Applications of Cellular Automata

Alex Josué Flórez Farfán

Tópicos em Visualização e Imagens

Profa. Maria Cristina Ferreira de Oliveira

ICMC USP
SÃO CARLOS

Contents

- Cellular Automata
- Applications
  - Authorship attribution based on Life-Like Network Automata
  - Exploring Spatio-temporal Dynamics of Cellular Automata for Pattern Recognition in Networks
  - Cellular automata rule characterization and classification using texture descriptors

Authorship attribution

PLOS | ONE

RESEARCH ARTICLE

Authorship attribution based on Life-Like Network Automata

Jeaneth Machicao[1], Edilson A. Corrêa, Jr.[2], Gisele H. B. Miranda[2], Diego R. Amancio[2], Odemir M. Bruno[1] *

1 Scientific Computing Group, São Carlos Institute of Physics, University of São Paulo, PO Box 369, 13560-970, São Carlos, São Paulo, Brazil, **2** Institute of Mathematics and Computer Science, University of São Paulo, Avenida Trabalhador são-carlense, 400, 13566-590, São Carlos, São Paulo, Brazil

Authorship attribution task

- Assign or discover authors to anonymous or disputed documents

- Traditional techniques use text analytics and natural language processing to characterize authors' writing styles
- Not simple: syntactical and semantical features

Text as network

- Network science and Cellular automata
- Discrete dynamical system: network automata
- Work based on Life-like Networkk Automata (LLNA)

- Text modeled as networks
- In which nodes are words and edges are linked among them

Authorship attribution based on Life-Like Network Automata

- Modeling: entities of a problem as vertices and interaction between these entities as edges
- Each **node** represents a word
- **Edges** created if two words are adjacent in the text
- Word adjacency network as matrix A

$$A_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are connected, and} \\ 0, & \text{otherwise.} \end{cases}$$

Related concepts

Example of tokenization.

A swimmer likes swimming, thus he swims.

↓

a	swimmer	likes	swimming	thus	he	swims

Example of stop word removal.

A swimmer likes swimming, thus he swims.

↓

swimmer	likes	swimming	,	swims	.

Example of Porter stemming.

A swimmer likes swimming, thus he swims.

↓

a	swimmer	like	swim	,	thu	he	swim	.

Example of lemmatization.

A swimmer likes swimming, thus he swims.

↓

A	swimmer	like	swimming	,	thus	he	swim	.

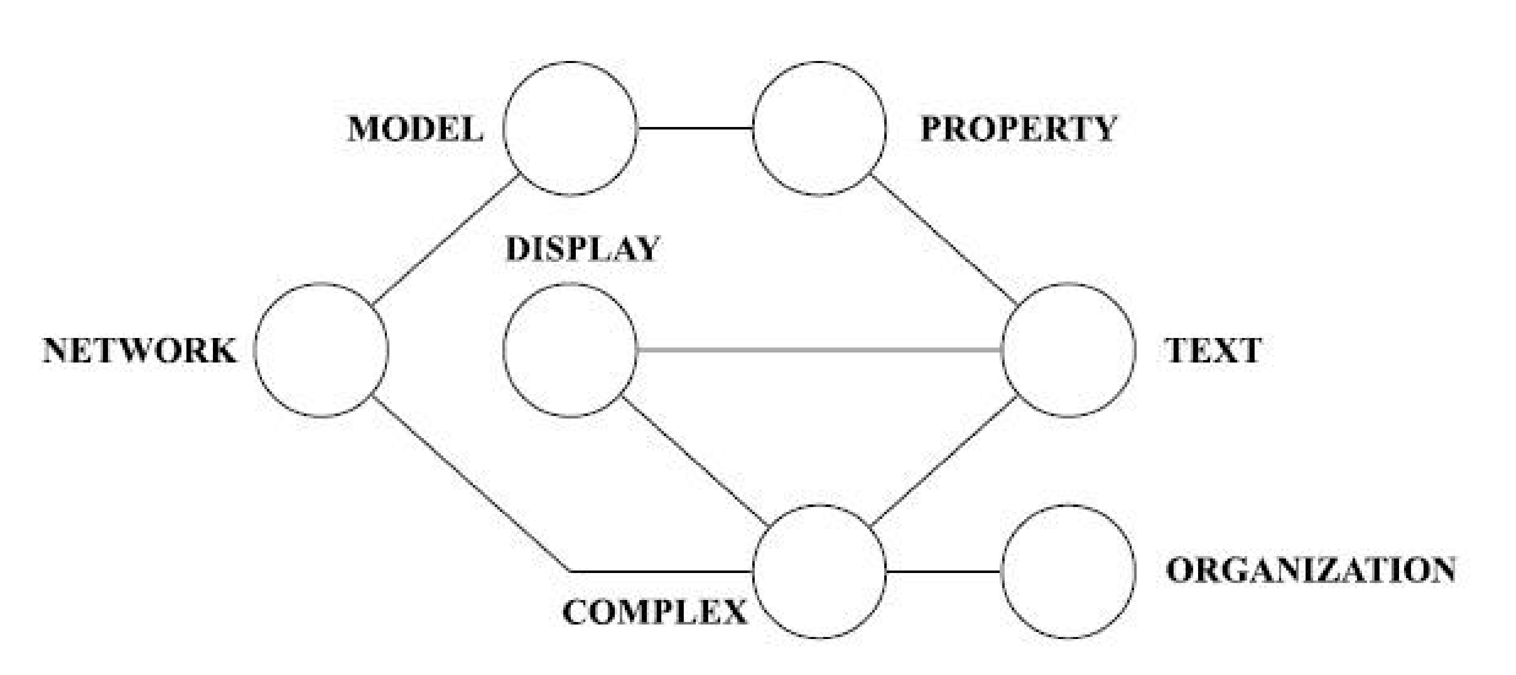
Network construction

- Tokenization: split document in meaningful words.
- Removal of stopwords.
- Punctuation marks are disregarded, convey no semantic meaning.

Modeling text as network

Complex networks model several properties of texts.

A complex text displays a complex organization.



