an optimum Ferroelectric layer thickness (Tfe) at VDD=0.3 V and 0.5V. Further, the NCFET 8T SRAM cell has been modified to realize energy efficient operations such as...

[88] Decision Manifolds: Classification Inspired by Self-Organization

Georg Pölzlbauer, ..., and A. Rauber. Unknown journal, 2007. 0 citations.

0% Topic Match

Abstract: We present a classifier algorithm that approximates the decision surface of labeled data by a patchwork of separating hyperplanes. The hyperplanes are arranged in a way inspired by how Self-Organizing Maps are trained. We take advantage of the fact that the boundaries can often be approximated by linear ones connected by a low-dimensional nonlinear manifold. The resulting classifier allows for a voting scheme that averages over neighboring hyperplanes. Our algorithm is computationally efficient both in terms of training and classification. Further, we present a model selection framework for estimation of the paratmeters of the classification boundary, and show results for...

[89] An energy-efficient VLSI architecture for pattern recognition via deep embedding of computation in SRAM

Mingu Kong, ..., and Ken Curewitz. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014. 161 citations.

0% Topic Match

No summary or abstract available

[90] EmbML Tool: Supporting the use of Supervised Learning Algorithms in Low-Cost Embedded Systems

Lucas Tsutsui da Silva, ..., and Gustavo E. A. P. A. Batista. 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019. 15 citations.

0% Topic Match

Abstract: Machine Learning (ML) is becoming a ubiquitous technology employed in many real-world applications. In some applications, sensors measure the environment while ML algorithms are responsible for interpreting the data. These systems often face three main restrictions: power consumption, cost, and lack of infrastructure. Therefore, we need highly-efficient classifiers suitable to execute in unresourceful hardware. However, this scenario conflicts to the state-of-practice of ML, in which classifiers are frequently implemented in high-level interpreted languages, make unrestricted use of floating-point operations and assume plenty of resources. In this paper, we present a software tool named EmbML that implements a pipeline to develop...

[91] Learning sparse low-threshold linear classifiers

Sivan Sabato, ..., and Tong Zhang. J. Mach. Learn. Res., 2012. 3 citations.

0% Topic Match

Abstract: We consider the problem of learning a non-negative linear classifier with a l1-norm of at most k, and a fixed threshold, under the hinge-loss. This problem generalizes the problem of learning a k-monotone disjunction. We prove that we can learn efficiently in this setting, at a rate which is linear in both k and the size of the threshold, and that this is the best possible rate. We provide an efficient online learning algorithm that achieves the optimal rate, and show that in the batch case, empirical risk minimization achieves this rate as well. The rates we show are tighter...

[92] Maximal Cutting Construction for Multiconlitron

Leng Qiang. Acta Automatica Sinica, 2014. 1 citations.

0% Topic Match

Abstract: Multiconlitron is a general framework for constructing piecewise linear classifiers. For convexly separable and commonly separable data sets, it can separate them correctly by using support conlitron algorithm(SCA) and support multiconlitron algorithm(SMA), respectively. On this basis, the paper proposes a maximal cutting construction method for multiconlitron design. The method consists of two training processes. In the first step, the maximal cutting process(MCP) is utilized iteratively to find a linear boundary such that it can obtain the maximum number of samples. Thus,the MCP can reduce the number of linear boundaries and construct a minimal set of decision functions, and ultimately simplify...

[93] The Perceptron Algorithm with Uneven Margins

Yaoyong Li, ..., and J. Kandola. Unknown journal, 2002. 173 citations.

0% Topic Match

Abstract: The perceptron algorithm with margins is a simple, fast and e ective learning algorithm for linear classi ers; it produces decision hyperplanes within some constant ratio of the maximal margin. In this paper we study this algorithm and a new variant: the perceptron algorithm with uneven margins, tailored for document categorisation problems (i.e. problems where classes are highly unbalanced and performance depends on the ranking of patterns). We discuss the interest of these algorithms from a theoretical point of view, provide a generalisation of Noviko 's theorem for uneven margins, give a geometrically description of these algorithms and show experimentally that both algorithms...

[94] Computing in-memory reconfigurable (accurate/approximate) adder design with negative capacitance FET 6T-SRAM for energy efficient AI edge devices

Birudu Venu, ..., and Ramesh Vaddi. Semiconductor Science and Technology, 2024. 0 citations.

0% Topic Match

Abstract: Computing in-memory (CiM) is an alternative to von-Neumann architectures for energy efficient AI edge computing architectures with CMOS scaling. Approximate computing in-memory (ACiM) techniques have also been recently proposed to further increase the energy efficiency of such architectures. In the first part of the work, a negative capacitance FET (NCFET) based 6T-SRAM CiM accurate full adder has been proposed, designed and performance benchmarked with equivalent baseline 40 nm CMOS design. Due to the steep slope characteristics of NCFET, at an increased ferroelectric layer thickness, T fe of 3 nm, the energy consumption of the proposed accurate NCFET based CiM design...

[95] Maximum Margin Logistic Vector Machine

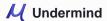Hu Wen. Journal of Software, 2012. 0 citations.

0% Topic Match

Abstract: The L2-kernel classifier does not consider explicitly its classification margin when approximating the difference of densities(DoD) with the integrated squared error(ISE) criterion of probability densities,which is disadvantageous for improving the performance of classifiers to a certain extent.Its weights can simply be obtained by solving the corresponding QP problem which results in the comparatively slow training speed and is impractical especially for large datasets.With the aim of overcoming the above drawbacks,a new classification method is proposed in this paper,called the maximum margin logistic vector machine(MMLVM),which maximizes the DoDbased classification margin and finds the corresponding weight vector by solving a logistic optimization...

[96] Linear dimensionality reduction: survey, insights, and generalizations

J. Cunningham and Zoubin Ghahramani. J. Mach. Learn. Res., 2014. 595 citations.

0% Topic Match

Abstract: Linear dimensionality reduction methods are a cornerstone of analyzing high dimensional data, due to their simple geometric interpretations and typically attractive computational properties. These methods capture many data features of interest, such as covariance, dynamical structure, correlation between data sets, input-output relationships, and margin between data classes. Methods have been developed with a variety of names and motivations in many fields, and perhaps as a result the connections between all these methods have not been highlighted. Here we survey methods from this disparate literature as optimization programs over matrix manifolds. We discuss principal component analysis, factor analysis, linear multidimensional scaling,...

[97] Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM)

Darmatasia and M. I. Fanany. 2017 5th International Conference on Information and Communication Technology (ICoIC7), 2017. 61 citations.

0% Topic Match

No summary or abstract available

[98] MULTI CLASS CLASSIFICATION

Surendra Gupta. Journal Not Provided, 2012. 5 citations.

0% Topic Match

Abstract: Generalization error of classifier can be reduced by larger margin of separating hyperplane. The proposed classification algorithm implements margin in classical perceptron algorithm, to reduce generalized errors by maximizing margin of separating hyperplane. Algorithm uses the same updation rule with the perceptron, to converge in a finite number of updates to solutions, possessing any desirable fraction of the margin. This solution is again optimized to get maximum possible margin. The algorithm can process linear, non-linear and multi class problems. Experimental results place the proposed classifier equivalent to the support vector machine and even better in some cases. Some preliminary experimental...

