USE OF 5-NEIGHBORHOOD CELLULAR AUTOMATA FOR PATTERN CLASSIFIER

Sukanya Mukherjee and Priya Kumari
Institute of Engineering & Management, Saltlake

ASCAT 2025

Motivation

Convergence property of Cellular Automaton

- ability to evolve to fixed point configurations (attractors) – theoretical aspect

- to effectively separate and classify patterns – good classifier

- Effectiveness of Elementary CA (2-state 3-neighborhood)

- 3-state 3-neighborhood – strictly single length cycle CA

Challenge

Count of states = 2; not restricted to 3-neighborhood – extended for *5-neighborhood*

CA Model

Strictly single length cycle CA

Multi Attractor CA (MACA)

CA Model

Strictly single length cycle CA

Multi Attractor CA (MACA)

Stable Cyclic behaviour

Moderate Number of Attractors

CA Model

Strictly single length cycle CA

Multi Attractor CA (MACA)

Stable Cyclic behaviour

Moderate Number of Attractors

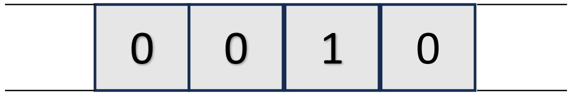
Balanced Rules

Unbalanced rules

Contributions

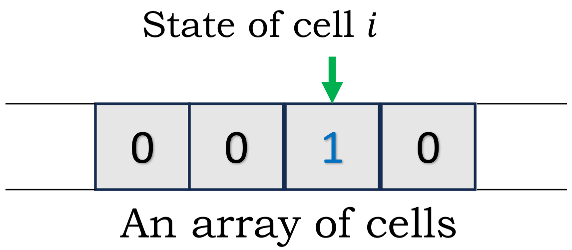
- Study of cycle structure of 2-state 5-neighborhood Strictly single length cycle CA - balanced and unbalanced rules
- Effectiveness in pattern classification

Terminologies



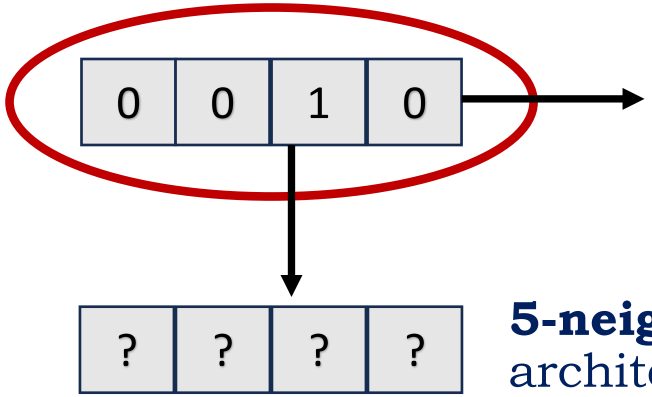
An array of cells

Terminologies



2-state CA: 0, 1

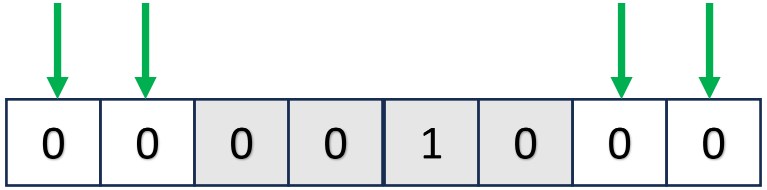
Terminologies



Configuration

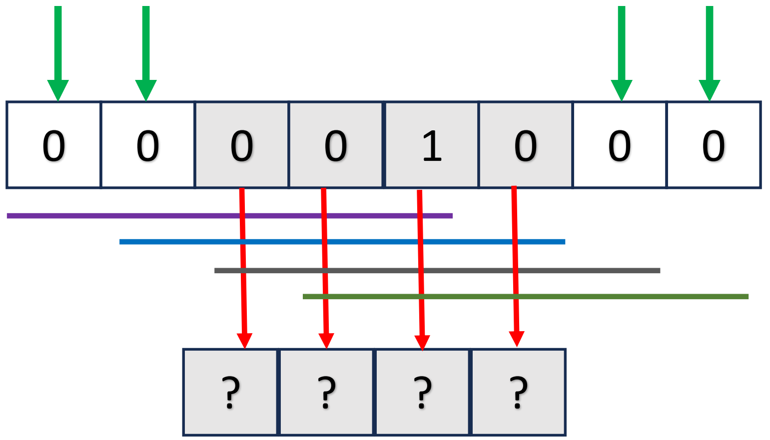
5-neighborhood architecture and **Rules**

Terminologies



Finite CA: Null boundary

Terminologies



Finite CA: Null boundary

Terminologies

| 0 | 0 | 1 | 0 |

Configuration

| ? | ? | ? | ? |

5-neighborhood architecture and **Rules**

0 0 | 0 | 0 | 1 | 0 | 0 0

Rule Min Term (RMT): states of cells $i-2$, $i-1$, i, $i+1$ and $i+2$.