Cellular Automata

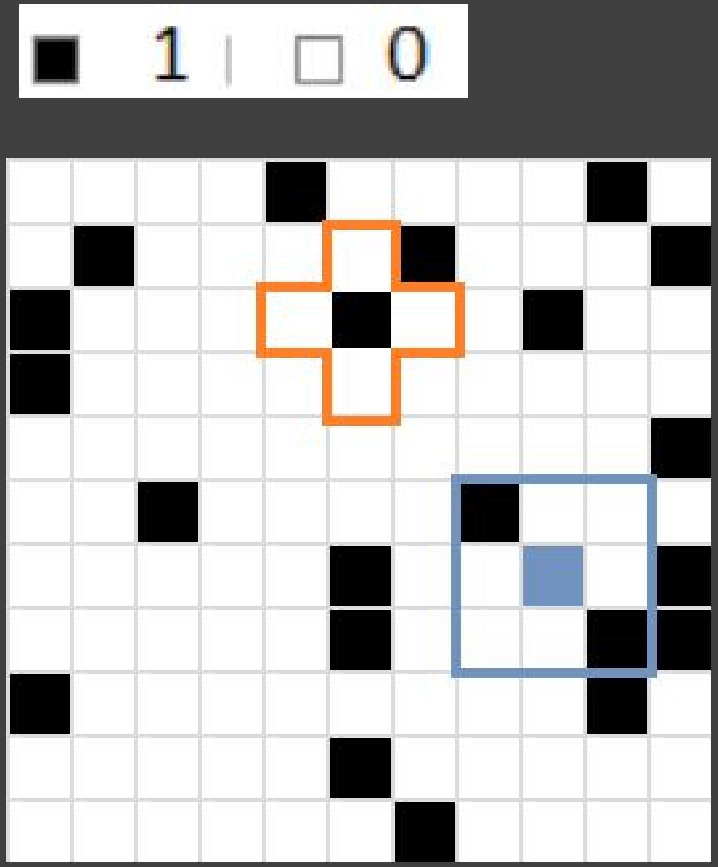
T: tesselletion

S: states

s_0: initial configuration

N: neighborhood

ϕ: transition rules



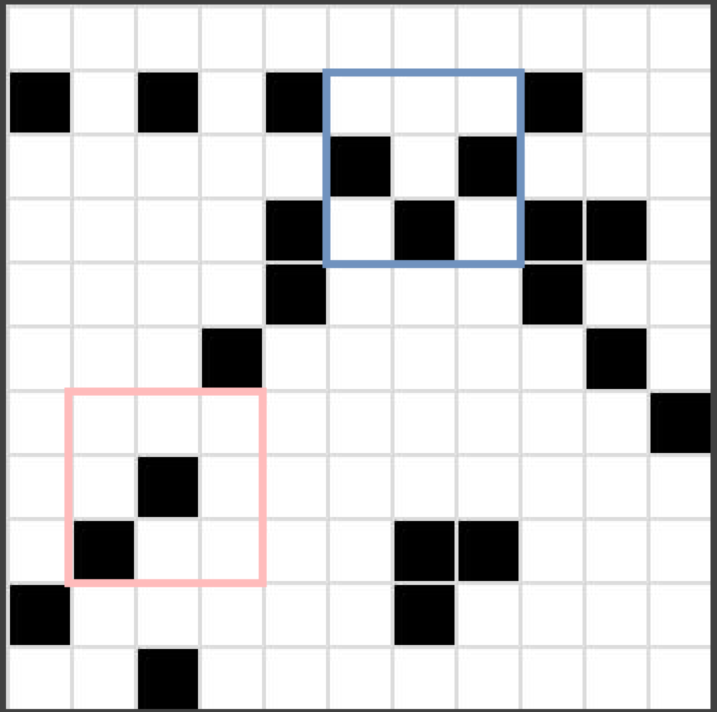
Von Neumann: 4 neighbors

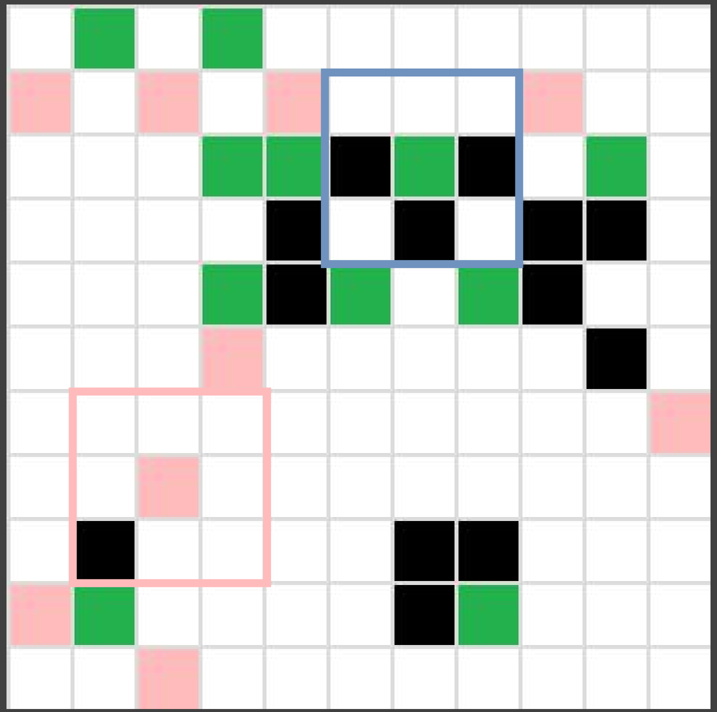
Moore: 8 neighbors

Cellular Automata: Transition rules

Conway's Game of Life B3 / S23

Life like rule: Birth: 3 / Survival: 2 or 3





Dead cells

Born cells

Life-like Network Automata (LLNA)

Modeling CAs over irregular tessellations.