[99] Multiconlitron: A General Piecewise Linear Classifier

Yujian Li, ..., and Houjun Li. IEEE Transactions on Neural Networks, 2011. 45 citations.

0% Topic Match

No summary or abstract available

[100] Injecting Logical Constraints into Neural Networks via Straight-Through Estimators

Zhun Yang, ..., and Chi-youn Park. ArXiv, 2023. 20 citations.

0% Topic Match

Abstract: Injecting discrete logical constraints into neural network learning is one of the main challenges in neuro-symbolic AI. We find that a straight-through-estimator, a method introduced to train binary neural networks, could effectively be applied to incorporate logical constraints into neural network learning. More specifically, we design a systematic way to represent discrete logical constraints as a loss function; minimizing this loss using gradient descent via a straight-through-estimator updates the neural network's weights in the direction that the binarized outputs satisfy the logical constraints. The experimental results show that by leveraging GPUs and batch training, this method scales significantly better than...

[101] Implementation of an On-Chip Learning Neural Network IC Using Highly Linear Charge Trap Device

Jong-Moon Choi, ..., and K. Kwon. IEEE Transactions on Circuits and Systems I: Regular Papers, 2021. 1 citations.

0% Topic Match

Abstract: This paper presents an IC implementation of on-chip learning neural network accelerator using highly linear CMOS-compatible floating gate charge trap devices. A simple learning algorithm utilizing winner-take-all and competitive learning is proposed to design fast and power-efficient hardware. This algorithm was analyzed with behavioral model of emerging non-volatile memory via MATLAB. The linearity, symmetry, and cycle-to-cycle variation of multi-bit switching characteristic affects training accuracy. The proposed content-aware programming technique of modulated column line driver provides flexibility for real-time training while maintaining device linearity, despite having to update a different step for every unit cell and training. The prototype IC is...

[102] The Role of Weight Shrinking in Large Margin Perceptron Learning

C. Panagiotakopoulos and Petroula Tsampouka. ArXiv, 2012. 0 citations.

0% Topic Match

Abstract: We introduce into the classical perceptron algorithm with margin a mechanism that shrinks the current weight vector as a first step of the update. If the shrinking factor is constant the resulting algorithm may be regarded as a margin-error-driven version of NORMA with constant learning rate. In this case we show that the allowed strength of shrinking depends on the value of the maximum margin. We also consider variable shrinking factors for which there is no such dependence. In both cases we obtain new generalizations of the perceptron with margin able to provably attain in a finite number of steps...

[103] Supervised Discrete Hashing

Fumin Shen, ..., and Heng Tao Shen. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. 1106 citations.

0% Topic Match

Abstract: Recently, learning based hashing techniques have attracted broad research interests because they can support efficient storage and retrieval for high-dimensional data such as images, videos, documents, etc. However, a major difficulty of learning to hash lies in handling the discrete constraints imposed on the pursued hash codes, which typically makes hash optimizations very challenging (NP-hard in general). In this work, we propose a new supervised hashing framework, where the learning objective is to generate the optimal binary hash codes for linear classification. By introducing an auxiliary variable, we reformulate the objective such that it can be solved substantially efficiently by...

[104] A Novel Hybrid Linear/Nonlinear Classifier for Two-Class Classification: Theory, Algorithm, and Applications

Weijie Chen, ..., and K. Drukker. IEEE Transactions on Medical Imaging, 2010. 12 citations.

0% Topic Match

No summary or abstract available

[105] Binary Support Vector Machines for Classification

C. Tomasi. Journal Not Provided, 2018. 16 citations.

0% Topic Match

Abstract: Any linear binary classifier separates the space R of all data points x_n into two half-spaces, ideally splitting data points with positive label $y_n = 1$ from those with negative label $y_n = -1$. Of course, the joint model $p(x, y)$ from which the data are drawn may not be linearly separable. As a consequence, linear classifiers are typically formulated so as to allow for some misclassification. Consider placing a decision boundary (a hyperplane in R) where it splits as much of the data as possible correctly into the two classes. After this choice is made, there is often still...

[106] A Greedy Method for Constructing Minimal Multiconlitron

Q. Leng, ..., and Y. Qin. IEEE Access, 2019. 1 citations.

0% Topic Match

Abstract: Multiconlitron is a general theoretical framework for constructing piecewise linear classifier. However, it contains a relatively large number of linear functions, resulting in complicated model structure and poor generalization ability. Learning to prune redundant or excessive components may be a very necessary progression. We propose a novel greedy method, i.e., greedy support multiconlitron algorithm (GreSMA) to simplify the multiconlitron. In GreSMA, a procedure of greedy selection is first used. It generates the initial linear boundaries, each of which can separate maximum number of training samples under the current iteration. In this way, a minimal set of decision functions is established....
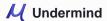
[107] Optimizing ψ-learning via mixed integer programming

Yufeng Liu and Yichao Wu. Statistica Sinica, 2006. 9 citations.

0% Topic Match

Abstract: As a new margin-based classier, ψ-learning shows great potential for high accuracy. However, the optimization of ψ-learning involves non-convex min- imization and is very challenging to implement. In this article, we convert the optimization of ψ-learning into a mixed integer programming (MIP) problem. This enables us to utilize the state-of-art algorithm of MIP to solve ψ-learning. More- over, the new algorithm can solve ψ-learning with a general piecewise linear loss and does not require continuity of the loss function. We also examine the variable selection property of 1-norm ψ-learning and make comparisons with the SVM.

[108] Convolutional-Neural-Network-Based Partial Discharge Diagnosis for Power Transformer Using UHF Sensor

The-Duong Do, ..., and Yong-Hwa Kim. IEEE Access, 2020. 49 citations.

0% Topic Match

Abstract: Given the enormous capital value of power transformers and their integral role in the electricity network, increasing attention has been given to diagnostic and monitoring tools as a safety precaution measure to evaluate the internal condition of transformers. This study overcomes the fault diagnosis problem of power transformers using an ultra high frequency drain valve sensor. A convolutional neural network (CNN) is proposed to classify six types of discharge defects in power transformers. The proposed model utilizes the phase–amplitude response from a phase-resolved partial discharge (PRPD) signal to reduce the input size. The performance of the proposed method is verified...

[109] Isosceles Constraints for Person Re-Identification

Furong Xu, ..., and S. Shan. IEEE Transactions on Image Processing, 2020. 12 citations.

0% Topic Match

Abstract: In the existing works of person re-identification (ReID), batch hard triplet loss has achieved great success. However, it only cares about the hardest samples within the batch. For any probe, there are massive mismatched samples (crucial samples) outside the batch which are closer than the matched samples. To reduce the disruptive influence of crucial samples, we propose a novel isosceles contraint for triplet. Theoretically, we show that if a matched pair has equal distance to any one of mismatched sample, the matched pair should be infinitely close. Motivated by this, the isosceles constraint is designed for the two mismatched pairs...

[110] EFFICIENT MACHINE LEARNING INFERENCE FOR EMBEDDED SYSTEMS WITH INTEGER BASED RESTRICTED BOLTZMANN MACHINES CLASSIFIERS

Bryan Sosa Barillas. Unknown journal, 2019. 2 citations.