RMT

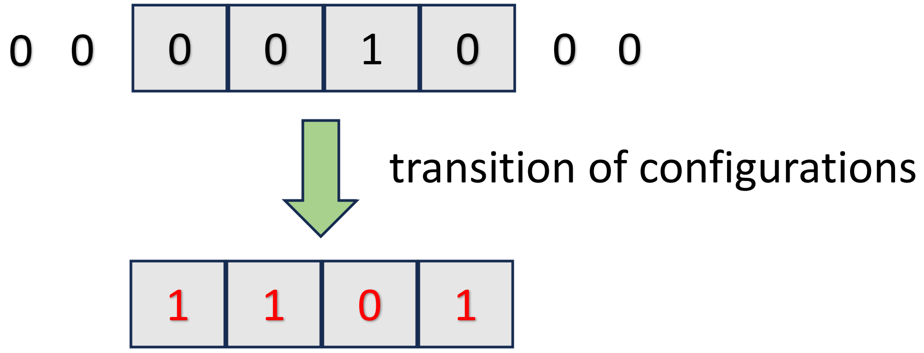
00100

$2^4.(0) + 2^3.(0) + 2^2.(1) + 2^1.(0) + 2^0.(0) = 4$

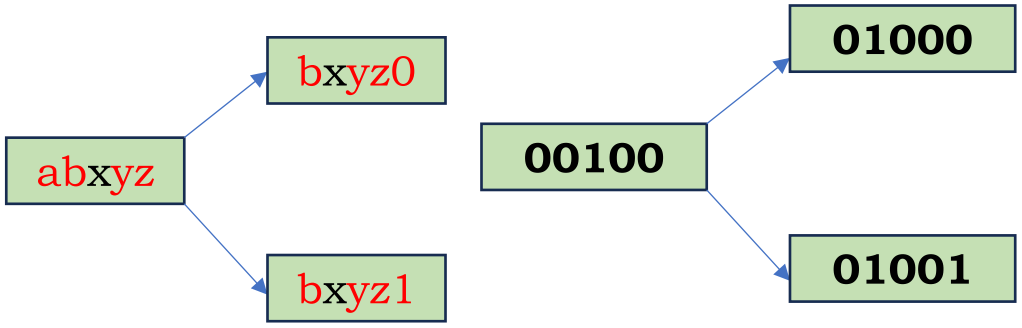
$R(0,0,1,0,0) = R[4] = 0$

RMT	11111 (31)	11110 (30)	11101 (29)	11100 (28)	11011 (27)	…….	01000 (8)	00111 (7)	00110 (6)	00101 (5)	00100 (4)	00011 (3)	00010 (2)	00001 (1)	00000 (0)
R	0	1	1	1	0	……..	1	0	1	1	0	1	1	1	0

Terminologies



RMTs for consecutive cells



RMT	11111 (31)	11110 (30)	11101 (29)	11100 (28)	11011 (27)	....	01000 (8)	00111 (7)	00110 (6)	00101 (5)	00100 (4)	00011 (3)	00010 (2)	00001 (1)	00000 (0)
R	0	1	1	1	0	....	1	0	1	1	0	1	1	1	0

Terminologies

0 0 | 0 | 0 | 1 | 0 | 0 0

RMT Sequence – (1, 2, 4, 8)

| 1 | 1 | 0 | 1 |

RMT	11111 (31)	11110 (30)	11101 (29)	11100 (28)	11011 (27)	…….	01000 (8)	00111 (7)	00110 (6)	00101 (5)	00100 (4)	00011 (3)	00010 (2)	00001 (1)	00000 (0)
R	0	1	1	1	0	……..	1	0	1	1	0	1	1	1	0

Terminologies

Possible RMTs $2^5 = 32$ → Rule space $2^{32} = 4294967296$

Uniform CA; Every cell uses same rule R.

2^{32} cellular automata for given CA size n.

Terminologies

Balanced Rule: Equal number of 0s and 1s: 00110011000011110110001101010110

Unbalanced Rule: Equal number of 0s and 1s: **1**0**11**00**11**0000**1111**0**11**000**11**0**1**0**1**0**11**0

Terminologies

Self Replicating RMT: If $ab0xy$ $(resp.\ ab1xy)$ evaluates 0 $(resp.\ 1)$ respectively;

$$R(0,0,0,0,0) = R[0] = 0$$

$$\boldsymbol{S_0} = \{0,1,2,3,8,9,10,11,16,17,18,19,24,25,26,27\}; \boldsymbol{S_1} = \{4,5,6,7,12,13,14,15,20,21,22,23,28,29,30,31\}$$

$$R[r] = 0 \text{ if } r \in S_0$$
$$R[r] = 1 \text{ if } r \in S_1$$

RMT	11111 (31)	11110 (30)	11101 (29)	11100 (28)	11011 (27)	....	01000 (8)	00111 (7)	00110 (6)	00101 (5)	00100 (4)	00011 (3)	00010 (2)	00001 (1)	00000 (0)
R	0	1	1	1	0	....	1	0	1	1	0	1	1	1	0

Non self replicating RMT

self replicating RMT

Terminologies

If every RMT of a RMT sequence is self replicating then the corresponding configuration is a **single length cycle.**

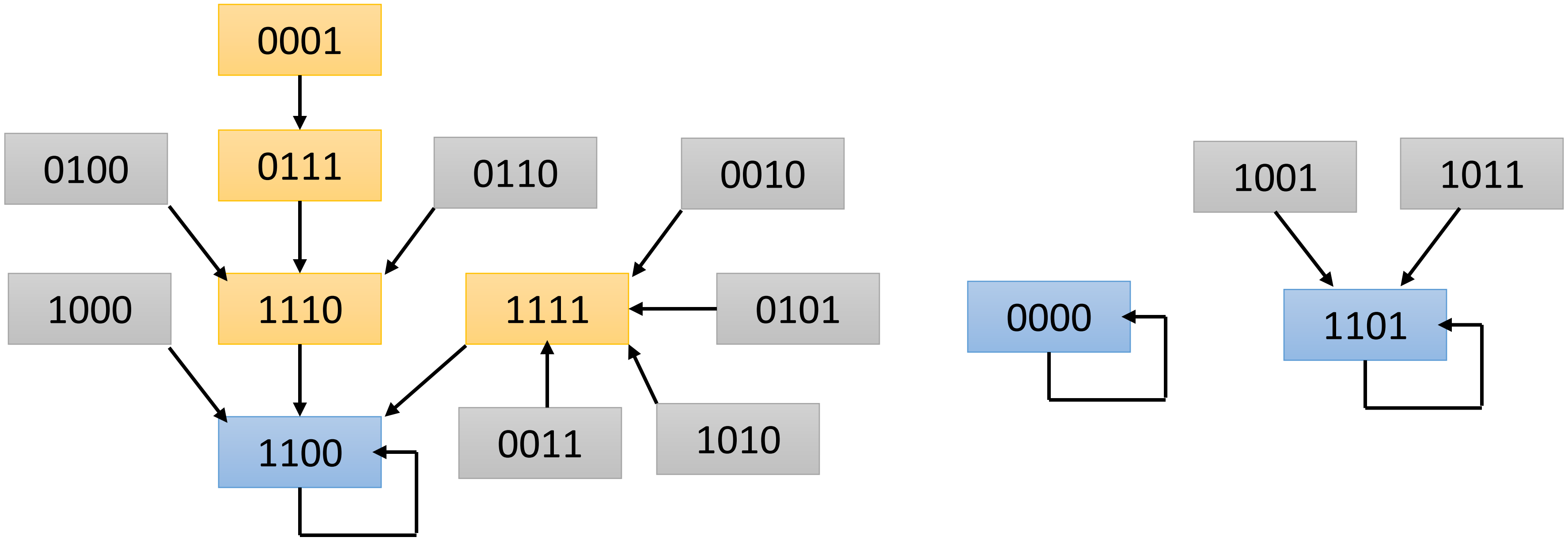
RMT	11111 (31)	11110 (30)	11101 (29)	11100 (28)	11011 (27)	…….	01000 (8)	00111 (7)	00110 (6)	00101 (5)	00100 (4)	00011 (3)	00010 (2)	00001 (1)	00000 (0)
R	0	1	1	1	0	……..	1	0	1	1	0	1	1	1	0