Network topology as a tessellation for CAs

Evolving networks through the use of CAs

The space-temporal pattern formed from the evolution is used as feature vector

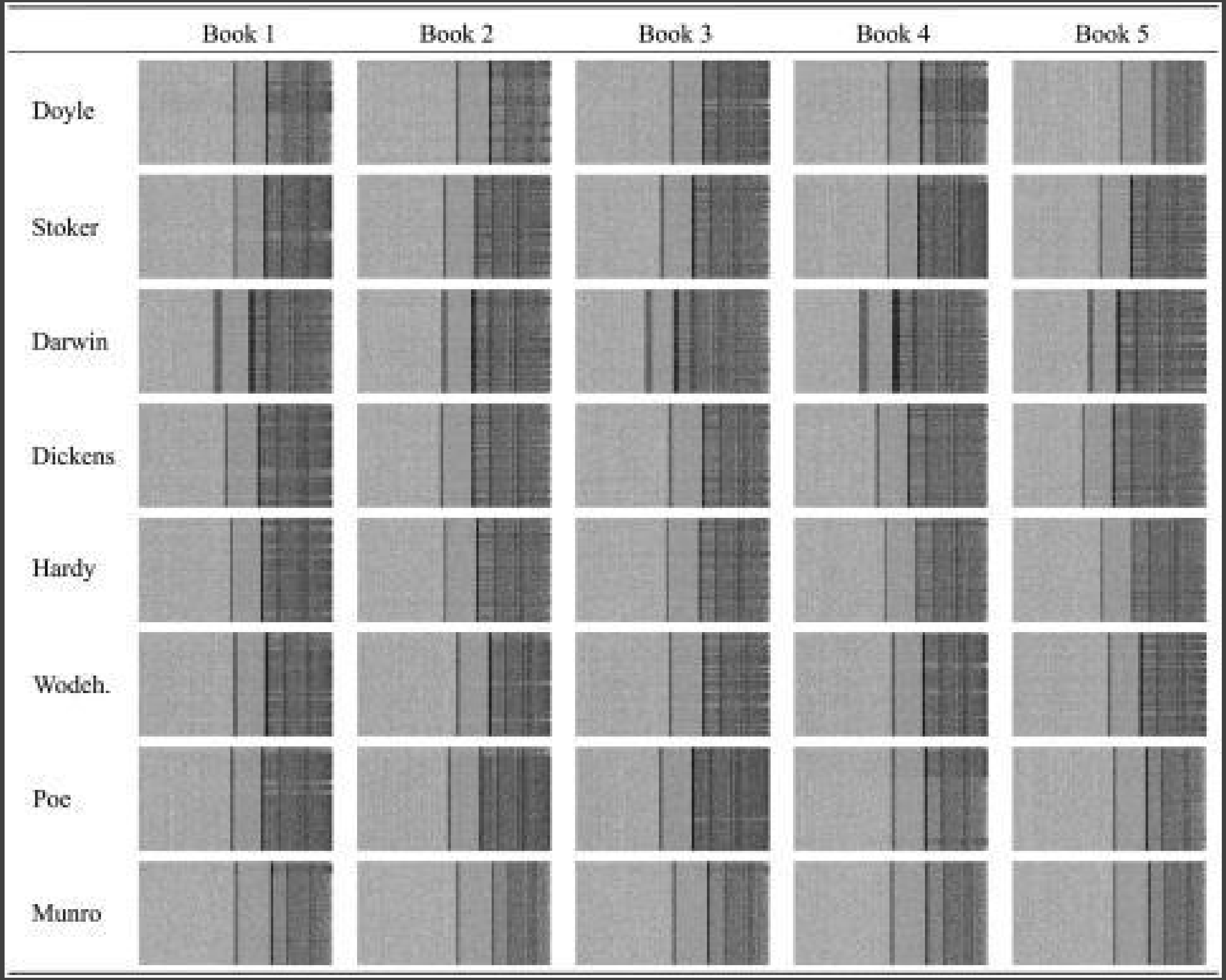
LLNA

T: tesselletion:

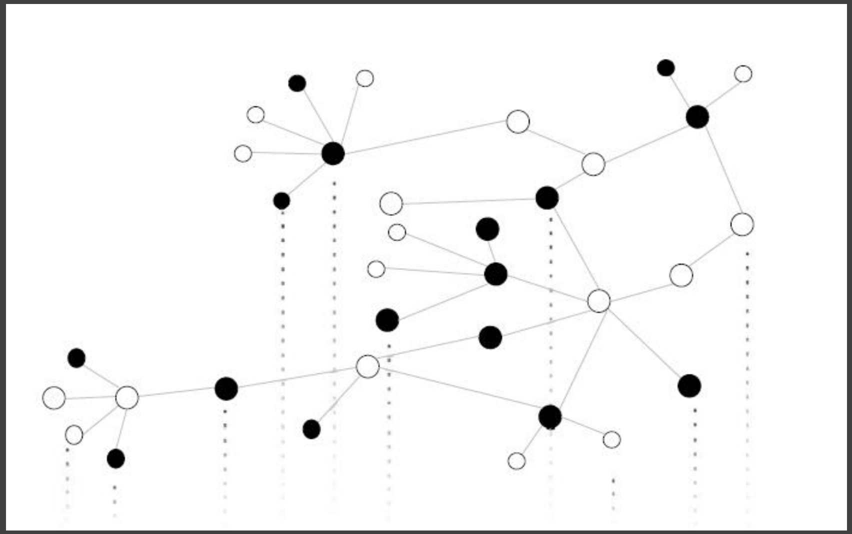
topology network

S: states: 0 or 1

Nodes ordered by degree



LLNA



N: neighborhood

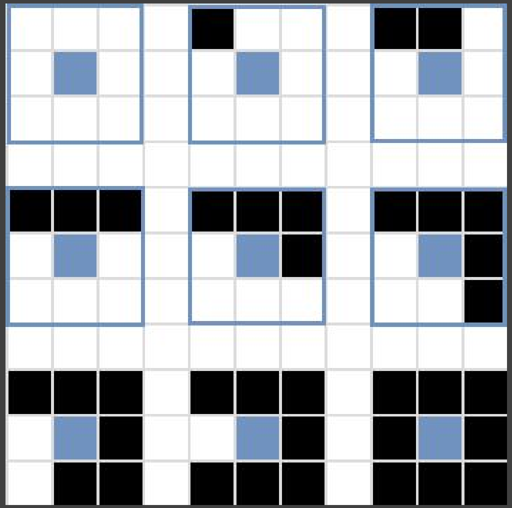
Mapping: density of living neighbors to Life-like rule

ϕ: transition rules

B_x / S_y x: {0, 1, 2, ..., 8}, y: {0, 1, 2, ..., 8}

$2^{(9+9)}$ = 262144 possible Life-like rules

Provides a vast space of optimal solutions for a specific problem.



Dataset Authorship attribution

- Books from Project Gutenberg
- Corpus: 100 books in English
- 20 authors
- Each author with 5 books

1. none-dataset: no lemmatization
2. partial-dataset: lemmatization only in nouns
3. full-dataset: lemmatization applied to all words

Experiments: Authorship attribution

- Learning:
  - rule-selection dataset: 12 authors
- Evaluation:
  - classification dataset: 8 authors

Experiments: Authorship attribution

Exploring space of possible rules to select those resulting in best accuracy

ϕ: transition rules

- B_x / S_y x: {0, 1, 2, ..., 8}, y: {0, 1, 2, ..., 8}
- $2^{(9+9)}$ = 262144 possible Life-like rules

