

UNIVERSIDADE PRESBITERIANA MACKENZIE
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E COMPUTAÇÃO

CLASSIFICAÇÃO AUTOMÁTICA DO COMPORTAMENTO
DINÂMICO DE AUTÔMATOS CELULARES BINÁRIOS
UNIDIMENSIONAIS

São Paulo

2019

MARCELO ARBORI NOGUEIRA

CLASSIFICAÇÃO AUTOMÁTICA DO COMPORTAMENTO DINÂMICO DE
AUTÔMATOS CELULARES BINÁRIOS UNIDIMENSIONAIS

Projeto de doutorado apresentado ao Programa de Pós-graduação em Engenharia Elétrica e Computação da Universidade Presbiteriana Mackenzie como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica, na área de concentração em Engenharia da Computação.

Aprovado em 02 de outubro de 2019.

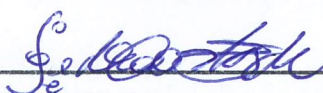
BANCA EXAMINADORA



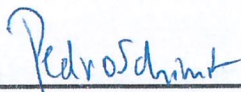
Prof. Dr. Pedro Paulo Balbi de Oliveira
Universidade Presbiteriana Mackenzie



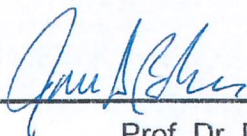
Prof. Dr. Eurico Luiz Prospero Ruivo
Universidade Presbiteriana Mackenzie



Prof. Dr. Pedro Contino da Silva Costa
Macquarie University



Prof. Dr. Pedro Henrique Triguis Schimit
Universidade Nove de Julho



Prof. Dr. Dario Bahamon
Universidade Presbiteriana Mackenzie

**UNIVERSIDADE PRESBITERIANA MACKENZIE
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E COMPUTAÇÃO**

Marcelo Arbori Nogueira

**CLASSIFICAÇÃO AUTOMÁTICA DO COMPORTAMENTO
DINÂMICO DE AUTÔMATOS CELULARES BINÁRIOS
UNIDIMENSIONAIS**

Projeto de doutorado apresentado ao Programa de Pós-graduação em Engenharia Elétrica e Computação da Universidade Presbiteriana Mackenzie como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica, na área de concentração em Engenharia da Computação.

Orientador: Prof. Dr. Pedro Paulo Balbi de Oliveira

São Paulo
2019

Aos meus familiares e amigos.

Dedico.

Agradecimentos

É difícil agradecer a conclusão de uma caminhada tão significativa quanto a que busco concluir com este trabalho. Mas é importante destacar o papel de meus pais, que se não puderam me levar pelas mãos, em nenhum momento deixaram de me incentivar. Minhas irmãs, como as melhores amigas, que também me incentivaram e se alegraram com minha decisão de abraçar o desafio de fazer o doutorado.

Além da família, o papel de amigos e os colegas do curso também foi muito importante, pois em vários momentos ofereceram seus comentários que muito me auxiliaram e por isso sou grato.

Sobretudo agradeço ao meu orientador que sempre esteve presente paciente e dedicado em momentos importantes.

Finalmente, agradeço ao IPM – Instituto Presbiteriano Mackenzie, pela bolsa de isenção de mensalidade que me foi concedida, e registro que este trabalho se beneficiou de auxílios de pesquisa concedidos a meu orientador, pela CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, através dos projetos STIC-AmSud (CoDANet) no. 88881.197456/2018-01 e PrInt no. 88887.310281/2018-00.

Histórias de sucesso são raras, por serem doloridas.

RESUMO

A variabilidade de evoluções temporais geradas por autômatos celulares advém do grande número de regras possíveis, à sua configuração inicial, quantidade de estados, quantidade de células na vizinhança e dimensão do reticulado. Mesmo para os mais simples, o número de regras do autômato celular pode facilmente ser da ordem de bilhões, e ainda que o reticulado seja unidimensional, a quantidade de evoluções temporais possíveis cresce exponencialmente com relação ao tamanho do reticulado. Portanto, classificar a dinâmica típica das evoluções temporais é trabalho de difícil realização, sendo útil qualquer processo automatizado para a tarefa.

Relata-se aqui o desenvolvimento de dois classificadores da dinâmica apresentada pelas evoluções temporais, segundo o esquema de classificação nas 4 classes de Wolfram, tendo como base o espaço elementar, mas tendo também por objetivo aplicá-lo em um espaço maior, cuja classificação é desconhecida, qual seja, aquele com 4 células na vizinhança e 2 estados possíveis.

Inicialmente foi feita uma revisão da classificação das regras do espaço elementar desenvolvida por Wuensche (1998), no qual, a cada iteração do autômato celular, a variação da entropia de evoluções temporais foi associada às classes das regras geradoras. Os resultados obtidos serviram como referência para os dois classificadores desenvolvidos posteriormente.

Dos dois classificadores, um usou rede neural convolucional, treinada para prever a classe da regra associada a uma evolução temporal. Como as 4 classes não possuem a mesma quantidade de regras, o que afeta o treinamento da rede, as regras foram escolhidas aleatoriamente, mas mantendo a mesma proporção para cada classe.

O segundo classificador usou análise de textura para extrair, das evoluções temporais, informação quanto às configurações de vizinhanças das células, de forma a se poder construir um espectro de frequência destas configurações. Um único espectro, com a frequência média de cada configuração possível, associado à regra geradora, foi incluído em um conjunto de dados para ser utilizado no algoritmo k -NN, a fim de se obter a predição da classe em questão.