0 0 | 0 | 0 | 0 | 0 | 0 0

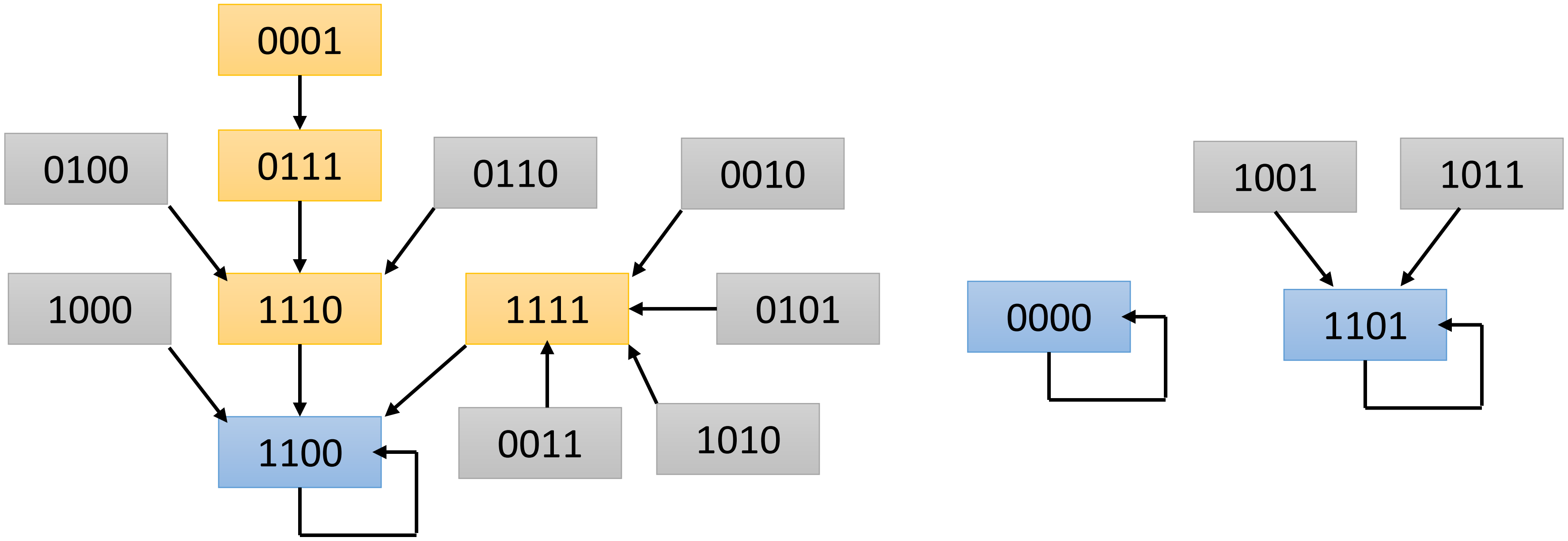
(0, 0, 0, 0) RMT Sequence

| 0 | 0 | 0 | 0 |

Cyclic Space of CA 1114110



Cyclic Space of CA 1114110



Cycle Structure = [3(1)]

Useful 5-neighborhood Rules for Pattern Classification

Rule space
$2^{32} = 4294967296$

Single length cycles only

Stable Cyclic behaviour

Moderate number of self replicating RMTs

Useful 5-neighborhood Rules for Pattern Classification

Rule space
$2^{32} = 4294967296$

Single length cycles only

Stable Cyclic behaviour

50% self replicating RMTs

Moderate number of self replicating RMTs

CS = [1(1)]

Rule 0: 00000000000000000000000000000000

Useful 5-neighborhood Rules for Pattern Classification

Strictly single length cycle CA as classifier: at least two cycles.