Evaluating rules

Explore all possible rules Birth/Survival

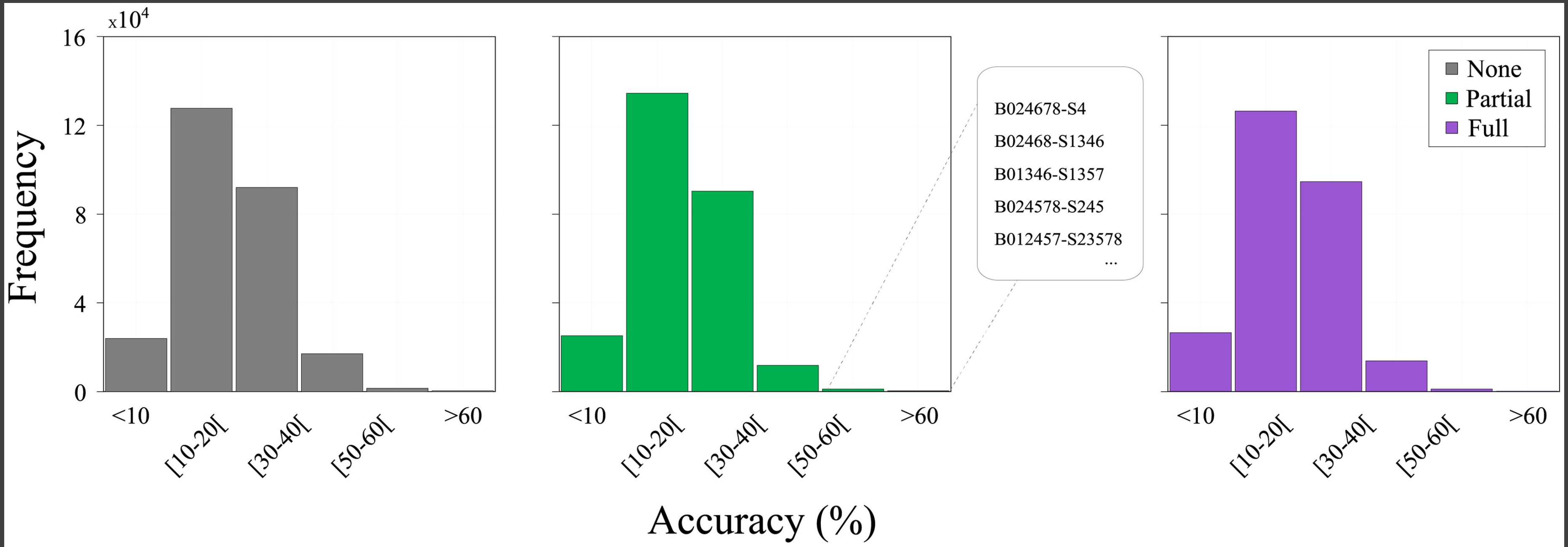
Classify dataset with each rule

Choose rules that produce best results

Birth	Survival
{0, 1, 2, 3, 4, 5, 6, 7, 8}	{0, 1, 2, 3, 4, 5, 6, 7, 8}
0	0
0	1
0	2
...	...
1	0
12	01
123	012
...	...
2	12
23	123
234	1234
...	...
012345678	012345678

Evaluating rules

- Histogram of the distribution of accuracy for all 262144 evaluated rules:
  - B024678 S4 B02468 S1346 B01346 S1357
  - B024578 S245 B012457 S23578



Classification Results

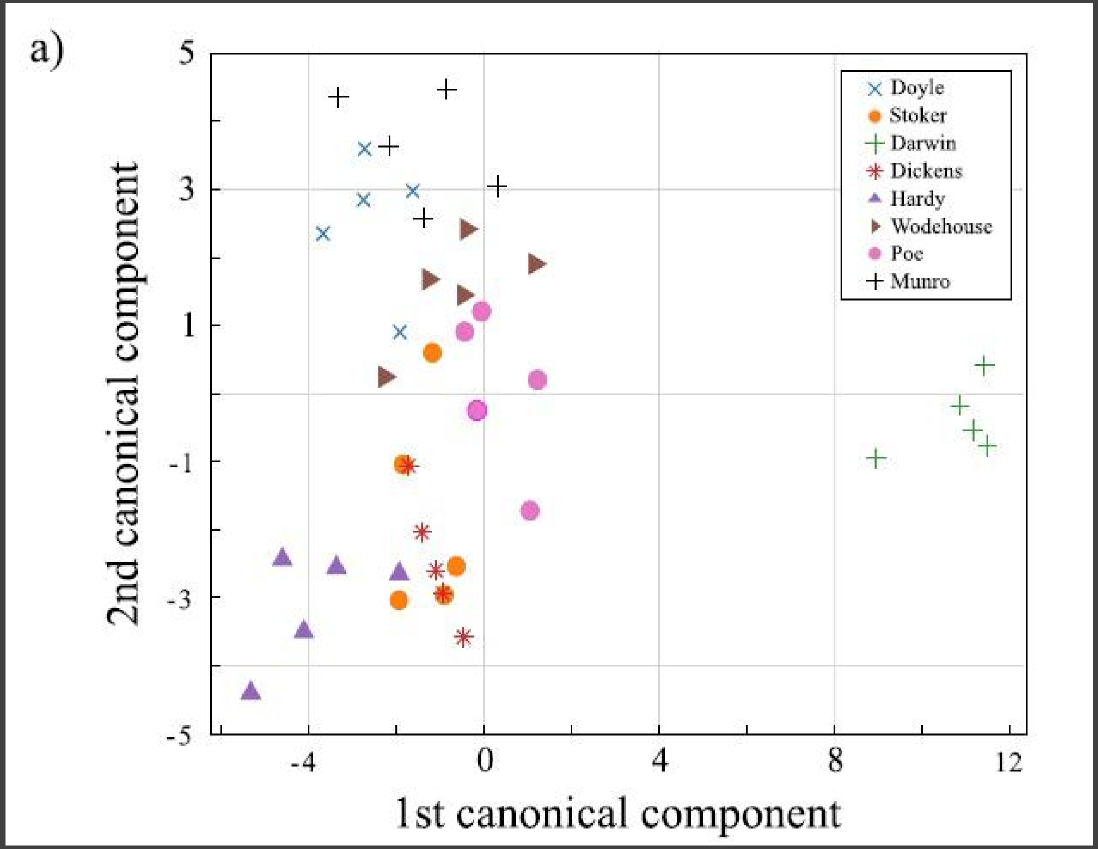
- u_s: Shannon entropy
- u_l: Lempel Ziv

Table 1. Accuracy rate (%) obtained using different measurements ($\vec{\mu}_S$ and $\vec{\mu}_L$) and their combinations as attributes ($[\langle\mu_S\rangle, \langle\mu_L\rangle]$) to classify 8 authors of the *classification-dataset*. To select the best rules, we used the kNN with $k = 1$ and 5-fold cross validation. The best result among all classifiers were also obtained with the kNN method.

Lemmatization	Rule	$\vec{\mu}_S$	$\vec{\mu}_L$	$[\langle\mu_S\rangle, \langle\mu_L\rangle]$
None	B124-S257	51.0(±16.11)	**68.5(±10.90)**	66.0(±12.25)
	B1245-S1245	37.5(±14.88)	62.0(±16.33)	57.5(±16.93)
	B245-S457	61.0(±14.12)	57.5(±17.31)	48.5(±15.44)
Partial	B2478-S25	43.5(±14.49)	**70.5(±13.44)**	50.0(±15.31)
	B026-S14	48.5(±15.02)	59.5(±15.00)	63.5(±14.84)
	B148-S6	62.5(±13.01)	36.0(±10.41)	53.5(±15.94)
Full	B3567-S03468	39.5(±11.79)	**68.0(±16.57)**	47.5(±20.41)
	B13568-S13	63.5(±15.28)	52.0(±14.74)	34.5(±14.11)
	B0134568-S0123568	36.5(±17.28)	55.0(±14.88)	61.5(±14.40)