0% Topic Match

Abstract: Nowadays, there exist many emerging applications for embedded systems, such as computer vision and speech recognition, which heavily rely on machine learning classification. The typical approach to carry out machine learning related tasks in embedded systems is to use the embedded device as a sensor which collects data, and then carry out the classification computations in the cloud. However, this approach presents issues in several aspects such as latency, power consumption, network bandwidth, privacy, and security. Thus, it would be ideal to perform the machine learning computations in the actual embedded device. Nonetheless, this is not a trivial task as...

[111] An Analog Hardware Classifier for Estimating Sea State or Wave Height from Inertial Sensor Data

Vassilis Alimisis, ..., and P. Sotiriadis. 2024 Panhellenic Conference on Electronics & Telecommunications (PACET), 2024. 0 citations.

0% Topic Match

Abstract: The related research presents a novel methodology for implementing power-efficient analog classifiers, achieving a power consumption of merely 176nW. These classifiers efficiently process multiple input features while sustaining a high degree of precision and minimizing power consumption. It is based on Threshold Machine Learning model, and integrates sigmoid activation circuit along with current comparators. Validation of the implementation was performed employing a real-time vessel dataset, achieving a good accuracy of 82.95% in predicting true waves from inertial sensor data. Furthermore, a comparative analysis was conducted with other analog classifiers employing the identical dataset. All related models were trained using a...

[112] An efficient, provably exact, practical algorithm for the 0-1 loss linear classification problem

Xi He and Max A. Little. ArXiv, 2023. 3 citations.

0% Topic Match

Abstract: Algorithms for solving the linear classification problem have a long history, dating back at least to 1936 with linear discriminant analysis. For linearly separable data, many algorithms can obtain the exact solution to the corresponding 0-1 loss classification problem efficiently, but for data which is not linearly separable, it has been shown that this problem, in full generality, is NP-hard. Alternative approaches all involve approximations of some kind, including the use of surrogates for the 0-1 loss (for example, the hinge or logistic loss) or approximate combinatorial search, none of which can be guaranteed to solve the problem exactly. Finding...

[113] Designing Energy-Efficient Decision Tree Memristor Crossbar Circuits using Binary Classification Graphs

Pranav Sinha and Sunny Raj. 2022 IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2022. 4 citations.

0% Topic Match

Abstract: We propose a method to design in-memory, energy-efficient, and compact memristor crossbar circuits for implementing decision trees using flow-based computing. We develop a new tool called binary classification graph, which is equivalent to decision trees in accuracy but uses bit values of input features to make decisions instead of thresholds. Our proposed design is resilient to manufacturing errors and can scale to large crossbar sizes due to the utilization of sneak paths in computations. Our design uses zero transistor and one memristor (0T1R) crossbars with only two resistance states of high and low, which makes it resilient to resistance drift...

[114] Gradient-based learning applied to document recognition

Yann LeCun, ..., and P. Haffner. Proc. IEEE, 1998. 54613 citations.

0% Topic Match

Abstract: Multilayer neural networks trained with the back-propagation algorithm constitute the best example of a successful gradient based learning technique. Given an appropriate network architecture, gradient-based learning algorithms can be used to synthesize a complex decision surface that can classify high-dimensional patterns, such as handwritten characters, with minimal preprocessing. This paper reviews various methods applied to handwritten character recognition and compares them on a standard handwritten digit recognition task. Convolutional neural networks, which are specifically designed to deal with the variability of 2D shapes, are shown to outperform all other techniques. Real-life document recognition systems are composed of multiple modules including...

[115] A Novel Multiclass Text Classification Algorithm Based on Multiconlitron

Yu-ping Qin, ..., and Aihua Zhang. Unknown journal, 2016. 0 citations.

0% Topic Match

Abstract: A novel multiclass text classification algorithm based on multiconlitron is proposed. The multiconlitron is constructed for each possible pair of classes in sample space, each of which is used to separate two classes. For the sample to be classified, every multiconlitron is used to judge its classman vote for the corresponding class. The final class of the sample is determined by the number of votes. The classification experiments are done on Reuters 21578, and the experimental results show that the proposed algorithm has better performance. Compared with multiclass support vector machines 1-a-1, 1-a-r and DAGSVM, while ensuring the classification accuracy,...

[116] Introduction to Support Vector Machines

D. Boswell. Unknown journal, 2002. 2748 citations.

0% Topic Match

Abstract: Support Vector Machines (SVM's) are a relatively new learning method used for binary classification. The basic idea is to find a hyperplane which separates the d-dimensional data perfectly into its two classes. However, since example data is often not linearly separable, SVM's introduce the notion of a "kernel induced feature space" which casts the data into a higher dimensional space where the data is separable. Typically, casting into such a space would cause problems computationally, and with overfitting. The key insight used in SVM's is that the higher-dimensional space doesn't need to be dealt with directly (as it turns out,...

[117] On a Simple Minkowski Metric Classifier

G. Toussaint. IEEE Trans. Syst. Sci. Cybern., 1970. 5 citations.

0% Topic Match

Abstract: A classifier which, in general, implements a nonlinear decision boundary is shown to be equivalent to a linear discriminant function when the measurements are binary valued; its relation to the Bayes classifier is derived. The classifier requires less computation than a similar one based on the Euclidean distance and can perform equally well.

[118] A VLSI neural network classifier based on integer-valued weights
S. Dr ghici and Damon A. Miller IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339), 1999. 2 citations.
0% Topic Match
No summary or abstract available

[119] Modifying linearly non-separable support vector machine binary classi er to account for the centroid mean vector
No author found. Journal Not Provided, Unknown year. 1 citations.
0% Topic Match
Abstract: This study proposes a modi cation to the objective function of the support vector machine for the linearly non-separable case of a binary classi er y i { 1 , 1 } . The modi cation takes into account the position of each data item x i from its corresponding class centroid. The resulting optimization function involves the centroid mean vector, and the spread of data besides the support vectors, which should be minimized by the choice of hyper-plane ² .Theoretical assumptions have been tested to derive an optimal separable hyperplane that yields the minimal misclassi cation rate. The proposed method has...

[120] A charge-based neural Hamming classifier
U. Çilingiro lu IEEE J. Solid State Circuits, 1993. 41 citations.
0% Topic Match
Abstract: A charge-based fixed-weight neural Hamming classifier with an on-chip normalization facility is described. The classifier utilizes a purely capacitive synapse matrix for quantization and a multiport sense amplifier for discrimination. The discriminator is compatible with variable-weight synapses as well. A detailed analysis of the classifier configuration is presented; design issues are addressed, and limitations are identified. It is shown that the ratio of the maximum Hamming weight to the minimum Hamming distance that can be handled by the classifier has an upper bound. As long as the exemplars comply with this upper bound, the network does not impose any limitation...

[121] Analytical Guarantees on Numerical Precision of Deep Neural Networks
Charbel Sakr, ..., and Naresh R Shanbhag. Unknown journal, 2017. 86 citations.
0% Topic Match
Abstract: The acclaimed successes of neural networks often overshadow their tremendous complexity. We focus on numerical precision a key parameter defining the complexity of neural networks. First, we present theoretical bounds on the accuracy in presence of limited precision. Interestingly, these bounds can be computed via the back-propagation algorithm. Hence, by combining our theoretical analysis and the backpropagation algorithm, we are able to readily determine the minimum precision needed to preserve accuracy without having to resort to timeconsuming fixed-point simulations. We provide numerical evidence showing how our approach allows us to maintain high accuracy but with lower complexity than state-of-the-art binary...