Useful 5-neighborhood Rules for Pattern Classification

Adding Impurities:

Number of strictly single length cycle rules with *majority 0s* where the count of cycles ≥ 2

# of 1s in rule	# of 0s in rule	Total possible rules	# of strictly single length cycle rules
1	31	32	1
2	30	496	32
3	29	4960	482
4	28	35960	4640
5	27	201376	30189

Useful 5-neighborhood Rules for Pattern Classification

Adding Impurities:

Number of strictly single length cycle rules with *majority 1s* where the count of cycles ≥ 2

# of 0s in rule	# of 1s in rule	Total possible rules	# of strictly single length cycle rules
1	31	32	6
2	30	496	109
3	29	4960	1123
4	28	35960	7282

Useful 5-neighborhood Rules for Pattern Classification

Balanced Rules: ${}^{32}C_{16}$

Balanced + Unbalanced (up to 25% 0s and 12.5% 1s)

2130 out 0f 601080390

# of 1s in rule	# of 0s in rule	Total possible rules	# of strictly single length cycle rules
1	31	32	1
2	30	496	32
3	29	4960	482
4	28	35960	4640
5	27	201376	30189

# of 0s in rule	# of 1s in rule	Total possible rules	# of strictly single length cycle rules
1	31	32	6
2	30	496	109
3	29	4960	1123
4	28	35960	7282

Useful 5-neighborhood Rules for Pattern Classification

Stable Cyclic Behaviour: Cycle Structure: Either constant of gradually increasing (linear with CA sizes)

A very few balanced and unbalanced 5-neighborhood rules with stable cyclic behavior

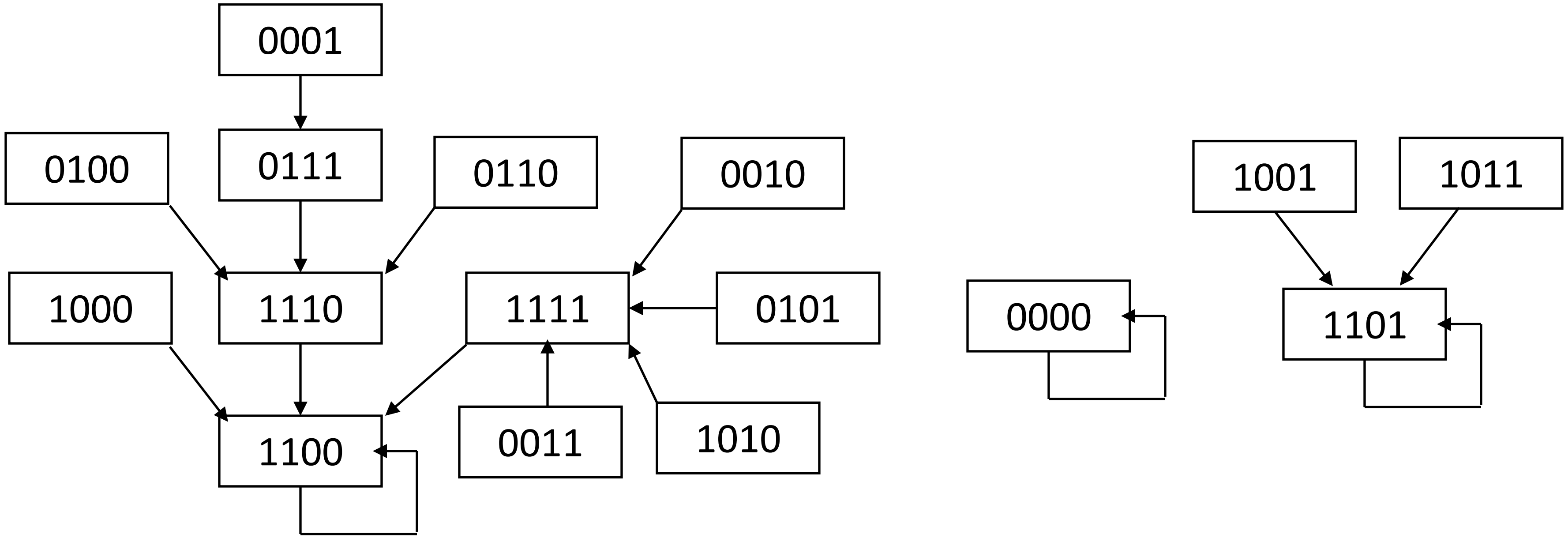
Rule	# of 0s in rule	# of 1s in rule	n = 4	n = 5	n = 6	n = 7	n = 8	n = 9	n = 10
84	29	3	3	3	3	4	4	4	5
1244158	16	16	3	3	3	4	4	4	4
118	26	6	3	3	3	4	4	4	5
4277534718	4	28	3	3	3	3	3	3	3

Useful 5-neighborhood Rules for Pattern Classification

The 150 rules used during training and testing

16	148	1048864	1046	135240	662	6472	268456208	1714	39496	1114103	1228798	1572078	4278190071	4294901726
1024	304	2097172	2328	2148532512	920	3146528	4227124	2714	4603922	1114110	1243127	1702143	4294965243	3755997175
2048	532	30	4170	263188	1110	2637844	1180692	28746	13648144	1177087	1244158	2027258	3607101439	3758028791
8192	784	60	4172	1048928	1236	9699364	126	28952	196606	1177591	1245171	4278190079	3758095351	3758096376
131072	4168	86	4400	2099472	1588	10617104	190	32862	450558	1178615	1277942	4294967294	4152358911	4194287614
20	16404	90	4880	62	2198	25170496	219	33394	720638	1179133	1307639	4294965247	4194303991	4194299895
272	131092	180	8724	118	3220	34603304	574	37426	712702	1179635	1308662	4211081215	4219469567	4225495039
28	262164	212	12360	182	4298	33817620	754	43288	942078	1212407	1309438	4227856383	4210950143	4261409790
52	262416	408	16916	244	4552	67129416	1118	42036	974590	1212414	1310454	4227857407	4227792879	4277534718
84	1048612	596	24848	572	5704	134225948	1238	41116	974844	1228791	1482750	4278124543	4276092926	4290772918