Classification Results

- Canonical analysis on classification dataset
- Rule B2478 S25



Classification Results

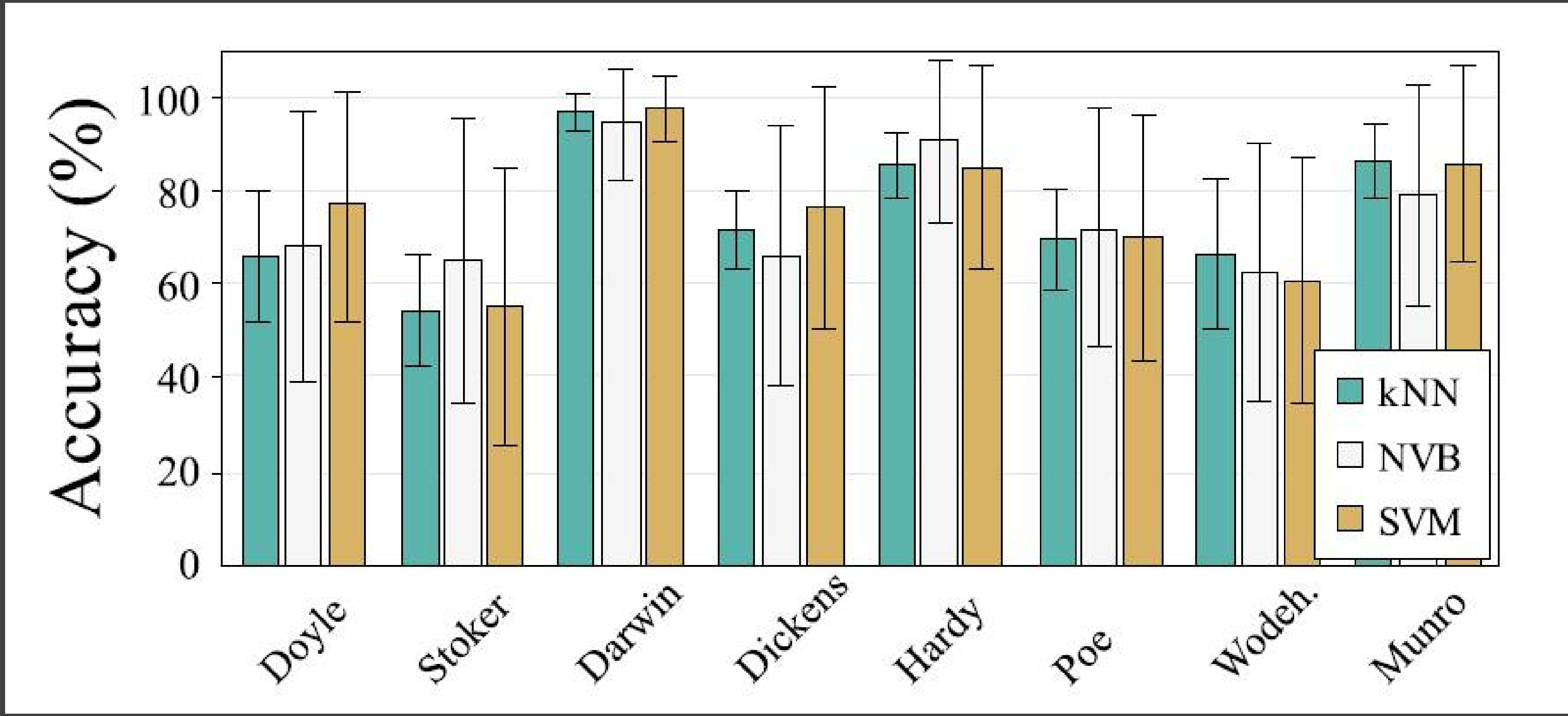
- Confusion matrix using kNN classifier

Predicted Class

Real Class	Doyle	Stoker	Darwin	Dickens	Hardy	Poe	Wodeh.	Munro
Doyle	3					1		1
Stoker		2		1	1	1		
Darwin			5					
Dickens		1		3				
Hardy		1			4			
Wodeh.	1					4		
Poe				1			3	
Munro								5

Classification Results

- Comparison of 3 classifiers



Cellular automata rule characterization and classification using texture descriptors

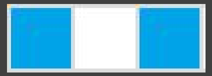
- Classify Elementary Cellular Automata (CA)

- Wolfram classified 256 elementary CA in 4 classes
- Two schemes:
  - Li & Packard 6 classes
  - 88 classes