[122] A piecewise linear classifier
A. A. Abdurrab. Unknown journal, 2007. 1 citations.
0% Topic Match
Abstract: A PIECEWISE LINEAR CLASSIFIER Publication No. _____ Abdul Aziz Abdurrab, MSEE The University of Texas at Arlington, 2007 Supervising Professor: Dr. Michael T. Manry A piecewise linear classifier is discussed which classifies N-dimensional input vectors. The network uses a distance measure to assign incoming input vectors to an appropriate cluster. Each cluster has a linear classifier for generating class discriminants. A training algorithm is described for generating the clusters and discriminants. A pruning algorithm is also described. The algorithm is applied after the network has grown completely, i.e, it has achieved the maximum number of clusters. The pruning algorithm eliminates...

[123] On the Generalisation of SoftMargin AlgorithmsJohn
J. Shawe-Taylor and N. Cristianini. Journal Not Provided, 2000. 0 citations.
0% Topic Match
Abstract: Generalisation bounds depending on the margin of a classiier are a relatively recent development. They provide an explanation of the performance of state-of-the-art learning systems such as Support Vector Machines (SVM) and Adaboost. The diiculty with these bounds has been either their dependence on the minimal margin or their agnostic form. The paper presents a technique for correcting those points that fail to meet a target margin, hence creating a large margin classiier at the expense of additional functional complexity. Analysis of this technique leads to bounds that motivate the previously heuristic soft margin SVM algorithms as well as justifying...

[124] LIBSVM: A library for support vector machines
Chih-Chung Chang and Chih-Jen Lin. ACM Trans. Intell. Syst. Technol., 2011. 44062 citations.
0% Topic Match
No summary or abstract available

[125] Transductive Polyhedral Conic Classifiers
Halil Saglamlar and Hakan Cevikalp. 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021. 1 citations.
0% Topic Match
Abstract: In this study, transductive (semi-supervised) polyhedral conic classifiers are proposed. In the proposed method, concave-convex procedure is used to solve the optimization problem as in Robust Transductive Support Vector Machines. This procedure decomposes a non-convex optimization problem into concave and convex parts. Unlike RTSVM that uses SVM formulation, our proposed method uses polyhedral conic classifiers (PCC) that provides tight and closed decision boundaries compared to SVM. In this way, effective, fast and better results can be obtained in large-scale datasets.

[126] A Compact Sequential Classifier for Digital Implementations
R. Dogaru, ..., and M. Glesner. EUROCON 2005 - The International Conference on "Computer as a Tool", 2005. 0 citations.
0% Topic Match
No summary or abstract available

[127] Energy-Ef cient Classi cation: Adaptive Approach
Zafar Takhirov, ..., and A. Joshi. Journal Not Provided, 2017. 0 citations.
0% Topic Match
Abstract: —Designing energy-ef cient machine learning-based systems for mobile and embedded applications is a hard problem – it is not straight forward to meet the energy-delay constraints, while achieving the accuracy goals. Moreover, the energy, delay and accuracy requirements could change at run-time, which could make the design-time optimization sub-optimal. In this work we propose an adaptive classi er that leverages the wide variability in data complexity to dynamically allocate computational resources and improve energy-ef ciency of a hardware accelerator for machine learning. On average, our proposed adaptive classi er can achieve up to H 100 × better energy-ef ciency by trading off H 1% accuracy...

[128] Hinge-Minimax Learner for the Ensemble of Hyperplanes
Dolev Raviv, ..., and Margarita Osadchy. J. Mach. Learn. Res., 2018. 7 citations.
0% Topic Match
Abstract: In this work we consider non-linear classifiers that comprise intersections of hyperplanes. We learn these classifiers by minimizing the "minimax" bound over the negative training examples and the hinge type loss of the positive training examples. These classifiers fit typical real-life datasets that consist of a small number of positive data points and a large number of negative data points. Such an approach is computationally appealing since the majority of training examples (belonging to the negative class) are represented by the statistics of their distribution, which is used in a single constraint on the empirical risk, as opposed to SVM,...

[129] Support-Vector Networks
Corinna Cortes and V. Vapnik. Machine Learning, 1995. 40415 citations.
0% Topic Match
No summary or abstract available

[130] A PAC-Bayesian study of linear classifiers: why SVMs work

R. Herbrich, ..., and P. Bollmann-Sdorra. Unknown journal, 2000. 1 citations.

0% Topic Match

Abstract: In this paper we present a bound on the generalisation error of linear classi ers in terms of a re ned margin quantity on the training set. The result is obtained by a fundamentally di erent reasoning than in the classical PAC framework. We show that the generalisation error of a classi er can be bounded purely by geometrical arguments. In contrast to the classical results our bound does not contain awkward constants (see, e.g. [10, p. xxi]) and is an exponential improvement of the result in [8] by scaling inverse logarithmically in the margin. In the form presented the...

[131] BinaryNet: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1

Matthieu Courbariaux and Yoshua Bengio. ArXiv, 2016. 683 citations.

0% Topic Match

Abstract: We introduce BinaryNet, a method which trains DNNs with binary weights and activations when computing parameters' gradient. We show that it is possible to train a Multi Layer Perceptron (MLP) on MNIST and ConvNets on CIFAR-10 and SVHN with BinaryNet and achieve nearly state-of-the-art results. At run-time, BinaryNet drastically reduces memory usage and replaces most multiplications by 1-bit exclusive-not-or (XNOR) operations, which might have a big impact on both general-purpose and dedicated Deep Learning hardware. We wrote a binary matrix multiplication GPU kernel with which it is possible to run our MNIST MLP 7 times faster than with an unoptimized...

[132] Solving a graph layout problem using strictly digital neural networks with virtual slack-neurons

K. Murakami and T. Nakagawa. IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339), 1999. 1 citations.

0% Topic Match

No summary or abstract available

[133] DURIAN RECOGNITION BASED ON MULTIPLE FEATURES AND LINEAR DISCRIMINANT ANALYSIS

M. Mustaffa, ..., and Nurul Amelina Nasharuddin. Malaysian Journal of Computer Science, 2018. 10 citations.

0% Topic Match

Abstract: Many fruit recognition approaches today are designed to classify different type of fruits but there is little effort being done for content-based fruit recognition specifically focuses on durian species. Durian, known as the king of tropical fruits, have few similar characteristics between different species where the skin have almost the same colour from green to yellowish brown with just slightly different shape and pattern of thorns. Therefore, it is hard to differentiate them with the current methods. It would be valuable to have an automated content-based recognition framework that can automatically represent and recognise a durian species given a durian...

[134] A 10T Non-Precharge Two-Port SRAM for 74% Power Reduction in Video Processing

Hiroki Noguchi, ..., and M. Yoshimoto. IEEE Computer Society Annual Symposium on VLSI (ISVLSI '07), 2007. 58 citations.

0% Topic Match

No summary or abstract available

[135] Tripos Part IIB FOURTH YEAR Paper 4 F 10 : Statistical Pattern Processing STATISTICAL PATTERN RECOGNITION Solutions to Examples Paper 2 1

No author found. Journal Not Provided, Unknown year. 0 citations.