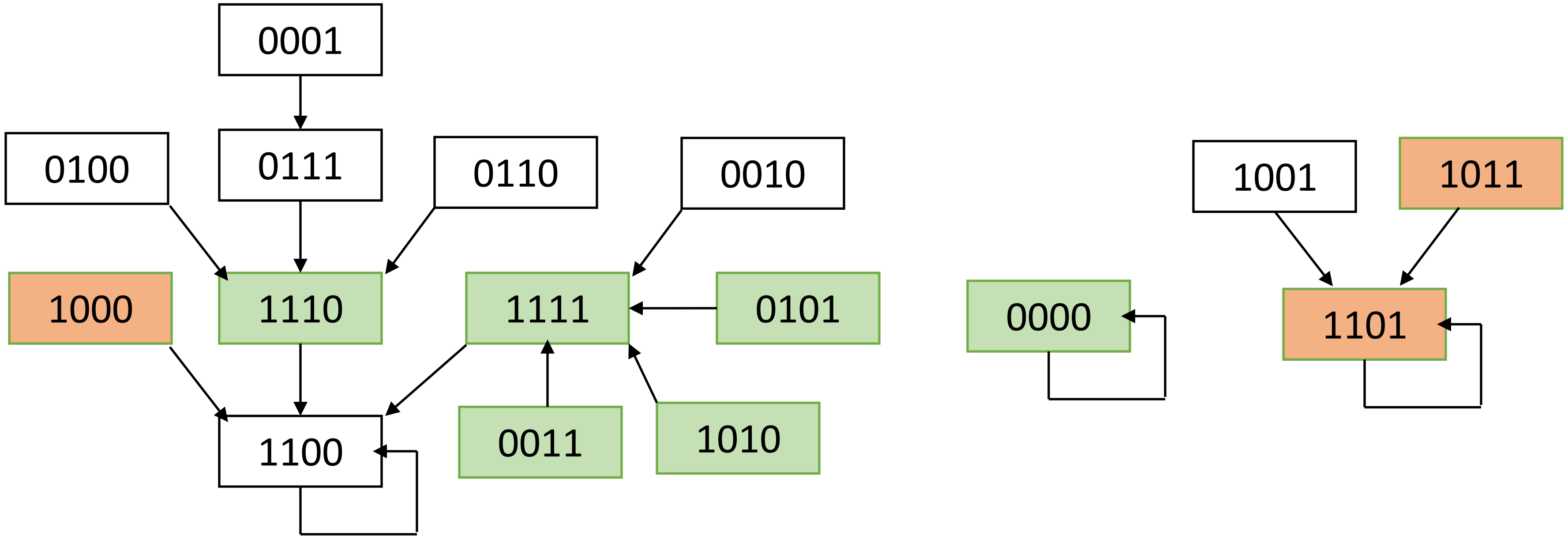
CA based Pattern Classification



Class A: 0011, 1110, 0000, 1111, 0101 and 1010

Class B: 1000, 1011 and 1101

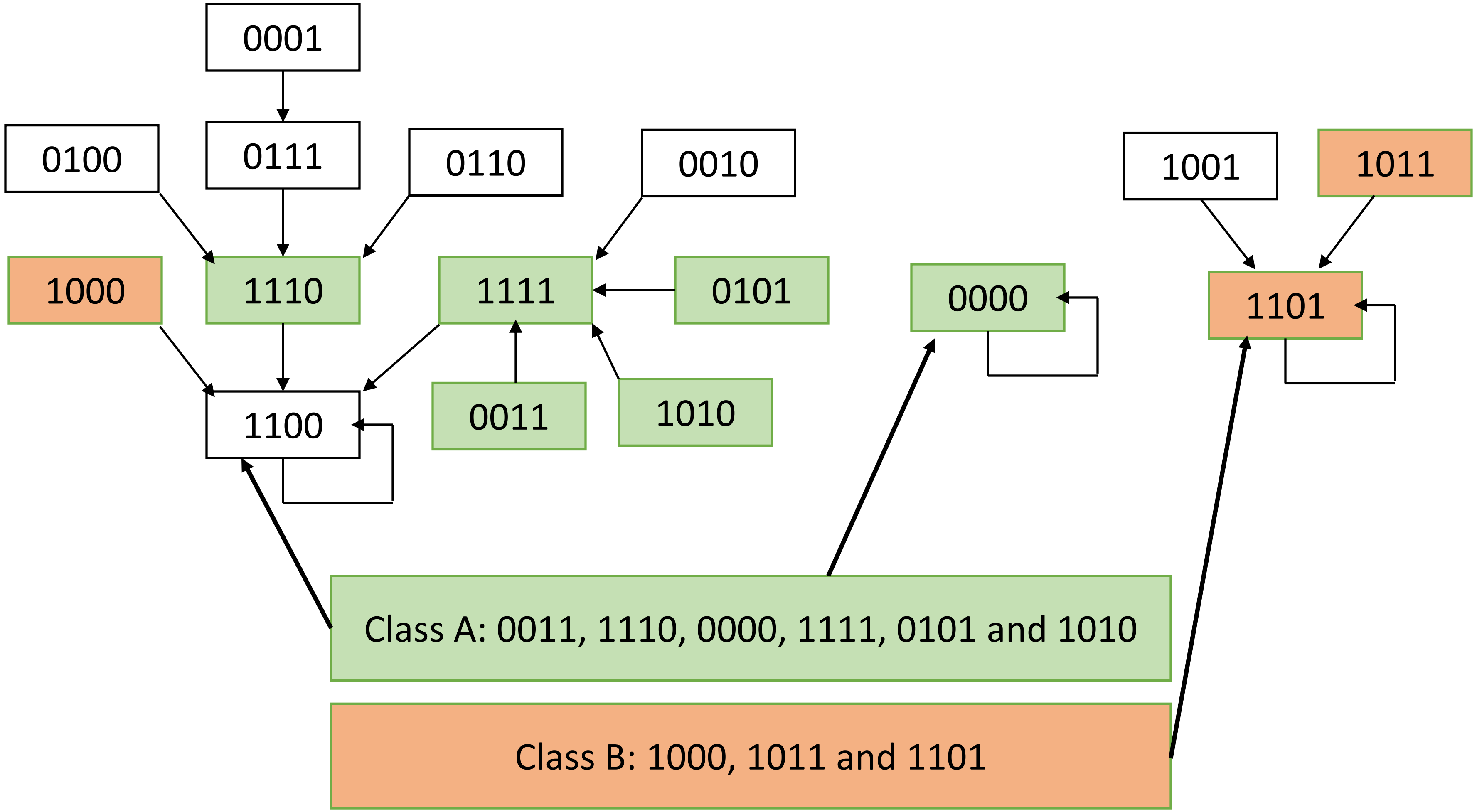
CA based Pattern Classification



Class A: 0011, 1110, 0000, 1111, 0101 and 1010

Class B: 1000, 1011 and 1101

CA based Pattern Classification



CA based Pattern Classification - Encoding

Encoding real-world data for use with CA requires careful transformation:

- **Frequency Encoding for Continuous Attributes:**

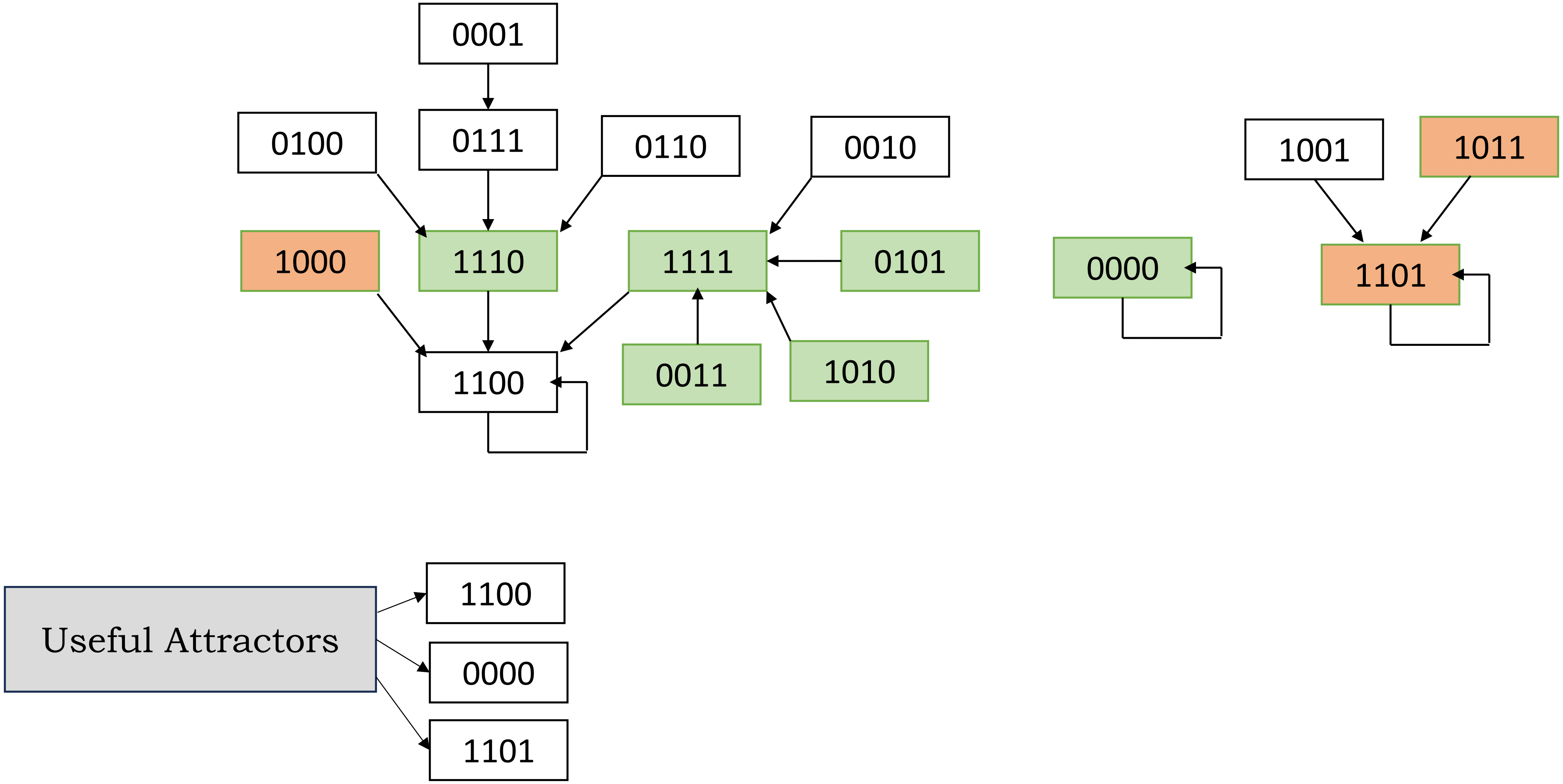
  – The range of each continuous attribute is divided into bins (e.g., 3 bins).
  – Each bin is represented by a unique binary string