Elementary Cellular Automata, ECAs

One dimensional

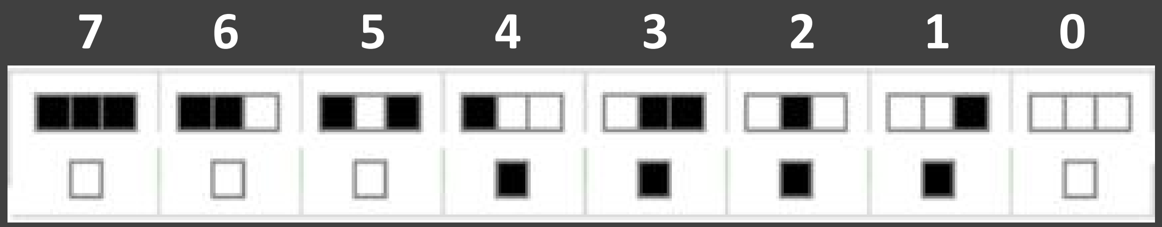
Two neighbors



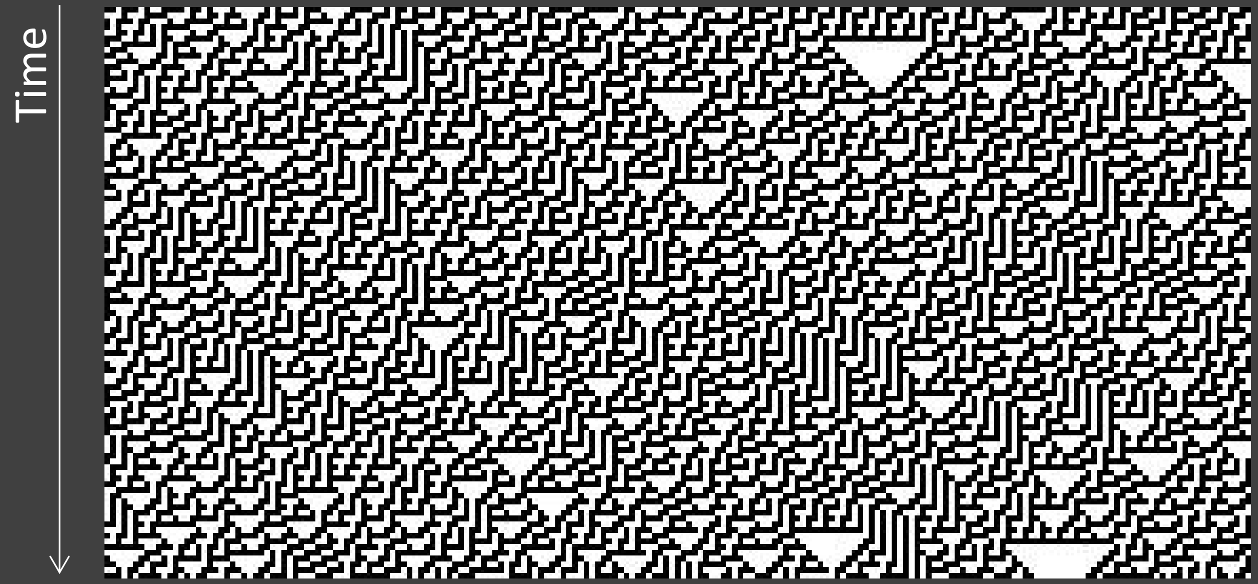
Eight configurations

Two output states: 0 or 1

$2^8 = 256$ possible ECAs



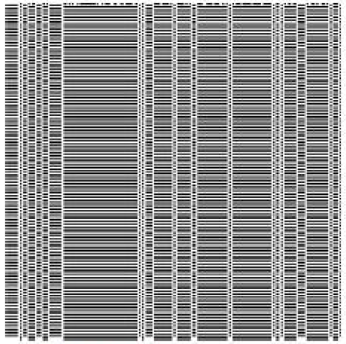
Rule 30



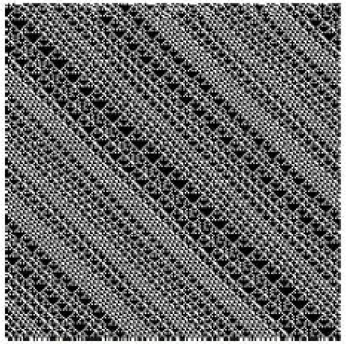
Li - Packard scheme 6 classes

(a) Rule 0 (null).

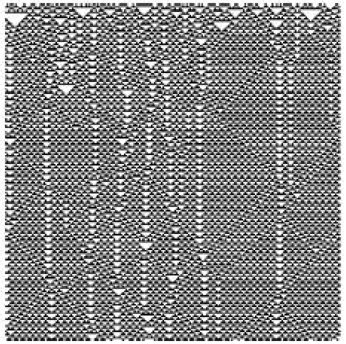
(b) Rule 247 (fixed-point).



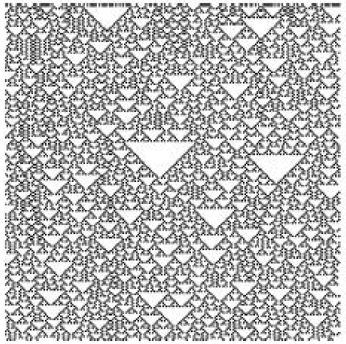
(c) Rule 127 (two-cycle).



(d) Rule 82 (periodic).



(e) Rule 147 (complex).

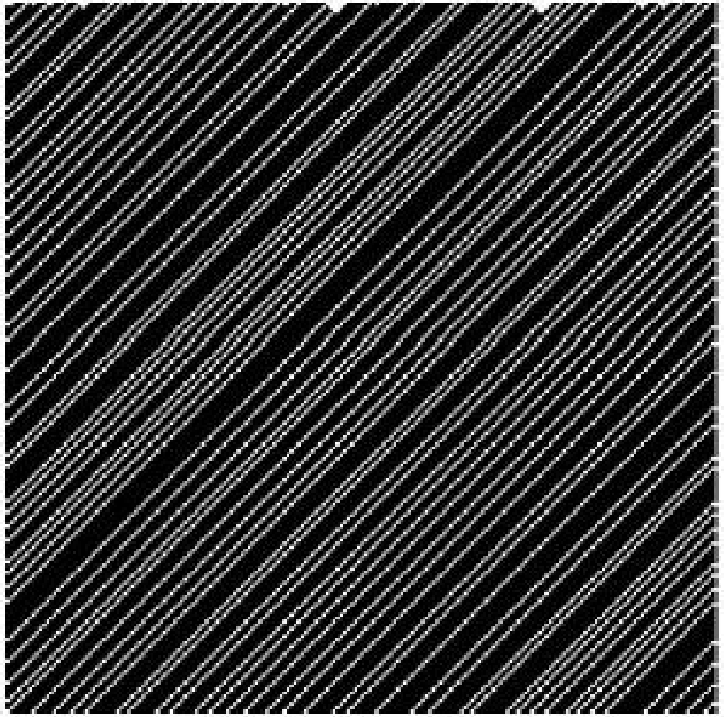


(f) Rule 183 (chaotic).

Cellular automata rule characterization and classification using texture descriptors

- Generate CA spatio-temporal pattern into a binary image
- Convert binary image pattern into gray-scale image
- Apply texture descriptor
- Use a classifier

Conversion spatio-temporal pattern (binary) to gray scale

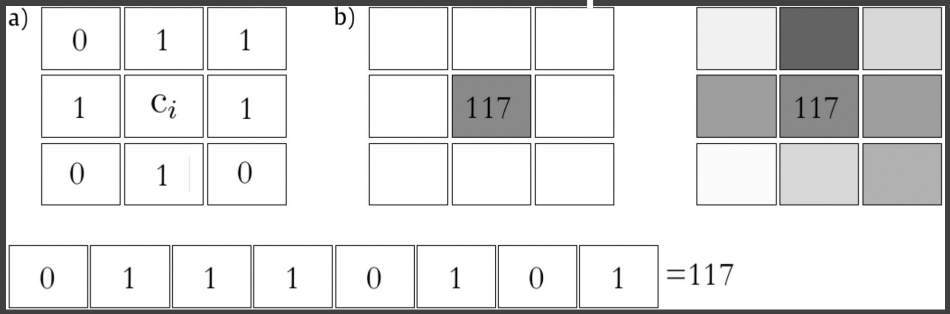


(a) Original ECA rule 130 and $p = 0.5$.

(b) Gray-scale conversion of (a).

Technique proposed

Spatio temporal pattern (binary) → Conversion to gray scale → Texture descriptor → Classifier