0% Topic Match

Abstract: 1. As the name implies linear classifiers only generate decision boundaries of the form w 2 x + b = 0. Non linear mappings of the feature can increase the effective dimen-sionality. A linear decision boundary in this mapped space will be non-linear in the original space. Note there is an increase in the number of model parameters that need to be trained for the decision boundary. A mapping will exist if the points have distinct labels (i.e. no point has multiple class labels associated with it.) 2. The conditions that must be satisfied are: ⊥ e 0 m...

[136] Fuzlearn: A Fuzzy Clusterization based Machine Learning using Learning Automata

Ivan Shmarov, ..., and A. Yakovlev. 2022 International Symposium on the Tsetlin Machine (ISTM), 2022. 0 citations.

0% Topic Match

Abstract: Existing machine learning (ML) algorithms, such as Neural Networks (NNs), typically consist of complex data encoding and preprocessing on real-world data followed by extensive arithmetic. As such when implemented in hardware or software they incur high energy consumption as well as increased data-to-decision latency. Recently, the Tsetlin Machine (TM) was proposed as a promising new alternative which takes advantage of logic propositions founded on the principle of Tsetlin Automata, leading to higher energy efficiency, faster training convergence and lower latency. However, TM interfaces require complex data encoding procedures which can add significant overheads.In this paper, we propose a new machine...

[137] The crossing number for neural networks

D. Durfee and J. Savage. International 1989 Joint Conference on Neural Networks, 1989. 1 citations.

0% Topic Match

No summary or abstract available

[138] Supervised Learning Methods

A. Einstein. Journal Not Provided, 2013. 3 citations.

0% Topic Match

Abstract: Today we discuss the classification setting in detail. Our setting is that we observe for each subject i a set of p predictors (covariates) xi, and qualitative outcomes (or classes) gi, which can takes values from a discrete set G. In this class we assume that predictors are genomic measurements (gene expression for now), and that we have many more measurements than samples (i.e. p << N , where N is the number of subjects). For gene expression, predictors are real numbers, and we can think of input, or feature, space, as Euclidean space. Since our prediction (x) will always...

[139] Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab

Eko Prasetyo. Unknown journal, 2015. 225 citations.

0% Topic Match

No summary or abstract available

[140] Neural computation for controlling the configuration of 2-dimensional truss structure

K. Tsutsumi, ..., and H. Matsumoto. IEEE 1988 International Conference on Neural Networks, 1988. 20 citations.
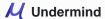
0% Topic Match

No summary or abstract available

[141] Deep Learning with Limited Numerical Precision

Suyog Gupta, ..., and P. Narayanan. Unknown journal, 2015. 2053 citations.

Not measured Topic Match

Abstract: Training of large-scale deep neural networks is often constrained by the available computational resources. We study the effect of limited precision data representation and computation on neural network training. Within the context of low-precision fixed-point computations, we observe the rounding scheme to play a crucial role in determining the network's behavior during training. Our results show that deep networks can be trained using only 16-bit wide fixed-point number representation when using stochastic rounding, and incur little to no degradation in the classification accuracy. We also demonstrate an energy-efficient hardware accelerator that implements low-precision fixed-point arithmetic with stochastic rounding.

[142] Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks
Yu-hsin Chen, ..., and V. Sze. IEEE Journal of Solid-State Circuits, 2016. 2502 citations.
Not measured Topic Match
No summary or abstract available

[143] Bitwise Neural Networks
Minje Kim and Paris Smaragdis. ArXiv, 2016. 218 citations.
Not measured Topic Match
Abstract: Based on the assumption that there exists a neural network that efficiently represents a set of Boolean functions between all binary inputs and outputs, we propose a process for developing and deploying neural networks whose weight parameters, bias terms, input, and intermediate hidden layer output signals, are all binary-valued, and require only basic bit logic for the feedforward pass. The proposed Bitwise Neural Network (BNN) is especially suitable for resource-constrained environments, since it replaces either floating or fixed-point arithmetic with significantly more efficient bitwise operations. Hence, the BNN requires for less spatial complexity, less memory bandwidth, and less power consumption...

[144] XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks
Mohammad Rastegari, ..., and Ali Farhadi. Unknown journal, 2016. 4378 citations.
Not measured Topic Match
Abstract: We propose two efficient approximations to standard convolutional neural networks: Binary-Weight-Networks and XNOR-Networks. In Binary-Weight-Networks, the filters are approximated with binary values resulting in $32\times$ memory saving. In XNOR-Networks, both the filters and the input to convolutional layers are binary. XNOR-Networks approximate convolutions using primarily binary operations. This results in $58\times$ faster convolutional operations (in terms of number of the high precision operations) and $32\times$ memory savings. XNOR-Nets offer the possibility of running state-of-the-art networks on CPUs (rather than GPUs) in real-time. Our binary networks are simple, accurate, efficient, and work on challenging visual tasks. We evaluate...

[145] Finite-precision analysis of the pipelined strength-reduced adaptive filter
M. Goel and Naresh R Shanbhag. IEEE Trans. Signal Process., 1998. 21 citations.
Not measured Topic Match
Abstract: In this correspondence, we compare the finite-precision requirements of the traditional cross-coupled (CC) and a low-power strength-reduced (SR) architectures. It is shown that the filter block (F block) coefficients in the SR architecture require 0.3 bits more than the corresponding block in the CC architecture. Similarly, the weight-update (WUD) block in the SR architecture is shown to require 0.5 bits fewer than the corresponding block in the CC architecture. This finite-precision architecture is then used as a near-end crosstalk (NEXT) canceller for 155.52 Mb/s ATM-LAN over unshielded twisted pair (UTP) category-3 cable. Simulation results are presented in support of the...

[146] A 4 + 2T SRAM for Searching and In-Memory Computing With 0.3-V V_{DDmin}
Qing Dong, ..., and D. Sylvester. IEEE Journal of Solid-State Circuits, 2018. 34 citations.
Not measured Topic Match
No summary or abstract available

[147] Error Adaptive Classifier Boosting (EACB): Leveraging Data-Driven Training Towards Hardware Resilience for Signal Inference
Zhuo Wang, ..., and N. Verma. IEEE Transactions on Circuits and Systems I: Regular Papers, 2015. 34 citations.
Not measured Topic Match
No summary or abstract available

[148] Deep learning with COTS HPC systems
Adam Coates, ..., and A. Ng. Unknown journal, 2013. 728 citations.
Not measured Topic Match
Abstract: Scaling up deep learning algorithms has been shown to lead to increased performance in benchmark tasks and to enable discovery of complex high-level features. Recent efforts to train extremely large networks (with over 1 billion parameters) have relied on cloudlike computing infrastructure and thousands of CPU cores. In this paper, we present technical details and results from our own system based on Commodity Off-The-Shelf High Performance Computing (COTS HPC) technology: a cluster of GPU servers with Infiniband interconnects and MPI. Our system is able to train 1 billion parameter networks on just 3 machines in a couple of days, and...