- **One-Hot Encoding for Categorical Attributes:**

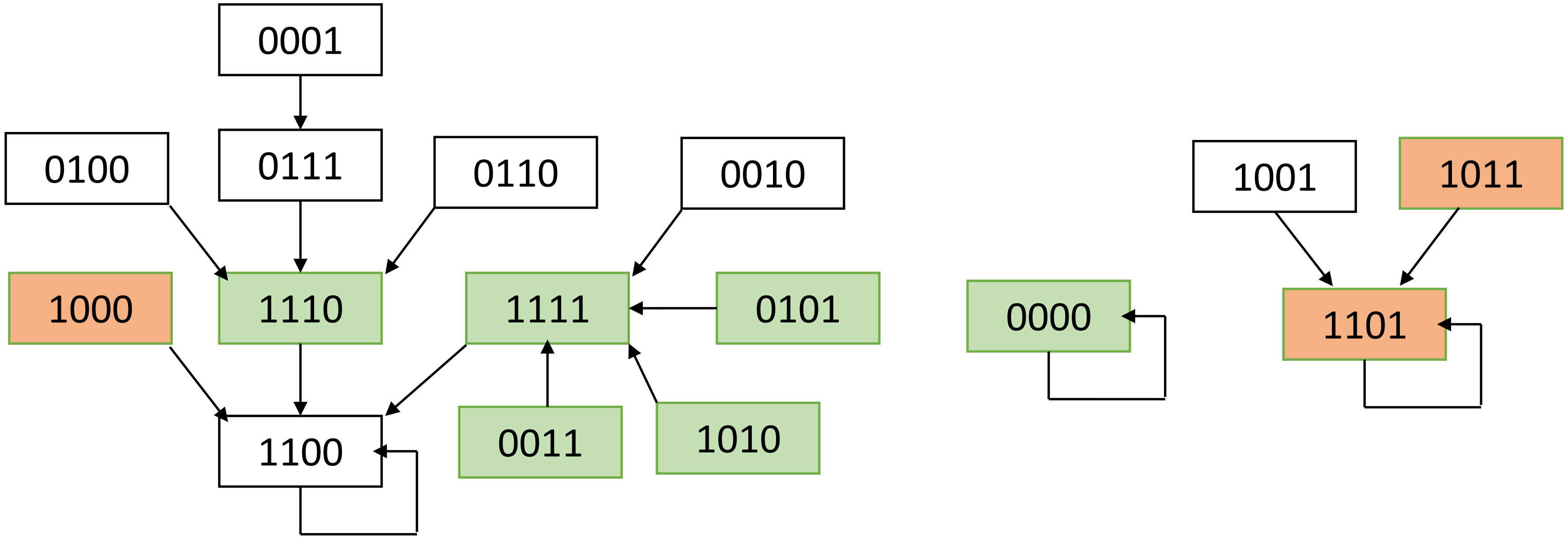
  – Each unique category is assigned a binary vector where only one bit is "1" (e.g., for colors: red → 001, green → 010, blue → 100).

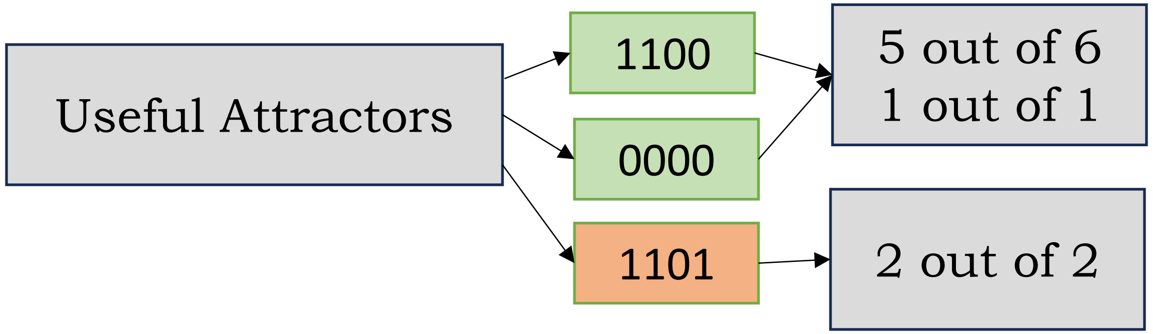
The final encoded binary string's length determines the CA size (n). For instance, a dataset with 7 continuous attributes and a bin size of 3 results in a CA of size 21. This systematic encoding bridges the gap between raw data and the CA framework.

CA based Pattern Classification - Training



CA based Pattern Classification - Training

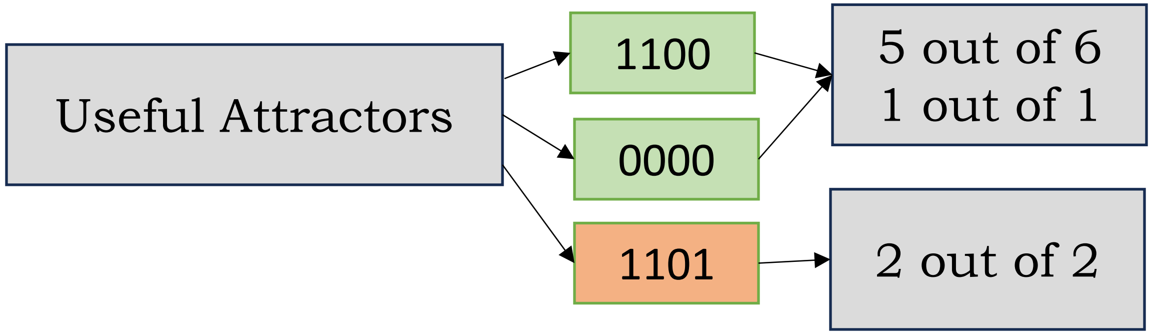




CA based Pattern Classification - Training

TRAINING EFFICIENCY $= \left(\left(\sum_{i=1}^{k} Ai\right)/n\right) \times 100$

$n =$ Total number of patterns in training set
$k =$ number of **useful** attractors
$Ai =$ The maximum (over classes) number of instances of a class recognized by attractor i

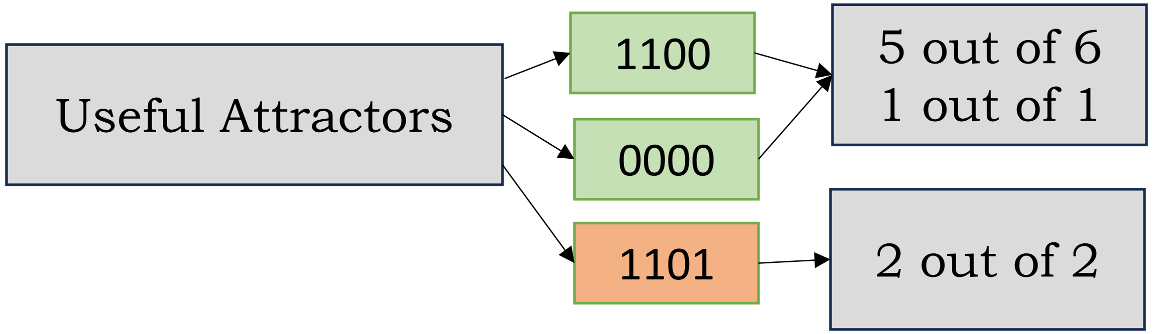


CA based Pattern Classification - Training

TRAINING EFFICIENCY $= \left(\left(\sum_{i=1}^{k} Ai\right)/n\right) \times 100$

$n =$ Total number of patterns in training set
$k =$ number of **useful** attractors
$Ai =$ The maximum (over classes) number of instances of a class recognized by attractor i