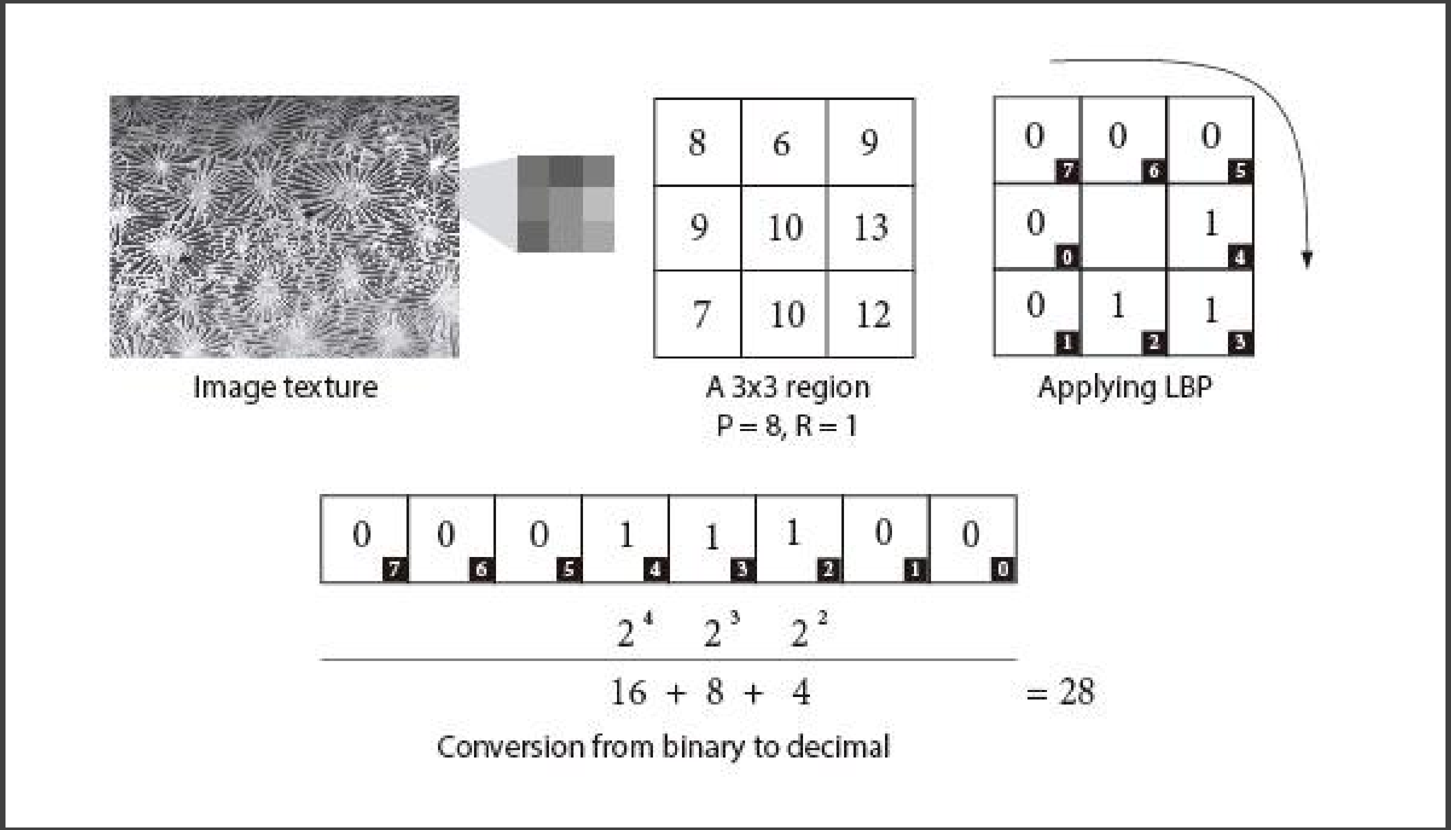


Texture descriptor:
- LBP
- Fourier
- Gabor
- GLCM

Classifier:
- SVM

Texture descriptor

- LBP: Local Binary Pattern



Classification results Li Packard 6 classes

Accuracy rate (%) of texture descriptors on binary and gray-scale CA images for the Li–Packard scheme of class experiment by varying the ECA initial configuration (ρ) and the number of pixels (N).

Config.		LBP		Fourier		Gabor		GLCM	
		Bin	Gray	Bin	Gray	Bin	Gray	Bin	Gray
	$\rho = 0.1$	91.57($\pm$ 0.83)	95.00($\pm$ 0.97)	84.57($\pm$ 1.54)	83.70($\pm$ 0.97)	84.00($\pm$ 0.94)	86.77($\pm$ 0.69)	87.33($\pm$ 0.68)	87.20($\pm$ 0.85)
	$\rho = 0.3$	98.13($\pm$ 0.39)	98.93($\pm$ 0.44)	85.37($\pm$ 1.21)	86.43($\pm$ 1.51)	87.97($\pm$ 1.05)	89.30($\pm$ 0.81)	88.67($\pm$ 1.12)	92.70($\pm$ 1.08)
$N = 200$	$\rho = 0.5$	98.07($\pm$ 0.62)	99.20($\pm$ 0.42)	85.80($\pm$ 1.06)	85.47($\pm$ 1.31)	87.73($\pm$ 1.11)	89.90($\pm$ 0.80)	88.63($\pm$ 1.43)	94.53($\pm$ 1.02)
	$\rho = 0.7$	97.13($\pm$ 0.57)	98.83($\pm$ 0.67)	81.50($\pm$ 1.57)	81.40($\pm$ 1.76)	86.43($\pm$ 1.74)	88.57($\pm$ 0.90)	89.13($\pm$ 1.06)	93.33($\pm$ 0.98)
	$\rho = 0.9$	90.27($\pm$ 1.53)	94.10($\pm$ 0.59)	75.40($\pm$ 1.36)	75.10($\pm$ 1.79)	77.63($\pm$ 1.35)	82.73($\pm$ 2.40)	83.03($\pm$ 1.97)	85.60($\pm$ 1.47)
	$\rho = 0.1$	91.67($\pm$ 0.99)	95.1($\pm$ 0.47)	83.7($\pm$ 1.35)	82.37($\pm$ 1.2)	84.3($\pm$ 0.85)	86.57($\pm$ 0.9)	87.5($\pm$ 1.1)	87.5($\pm$ 0.61)
	$\rho = 0.3$	98.33($\pm$ 0.22)	99.33($\pm$ 0.35)	84.33($\pm$ 1.01)	83.77($\pm$ 1.69)	87($\pm$ 0.7)	89.33($\pm$ 1.07)	88.57($\pm$ 1.14)	92.3($\pm$ 0.46)
$N = 250$	$\rho = 0.5$	**98.53**($\pm$ 0.39)	**99.57**($\pm$ 0.39)	82.57($\pm$ 1.92)	81.63($\pm$ 2.26)	88.4($\pm$ 1.08)	89.5($\pm$ 0.72)	88.5($\pm$ 1.03)	94($\pm$ 0.92)
	$\rho = 0.7$	96.77($\pm$ 0.32)	99.23($\pm$ 0.42)	80.4($\pm$ 1.68)	79.4($\pm$ 1.94)	87.47($\pm$ 1.58)	88.47($\pm$ 0.55)	87.73($\pm$ 1.3)	93.9($\pm$ 0.96)
	$\rho = 0.9$	90.5($\pm$ 0.93)	94.4($\pm$ 0.52)	74.1($\pm$ 1.66)	73.9($\pm$ 1.44)	77.37($\pm$ 0.92)	82.93($\pm$ 1.71)	84.9($\pm$ 2.23)	86.77($\pm$ 0.94)
	$\rho = 0.1$	91.77($\pm$ 0.97)	95.43($\pm$ 0.59)	84.13($\pm$ 1.44)	83.67($\pm$ 0.9)	84.67($\pm$ 0.83)	86.83($\pm$ 0.55)	87.83($\pm$ 0.81)	87.97($\pm$ 0.79)
	$\rho = 0.3$	98.33($\pm$ 0.52)	99.5($\pm$ 0.28)	85.73($\pm$ 1.65)	85.1($\pm$ 1.5)	87.17($\pm$ 0.97)	89.23($\pm$ 0.7)	88.8($\pm$ 1.55)	92.7($\pm$ 0.76)
$N = 300$	$\rho = 0.5$	98.4($\pm$ 0.52)	99.43($\pm$ 0.27)	85.93($\pm$ 1.19)	85.43($\pm$ 1.22)	87.6($\pm$ 0.75)	89.47($\pm$ 0.89)	88.97($\pm$ 0.69)	94.93($\pm$ 1.22)
	$\rho = 0.7$	96.77($\pm$ 0.45)	99.23($\pm$ 0.22)	82.27($\pm$ 1.68)	81.73($\pm$ 1.83)	86.23($\pm$ 1.98)	88($\pm$ 1.4)	88.77($\pm$ 0.97)	94.1($\pm$ 1.1)
	$\rho = 0.9$	91.17($\pm$ 1.21)	95($\pm$ 0.35)	75.1($\pm$ 0.92)	74.47($\pm$ 1.49)	77.5($\pm$ 1.34)	82.9($\pm$ 2.85)	85.9($\pm$ 1.49)	87.97($\pm$ 1)
Mean acc.		95.16 ($\pm$ 3.45)	**97.49** ($\pm$ 2.26)	82.06 ($\pm$4.07)	81.57 ($\pm$ 4.11)	84.76 ($\pm$ 3.99)	87.37 ($\pm$ 2.56)	87.62 ($\pm$ 1.74)	91.03 ($\pm$ 3.38)