[149] Low-energy Formulations of Support Vector Machine Kernel Functions for Biomedical Sensor Applications
Kyong-Ho Lee, ..., and N. Verma. Journal of Signal Processing Systems, 2012. 39 citations.
Not measured Topic Match
No summary or abstract available

[150] A roundoff error analysis of the LMS adaptive algorithm
C. Caraiscos and Bede Liu. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984. 226 citations.
Not measured Topic Match
Abstract: The steady state output error of the least mean square (LMS) adaptive algorithm due to the finite precision arithmetic of a digital processor is analyzed. It is found to consist of three terms: 1) the error due to the input data quantization, 2) the error due to the rounding of the arithmetic operations in calculating the filter's output, and 3) the error due to the deviation of the filter's coefficients from the values they take when infinite precision arithmetic is used. The last term is of particular interest because its mean squared value is inversely proportional to the adaptation step...

[151] An Analytical Method to Determine Minimum Per-Layer Precision of Deep Neural Networks
Charbel Sakr and Naresh R Shanbhag. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018. 35 citations.
Not measured Topic Match
No summary or abstract available

[152] Hybrid-cell register files design for improving NBTI reliability
Na Gong, ..., and R. Sridhar. Microelectron. Reliab., 2012. 15 citations.
Not measured Topic Match
No summary or abstract available

[153] Optimization Methods for Large-Scale Machine Learning
L. Bottou, ..., and J. Nocedal. ArXiv, 2016. 3231 citations.
Not measured Topic Match
Abstract: This paper provides a review and commentary on the past, present, and future of numerical optimization algorithms in the context of machine learning applications. Through case studies on text classification and the training of deep neural networks, we discuss how optimization problems arise in machine learning and what makes them challenging. A major theme of our study is that large-scale machine learning represents a distinctive setting in which the stochastic gradient (SG) method has traditionally played a central role while conventional gradient-based nonlinear optimization techniques typically falter. Based on this viewpoint, we present a comprehensive theory of a straightforward, yet...

[154] Supervised hashing with adaptive discrete optimization for multimedia retrieval
Sixiu Chen, ..., and Jingkuan Song. Neurocomputing, 2017. 12 citations.
Not measured Topic Match
No summary or abstract available

[155] Minimax rates for memory-bounded sparse linear regression

J. Steinhardt and John C. Duchi. Unknown journal, 2015. 47 citations.
Not measured Topic Match

Abstract: We establish a minimax lower bound of kd on the sample size needed to estimate parameters in a k-sparse linear regression of dimension d under memory restrictions to B bits, where is the '2 parameter error. When the covariance of the regressors is the identity matrix, we also provide an algorithm that uses ~ O(B +k) bits and requires ~ O(kd B 2) observations to achieve error . Our lower bound holds in a more general communication-bounded setting, where instead of a memory bound, at mostB bits of information are allowed to be (adaptively) communicated about each sample.

[156] Explaining and Harnessing Adversarial Examples
I. Goodfellow, ..., and Christian Szegedy. CoRR, 2014. 19198 citations.
Not measured Topic Match

Abstract: Several machine learning models, including neural networks, consistently misclassify adversarial examples---inputs formed by applying small but intentionally worst-case perturbations to examples from the dataset, such that the perturbed input results in the model outputting an incorrect answer with high confidence. Early attempts at explaining this phenomenon focused on nonlinearity and overfitting. We argue instead that the primary cause of neural networks' vulnerability to adversarial perturbation is their linear nature. This explanation is supported by new quantitative results while giving the first explanation of the most intriguing fact about them: their generalization across architectures and training sets. Moreover, this view yields...

[157] Deep learning in neural networks: An overview
J. Schmidhuber. Neural networks : the official journal of the International Neural Network Society, 2014. 16459 citations.
Not measured Topic Match

No summary or abstract available

[158] Countermeasures against NBTI degradation on 6T-SRAM cells
E. Glocker, ..., and S. Drapatz. Advances in Radio Science, 2011. 13 citations.
Not measured Topic Match

Abstract: Abstract. In current process technologies, NBTI (negative bias temperature instability) has the most severe aging effect on static random access memory (SRAM) cells. This degradation effect causes loss of stability. In this paper countermeasures against this hazard are presented and quantified via simulations in 90 nm process technologies by the established metrics SNMread, SNMhold, Iread and Write Level. With regard to simulation results and practicability best candidates are chosen and, dependent on individual preferences at memory cell design, the best countermeasure in each case is recommended.

[159] Impact of Negative-Bias Temperature Instability in Nanoscale SRAM Array: Modeling and Analysis
Kunhyuk Kang, ..., and M. Alam. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2007. 174 citations.
Not measured Topic Match

No summary or abstract available

[160] Penelope: The NBTI-Aware Processor
J. Abella, ..., and Antonio González. 40th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 2007), 2007. 194 citations.
Not measured Topic Match

No summary or abstract available

[161] NBTI resistant SRAM design
Fahad Ahmed and L. Milor. 2011 4th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), 2011. 20 citations.
Not measured Topic Match

No summary or abstract available

[162] Analysis and mitigation of NBTI aging in register file: An end-to-end approach
Saurabh Kothawade, ..., and Sanghamitra Roy. 2011 12th International Symposium on Quality Electronic Design, 2011. 31 citations.
Not measured Topic Match

No summary or abstract available

[163] An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches
Leland Chang, ..., and D. Jamsek. IEEE Journal of Solid-State Circuits, 2008. 445 citations.
Not measured Topic Match

No summary or abstract available

[164] Incremental and parallel proximal SVM algorithm tailored on the Jetson Nano for the ImageNet challenge
Thanh-Nghi Do. Int. J. Web Inf. Syst., 2022. 4 citations.
Not measured Topic Match

Abstract: Purpose This paper aims to propose the new incremental and parallel training algorithm of proximal support vector machines (Inc-Par-PSVM) tailored on the edge device (i.e. the Jetson Nano) to handle the large-scale ImageNet challenging problem. Design/methodology/approach The Inc-Par-PSVM trains in the incremental and parallel manner ensemble binary PSVM classifiers used for the One-Versus-All multiclass strategy on the Jetson Nano. The binary PSVM model is the average in bagged binary PSVM models built in undersampling training data block. Findings The empirical test results on the ImageNet data set show that the Inc-Par-PSVM algorithm with the Jetson Nano (Quad-core ARM A57...

[165] Predicting thermophysical properties of dialkylimidazolium ionic liquids from sigma profiles
Oscar Nordness, ..., and J. Brennecke. Journal of Molecular Liquids, 2021. 20 citations.
Not measured Topic Match

No summary or abstract available

[166] Comparison between fuzzy kernel k-medoids using radial basis function kernel and polynomial kernel function in hepatitis classification
G. Saragih, ..., and Zuherman Rustam. IAES International Journal of Artificial Intelligence, 2021. 0 citations.
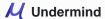Not measured Topic Match

Abstract: This paper compares the fuzzy kernel k-medoids using radial basis function (RBF) and polynomial kernel function in hepatitis classification. These two kernel functions were chosen due to their popularity in any kernel-based machine learning method for solving the classification task. The hepatitis dataset then used to evaluate the performance of both methods that were expected to provide an accurate diagnosis in patients to obtain treatment at an early phase. The data were obtained from two hospitals in Indonesia, consisting of 89 hepatitis-B and 31 hepatitis-C samples. The data were analyzed using several cases of k-fold cross-validation, and the performances were...