((5+1+2)/(6+1+2))/100=88.88%



CA based Pattern Classification – Testing & Performance Metrics

ACCURACY = $(TP + TN) \div (TP + FP + TN + FN)$
PRECISION = $TP \div (TP + FP)$
RECALL = $TP \div (TP + FN)$
F-MEASURE = $(2 \times PRECISION \times RECALL) \div (PRECISION + RECALL)$

TP=true positive
FP=false positive
TN=true negative
FN=false negative

CA based Pattern Classification

Rules Categorization

CA rules	No. of states containing ‘1’	No. of states containing ‘0’
84	3	29
1244158	16	16
304	3	29
86	4	28
1118	6	26
1114110	16	16
754	6	26
4277534718	28	4
4227792879	29	3

CA based Pattern Classification

Dataset Description

Name of the dataset	# of attributes	# of objects	Maximum bin size	# of Useful Configuration	CA size
Raisin	7	900	3	117	21
Rice	7	3810	3	252	21
Breast Cancer	9	116	2	71	18

CA based Pattern Classification

Raisin Dataset: 5-neighborhood CA based classifier results

CA rules	Efficiency	Accuracy	Precision1	Recall1	F-score1
84	70.73%	77.14%	75%	100%	85.71%
1244258	69.51%	80%	76.92%	95.24%	85.11%
304	74.39%	80%	85%	80%	82.93%
86	71.95%	82.86%	82.61%	90.48%	86.36%
1118	73.17%	85.71%	86.36%	90.48%	88.37%
1114110	**80.49%**	**85.71%**	**90.48%**	**90.48%**	**90.48%**
4227792879	69.51%	68.57%	67.86%	90.48%	77.55%

Comparison: CA 1114110 vs benchmark algorithms for Raisin dataset

Metrics	KNN	LDA	SVC	XGB	RIDGE	Rule 1114110
Accuracy	83.2%	85.7%	85.44%	86.88%	85.77%	85.71%
F-score1	83.6%	86.39%	83.43%	86.22%	87.05%	90.48%

CA based Pattern Classification

RICE Dataset: 5-neighborhood CA based classifier results

Rules	Efficiency	Accuracy	Precision1	Recall1	F-score1
84	68.18%	61.84%	67.5%	62.79%	65.06%
1244158	**84.09%**	**80.26%**	**81.82%**	**83.72%**	**82.76%**
86	67.61%	69.74%	79.41%	62.79%	70.13%
1118	76.70%	65.79%	75.76%	58.14%	65.79%
1114110	82.95%	75%	73.08%	88.37%	80%
4277534718	56.82%	60.53%	59.428%	95.35%	73.21%

Comparison: CA 1244158 vs benchmark algorithms for Rice dataset

Metrics	KNN	SVC	RF	NB	Rule 1244158
Accuracy	83.2%	72.6%	91.8%	90.6%	80.26%
Precision	88.3%	73.3%	91.8%	90.6%	81.82%
Recall	88.3%	72.6%	91.8%	90.6%	83.72%
F-score	88.3%	71.4%	91.8%	90.6%	82.76%

CA based Pattern Classification

BREAST CANCER Dataset: 5-neighborhood CA based classifier results

Rules	Efficiency	Accuracy	C2Precision	C2Recall	C2Fmeasure
84	**81.63%**	**63.64%**	**58.82%**	**90.91%**	**71.43%**
1244158	85.71%	50%	50%	100%	66.67%
86	75.51%	50%	50%	90.91%	64.52%
754	79.59%	63.64%	83.33%	45.45%	58.82%
1114110	85.71%	59.09%	55%	100%	70.97%
4277534718	75.5%	54.55%	52.63%	90.91%	66.7%

Table 11: Comparison: CA 84 vs benchmark algorithms for Breast Cancer dataset

Metrics	KNN	NB	DT	RF	Rule 84
Precision	48%	66%	70%	71%	58.82%
Recall	48%	60%	70%	71%	90.91%
F-score	48%	59%	70%	70%	71.436%

Comparative Analysis & Discussion

Analyzing results across all three datasets, we observe that:

• The selected CA rules maintain a stable and moderate number of attractors, regardless of the CA size.

• There is a consistent relationship between rule balance (or impurity) and classification performance.

• Our CA classifier not only matches but in some cases outperforms conventional methods on key metrics such as accuracy and F-score.

This comparative discussion highlights the strengths of the 5-neighborhood CA approach in capturing data patterns through stable attractor dynamics and reinforces its potential for broader applications.

Conclusion & Future Work

In conclusion, our work demonstrates that 2-state 5-neighborhood Cellular Automata—with strictly single length cycles—can serve as effective pattern classifiers.
Key contributions include:

- A novel rule selection strategy that identifies rules with stable cyclic behavior and a moderate number of attractors.
- A robust encoding methodology to convert real-world datasets into CA configurations.
- Competitive performance against benchmark classifiers across multiple datasets.

Looking ahead, further research may explore:

- Extending the approach to multi-state CA systems.
- Integrating hybrid models to further boost classification performance.
- Applying the methodology to a broader range of datasets and real-world problems.

THANK YOU