Classification results 88 classes

Accuracy rate (%) of texture descriptors on binary and gray-scale CA images for the Li–Packard scheme of class experiment by varying the ECA initial configuration (ρ) and the number of pixels (N).

Config.		LBP		Fourier		Gabor		GLCM	
		Bin	Gray	Bin	Gray	Bin	Gray	Bin	Gray
$N = 200$	$\rho = 0.1$	91.57($\pm$ 0.83)	95.00($\pm$ 0.97)	84.57($\pm$ 1.54)	83.70($\pm$ 0.97)	84.00($\pm$ 0.94)	86.77($\pm$ 0.69)	87.33($\pm$ 0.68)	87.20($\pm$ 0.85)
	$\rho = 0.3$	98.13($\pm$ 0.39)	98.93($\pm$ 0.44)	85.37($\pm$ 1.21)	86.43($\pm$ 1.51)	87.97($\pm$ 1.05)	89.30($\pm$ 0.81)	88.67($\pm$ 1.12)	92.70($\pm$ 1.08)
	$\rho = 0.5$	98.07($\pm$ 0.62)	99.20($\pm$ 0.42)	85.80($\pm$ 1.06)	85.47($\pm$ 1.31)	87.73($\pm$ 1.11)	89.90($\pm$ 0.80)	88.63($\pm$ 1.43)	94.53($\pm$ 1.02)
	$\rho = 0.7$	97.13($\pm$ 0.57)	98.83($\pm$ 0.67)	81.50($\pm$ 1.57)	81.40($\pm$ 1.76)	86.43($\pm$ 1.74)	88.57($\pm$ 0.90)	89.13($\pm$ 1.06)	93.33($\pm$ 0.98)
	$\rho = 0.9$	90.27($\pm$ 1.53)	94.10($\pm$ 0.59)	75.40($\pm$ 1.36)	75.10($\pm$ 1.79)	77.63($\pm$ 1.35)	82.73($\pm$ 2.40)	83.03($\pm$ 1.97)	85.60($\pm$ 1.47)
$N = 250$	$\rho = 0.1$	91.67($\pm$ 0.99)	95.1($\pm$ 0.47)	83.7($\pm$ 1.35)	82.37($\pm$ 1.2)	84.3($\pm$ 0.85)	86.57($\pm$ 0.9)	87.5($\pm$ 1.1)	87.5($\pm$ 0.61)
	$\rho = 0.3$	98.33($\pm$ 0.22)	99.33($\pm$ 0.35)	84.33($\pm$ 1.01)	83.77($\pm$ 1.69)	87($\pm$ 0.7)	89.33($\pm$ 1.07)	88.57($\pm$ 1.14)	92.3($\pm$ 0.46)
	$\rho = 0.5$	**98.53**($\pm$ 0.39)	**99.57**($\pm$ 0.39)	82.57($\pm$ 1.92)	81.63($\pm$ 2.26)	88.4($\pm$ 1.08)	89.5($\pm$ 0.72)	88.5($\pm$ 1.03)	94($\pm$ 0.92)
	$\rho = 0.7$	96.77($\pm$ 0.32)	99.23($\pm$ 0.42)	80.4($\pm$ 1.68)	79.4($\pm$ 1.94)	87.47($\pm$ 1.58)	88.47($\pm$ 0.55)	87.73($\pm$ 1.3)	93.9($\pm$ 0.96)
	$\rho = 0.9$	90.5($\pm$ 0.93)	94.4($\pm$ 0.52)	74.1($\pm$ 1.66)	73.9($\pm$ 1.44)	77.37($\pm$ 0.92)	82.93($\pm$ 1.71)	84.9($\pm$ 2.23)	86.77($\pm$ 0.94)
$N = 300$	$\rho = 0.1$	91.77($\pm$ 0.97)	95.43($\pm$ 0.59)	84.13($\pm$ 1.44)	83.67($\pm$ 0.9)	84.67($\pm$ 0.83)	86.83($\pm$ 0.55)	87.83($\pm$ 0.81)	87.97($\pm$ 0.79)
	$\rho = 0.3$	98.33($\pm$ 0.52)	99.5($\pm$ 0.28)	85.73($\pm$ 1.65)	85.1($\pm$ 1.5)	87.17($\pm$ 0.97)	89.23($\pm$ 0.7)	88.8($\pm$ 1.55)	92.7($\pm$ 0.76)
	$\rho = 0.5$	98.4($\pm$ 0.52)	99.43($\pm$ 0.27)	85.93($\pm$ 1.19)	85.43($\pm$ 1.22)	87.6($\pm$ 0.75)	89.47($\pm$ 0.89)	88.97($\pm$ 0.69)	94.93($\pm$ 1.22)
	$\rho = 0.7$	96.77($\pm$ 0.45)	99.23($\pm$ 0.22)	82.27($\pm$ 1.68)	81.73($\pm$ 1.83)	86.23($\pm$ 1.98)	88($\pm$ 1.4)	88.77($\pm$ 0.97)	94.1($\pm$ 1.1)
	$\rho = 0.9$	91.17($\pm$ 1.21)	95($\pm$ 0.35)	75.1($\pm$ 0.92)	74.47($\pm$ 1.49)	77.5($\pm$ 1.34)	82.9($\pm$ 2.85)	85.9($\pm$ 1.49)	87.97($\pm$ 1)
Mean acc.		95.16 ($\pm$ 3.45)	**97.49** ($\pm$ 2.26)	82.06 ($\pm$4.07)	81.57 ($\pm$ 4.11)	84.76 ($\pm$ 3.99)	87.37 ($\pm$ 2.56)	87.62 ($\pm$ 1.74)	91.03 ($\pm$ 3.38)