[167] Estimation of particulate matter (PM2.5, PM10) concentration and its variation over urban sites in Bangladesh
Amitesh Gupta, ..., and K. Mondal. SN Applied Sciences, 2020. 26 citations.
Not measured Topic Match

No summary or abstract available

[168] Critical exponents for total positivity, individual kernel encoders, and the Jain-Karlin-Schoenberg kernel.
A. Khare. arXiv: Functional Analysis, 2020. 5 citations.
Not measured Topic Match

Abstract: We prove the converse to a result of Karlin [Trans. AMS 1964], and also strengthen his result and two results of Schoenberg [Ann. of Math. 1955]. One of the latter results concerns zeros of Laplace transforms of multiply positive functions. The other results study which powers α of two specific kernels are totally non-negative of order $p\geq 2$ (denoted TN$\_p$); both authors showed this happens for $\alpha\geq p-2$, and Schoenberg proved that it does not for $\alpha p-2$, and is not TN$\_p$ for every $\alpha\in(0,p-2)\setminus\mathbb{Z}$. In particular, these results reveal a 'critical exponent' phenomenon in total positivity. (This exponent $(p-2)$$...

[169] Assessing the impact of parameters tuning in ensemble based breast Cancer classification
A. Idri, ..., and Ibtissam Abnane. Health and Technology, 2020. 30 citations.
Not measured Topic Match

No summary or abstract available

[170] A model to predict ammonia emission using a modified genetic artificial neural network: Analyzing cement mixed with fly ash from a coal-fired power plant
H. Jang and Shuli Xing. Construction and Building Materials, 2020. 27 citations.
Not measured Topic Match

No summary or abstract available

[171] N-GlyDE: a two-stage N-linked glycosylation site prediction incorporating gapped dipeptides and pattern-based encoding
Thejkiran Pitti, ..., and Ting-Yi Sung. Scientific Reports, 2019. 55 citations.
Not measured Topic Match

Abstract: N-linked glycosylation is one of the predominant post-translational modifications involved in a number of biological functions. Since experimental characterization of glycosites is challenging, glycosite prediction is crucial. Several predictors have been made available and report high performance. Most of them evaluate their performance at every asparagine in protein sequences, not confined to asparagine in the N-X-S/T sequon. In this paper, we present N-GlyDE, a two-stage prediction tool trained on rigorously-constructed non-redundant datasets to predict N-linked glycosites in the human proteome. The first stage uses a protein similarity voting algorithm trained on both glycoproteins and non-glycoproteins to predict a score for...

[172] Automatic Learning Algorithms for Local Support Vector Machines
Thanh-Nghi Do. SN Computer Science, 2019. 5 citations.
Not measured Topic Match

No summary or abstract available

[173] Data scarcity, robustness and extreme multi-label classification
Rohit Babbar and B. Scholkopf. Machine Learning, 2019. 150 citations.
Not measured Topic Match

Abstract: The goal in extreme multi-label classification (XMC) is to learn a classifier which can assign a small subset of relevant labels to an instance from an extremely large set of target labels. The distribution of training instances among labels in XMC exhibits a long tail, implying that a large fraction of labels have a very small number of positive training instances. Detecting tail-labels, which represent diversity of the label space and account for a large fraction (upto 80%) of all the labels, has been a significant research challenge in XMC. In this work, we pose the tail-label detection task in...

[174] Gene selection via BPSO and Backward generation for cancer classification
Ahmed Bir-Jmel, ..., and S. Bernoussi. RAIRO Oper. Res., 2019. 5 citations.
Not measured Topic Match

Abstract: Gene expression data (DNA microarray) enable researchers to simultaneously measure the levels of expression of several thousand genes. These levels of expression are very important in the classification of different types of tumors. In this work, we are interested in gene selection, which is an essential step in the data pre-processing for cancer classification. This selection makes it possible to represent a small subset of genes from a large set, and to eliminate the redundant, irrelevant or noisy genes. The combinatorial nature of the selection problem requires the development of specific techniques such as filters and Wrappers, or hybrids combining...

[175] Automatic Hyper-parameters Tuning for Local Support Vector Machines
Thanh-Nghi Do and Minh-Thu Tran-Nguyen. Unknown journal, 2018. 1 citations.
Not measured Topic Match

No summary or abstract available

[176] An Efficient Alternating Newton Method for Learning Factorization Machines
Wei-Sheng Chin, ..., and Chih-Jen Lin. ACM Transactions on Intelligent Systems and Technology (TIST), 2018. 8 citations.
Not measured Topic Match

Abstract: To date, factorization machines (FMs) have emerged as a powerful model in many applications. In this work, we study the training of FM with the logistic loss for binary classification, which is a nonlinear extension of the linear model with the logistic loss (i.e., logistic regression). For the training of large-scale logistic regression, Newton methods have been shown to be an effective approach, but it is difficult to apply such methods to FM because of the nonconvexity. We consider a modification of FM that is multiblock convex and propose an alternating minimization algorithm based on Newton methods. Some novel optimization...

[177] Galvanic Skin Response Data Classification for Emotion Detection
D. Setyohadi, ..., and A. S. Prabuwono. International Journal of Electrical and Computer Engineering (IJECE), 2018. 25 citations.
Not measured Topic Match

Abstract: Emotion detection is a very exhausting job and needs a complicated process; moreover, these processes also require the proper data training and appropriate algorithm. The process involves the experimental research in psychological experiment and classification methods. This paper describes a method on detection emotion using Galvanic Skin Response (GSR) data. We used the Positive and Negative Affect Schedule (PANAS) method to get a good data training. Furthermore, Support Vector Machine and a correct preprocessing are performed to classify the GSR data. To validate the proposed approach, Receiver Operating Characteristic (ROC) curve, and accuracy measurement are used. Our method shows that...

[178] Recurrent Neural Network for Partial Discharge Diagnosis in Gas-Insulated Switchgear
M. Nguyen, ..., and Yong-Hwa Kim. Energies, 2018. 73 citations.
Not measured Topic Match

Abstract: The analysis of partial discharge (PD) signals has been identified as a standard diagnostic tool for monitoring the condition of different electrical apparatuses. This study proposes an approach to detecting PD patterns in gas-insulated switchgear (GIS) using a long short-term memory (LSTM) recurrent neural network (RNN). The proposed method uses phase-resolved PD (PRPD) signals as input, extracts low-level features, and finally, classifies faults in GIS. In the proposed method, LSTM networks can learn temporal dependencies directly from PRPD signals. Most existing models use support vector machines (SVMs) and mainly focus on improving feature representation and extraction manually to analyze PRPD...

[179] Engagement Prediction in the Adaptive Learning Model for Students with Dyslexia
Siti Suhaila Abdul Hamid, ..., and A. Kamaruddin. Proceedings of the 4th International Conference on Human-Computer Interaction and User Experience in Indonesia, CHIuXiD '18, 2018. 6 citations.
Not measured Topic Match

No summary or abstract available

[180] Dyslexia Adaptive Learning Model: Student Engagement Prediction Using Machine Learning Approach
Siti Suhaila Abdul Hamid, ..., and A. Ghani. Unknown journal, 2018. 31 citations.

Not measured Topic Match

No summary or abstract available

[181] Parallel Algorithm of Local Support Vector Regression for Large Datasets

Le-Diem Bui, ..., and Thanh-Nghi Do. Unknown journal, 2017. 4 citations.

Not measured Topic Match

No summary or abstract available

[182] Material Classification using Neural Networks

Anca Sticlaru. ArXiv, 2017. 4 citations.

Not measured Topic Match

Abstract: The recognition and classification of the diversity of materials that exist in the environment around us are a key visual competence that computer vision systems focus on in recent years. Understanding the identification of materials in distinct images involves a deep process that has made usage of the recent progress in neural networks which has brought the potential to train architectures to extract features for this challenging task. This project uses state-of-the-art Convolutional Neural Network (CNN) techniques and Support Vector Machine (SVM) classifiers in order to classify materials and analyze the results. Building on various widely used material databases collected,...

[183] Etude expérimentale des dynamiques temporelles du comportement normal et pathologique chez le rat et la souris

Maria Isabel Carreno-Mun oz Unknown journal, 2017. 0 citations.

Not measured Topic Match

Abstract: Le developpement d'outils de phenotypage comportemental sophistiques est indispensable pour comprendre le fonctionnement cognitif. A partir d'une analyse elaboree de tests comportementaux classiques, mes resultats suggerent que l'hypersensibilite sensorielle associee a un canal potassique specifique (BkCa) participe aux divers troubles comportementaux du syndrome de l'X-Fragile et du spectre autistique. Grâce a un dispositif experimental nouveau et original, comprenant des capteurs de pression hyper-sensibles a meme de detecter les moindres mouvement d'un rat ou d'une souris avec une sensibilite et une precision temporelle exceptionnelles, j'ai pu identifier des composantes comportementales normales et pathologiques inedites, telles que des tremblements ou la dynamique...

[184] Time Series Classification by Sequence Learning in All-Subsequence Space

Thach Le Nguyen, ..., and Georgiana Ifrim. 2017 IEEE 33rd International Conference on Data Engineering (ICDE), 2017. 47 citations.

Not measured Topic Match

Abstract: Existing approaches to time series classification can be grouped into shape-based (numeric) and structure-based (symbolic). Shape-based techniques use the raw numeric time series with Euclidean or Dynamic Time Warping distance and a 1-Nearest Neighbor classifier. They are accurate, but computationally intensive. Structure-based methods discretize the raw data into symbolic representations, then extract features for classifiers. Recent symbolic methods have outperformed numeric ones regarding both accuracy and efficiency. Most approaches employ a bag-of-symbolic-words representation, but typically the word-length is fixed across all time series, an issue identified as a major weakness in the literature. Also, there are no prior attempts to...

[185] Parallel Learning of Local SVM Algorithms for Classifying Large Datasets

Thanh-Nghi Do and F. Poulet. Trans. Large Scale Data Knowl. Centered Syst., 2016. 22 citations.

Not measured Topic Match

No summary or abstract available

[186] Conjugate descent for the SMO algorithm

A. Torres-Barrán and José R. Dorronsoro. 2016 International Joint Conference on Neural Networks (IJCNN), 2016. 3 citations.

Not measured Topic Match

No summary or abstract available

[187] Classification methods for handwritten digit recognition: A survey

Ira Tuba, ..., and M. Veinovic. Vojnotehnicki glasnik, 2023. 1 citations.

Not measured Topic Match

Abstract: Introduction/purpose: This paper provides a survey of handwritten digit recognition methods tested on the MNIST dataset. Methods: The paper analyzes, synthesizes and compares the development of different classifiers applied to the handwritten digit recognition problem, from linear classifiers to convolutional neural networks. Results: Handwritten digit recognition classification accuracy tested on the MNIST dataset while using training and testing sets is now higher than 99.5% and the most successful method is a convolutional neural network. Conclusions: Handwritten digit recognition is a problem with numerous real-life applications. Accurate recognition of various handwriting styles, specifically digits is a task studied for decades and...

[188] Data-Enabled Distribution Grid Management

Z. Hosseini. Journal Not Provided, 2023. 0 citations.

Not measured Topic Match

Abstract: In 2020, U.S. electric utilities installed more than 94 million advanced meters, which brought the percentage of residential customers equipped with smart meters to 75%. This significant investment allows collecting extensive customer data at the distribution level, however, the data are not currently leveraged effectively to help with system operations. This dissertation aims to use the smart meters' data to improve the grid's reliability, stability, and controllability by solving two of the most challenging problems at the distribution level, namely distribution network phase identification and outage identification. Distribution networks have typically been the least observable and most dynamic and locally...

[189] Random Partition Based Adaptive Distributed Kernelized SVM for Big Data

Amrit Pal, ..., and Manish Kumar. IEEE Access, 2022. 2 citations.

Not measured Topic Match

Abstract: In this paper, we present a distributed classification technique for big data by efficiently using distributed storage architecture and data processing units of a cluster. While handling such large data, the existing approaches consider specific data partitioning techniques which demand complete data be processed before partitioning. This leads to an excessive overhead of high computation and data communication. The proposed method does not require any pre-structured data partitioning technique and is also adaptive to big data mining tools. We hypothesize that an effective aggregation of the information generated from data partitions by subprocesses of the complete learning process can lead...

[190] Argument Identification in Indonesian Tweets on the Issue of Moving the Indonesian Capital

Amalia Huwaidah, ..., and Said al Faraby. Procedia Computer Science, 2021. 3 citations.

Not measured Topic Match

No summary or abstract available

[191] Hybridisation of Optimised Support Vector Machine and Artificial Neural Network for Diabetic Retinopathy Classification

Nur Izzati Ab Kader, ..., and Maziani Sabudin. Unknown journal, 2020. 0 citations.

Not measured Topic Match

No summary or abstract available

[192] Parallel Learning Algorithms of Local Support Vector Regression for Dealing with Large Datasets

Thanh-Nghi Do and Le-Diem Bui. Trans. Large Scale Data Knowl. Centered Syst., 2019. 2 citations.

Not measured Topic Match

No summary or abstract available

[193] Diabetic Retinopathy Classification using Support Vector Machine with Hyperparameter Optimization

Nur Izzati, ..., and S. Naim. Journal Not Provided, 2019. 5 citations.

Not measured Topic Match

Abstract: Diabetic Retinopathy (DR) is one of the frequent comorbidities of diabetes worldwide. Diabetic eye screening has become a challenging task for ophthalmologist as they need to deal with large number of patients to be diagnosed, creating a need to develop tool that may help ophthalmologist to classify the severity of DR in order to establish an adequate therapy. Previous researchers have studied machine learning to propose an automatic DR classification using the clinical variables. However, it needs to be improvised especially in terms of accuracy. Hence, this paper aims to propose an optimal or near-optimal DR classifier using the Support...

[194] A Study of Diabetic Retinopathy Classification Using Support Vector Machine

Nur Izzati, ..., and S. Naim. Journal Not Provided, 2019. 0 citations.

Not measured Topic Match

Abstract: Diabetic Retinopathy (DR) is a diabetic complication which can cause blindness. As DR cases keep increasing, ophthalmologists are forced to diagnose a large number of retinal images daily. Generally, the diabetic eye screening is done manually using qualitative scale to detect abnormalities on the retina. Although this approach is useful, the detection is not accurate; and create a need for a tool that can help the experts to classify the severity of DR to establish adequate therapy. Previous researchers have studied machine learning to pro-pose an automatic DR classification. However, it needs to be improvised especially in terms of accuracy....