Os classificadores foram avaliados de duas formas: a princípio, para inferir as classes das regras do espaço elementar, tomando por base os comportamentos típicos delas, aqueles mais comumente apresentados em um conjunto de evoluções temporais. As classes preditas puderam ser comparadas com a classificação conhecida do espaço elementar e se observou acurácia total para ambos. Já para o espaço com 4 células na vizinhança, foi realizada uma classificação visual de todo o espaço. Neste caso, nenhum dos classificadores conseguiu acurácia elevada. Ainda assim, ambos foram capazes de extrair informações do espaço muito maior que o elementar. Por fim, foram usadas matrizes de confusão para avaliar a qualidade dos classificadores com dados dos dois espaços, com ambos os classificadores apresentando dificuldades na classificação do espaço com 4 células na vizinhança.

Palavras-chave: *Autômato celular, rede neural convolucional, classificação de comportamento dinâmico, classes de Wolfram, espectro de configuração de vizinhança, aprendizado profundo.*

ABSTRACT

The variability of temporal evolution generated by cellular automata comes from the large number of possible rules, their initial configuration, the number of states, the number of cells in the neighborhood and the dimension of the lattice. Even for the simplest cases, the number of rules in the space can easily reach billions, and even if the lattice is one-dimensional, the number of possible temporal evolutions grows exponentially as the lattice size grows. Therefore, to classify the typical dynamics of the temporal evolutions is a daunting endeavour, so that any automated process for the task is clearly useful.

We report here the development of two classifiers of the dynamics presented by the temporal evolutions, according to Wolfram's 4-class classification scheme, based on the elementary space, but also aiming to apply it to a larger space, whose classification is unknown, namely, the one with 4 cells in the neighbourhood and 2 possible states.

At first, a review was made of the classification method developed by Wuensche (1998), in which, at each time step of the cellular automaton, the entropy variation observed in the temporal evolution was associated with the generating rule classes. The results obtained served as a reference for the classifiers developed further on.

One of the two classifiers relied on a convolutional neural network, trained to predict the rule class that generated a temporal evolution. Since the 4 classes do not have the same amount of rules, which affects the network training, the rules were chosen randomly, while keeping the same proportion for each class.

The second classifier used texture analysis to extract, from the temporal evolutions, information of the neighborhood configurations of the cells, which allowed for the construction of a frequency spectrum of these configurations. A single spectrum, with the average frequency of each possible configuration associated with the generating rule was then included in a dataset, and used in the k -NN algorithm to obtain the prediction of the class at issue.

The classifiers were evaluated in two ways: at first, to define the classes of the elementary space, according to their typical behaviors, which are the most common ones displayed in a set of temporal evolutions. The predicted classes could be compared with the known classification of elementary space and total accuracy was observed for both. For the space with 4 cells in the neighbourhood, a visual classification of the entire space was performed. In this case, none of the classifiers achieved high accuracy. Still, they were able to extract information from that space, which is larger than the elementary space. Finally, confusion matrices were used to evaluate the quality of the classifiers with data from both spaces, with both classifiers having difficulties in classifying the space with 4 cells in the vicinity.

Keywords: *Cellular automaton, convolutional neural network, dynamic behaviour classification, Wolfram classes, spectrum of neighbourhood configuration, deep learning.*

Lista de símbolos e abreviaturas

AC	Autômato celular
ACE	Autômato celular elementar.
\mathbb{Z}^d	Reticulado celular de dimensão d .
S	Conjunto de estados possíveis para as células do reticulado celular.
f	Função de transição de estado.
G	Função de transição global do AC.
E	Evolução temporal.
\vec{N}	Vetor deslocamento que define as células da vizinhança.
r	Raio em torno da célula central.
d	Dimensão do reticulado celular.
m	Quantidade de células na vizinhança.
q	Um estado possível da célula.
t	Tempo discreto de atualização do AC.
$*$	Operador de convolução de duas funções.
V	Conjunto de estados das células na vizinhança.
V_i	Estado da célula na vizinhança V .
E_i	Valor transformado do pixel i em relação à célula central.
TU	Unidade de textura.
N_{TU}	Número da unidade de textura.
$d(\vec{a}, \vec{b})$	Distância Euclidiana entre dois vetores.
k -NN	Algoritmo k Nearest Neighborhood, usado na classificação.
VP_x	Verdadeiro Positivo da classe x
VN_x	Verdadeiro Negativo da classe x
FP_x	Falso Positivo da classe x
FN_x	Falso Negativo da classe x
AG	Acurácia Global da classe x

TFP_x	Taxa de falsos positivos da classe x
TFN_x	Taxa de falsos negativos da classe x
TFD_x	Taxa de falsas descobertas da classe x
TFO_x	Taxa de falsas omissões da classe x
H	Entropia medida do estado do AC.
\vec{s}_z	Espectro da evolução temporal.
s	Frequência da ocorrência de configuração de vizinhança.
\vec{t}_n	Soma de n histogramas.
\vec{f}_n	Espectro da evolução temporal, normalizado.
\vec{p}	Vetor proporção entre \vec{s}_z e \vec{s}_t .
F	Conjunto de espectros de evoluções temporais.

Figuras

2.1	Evolução temporal da regra 30 em um reticulado com 50 células, executada com a regra por 100 iterações.	8
2.2	Exemplos de evoluções temporais para cada classe de Wolfram. As evoluções classificadas como homogêneas podem ser tanto aquelas onde todas as células são 0, quanto apenas 1. Aqui a estrutura simples da classe 2 é exemplificada por linhas diagonais, mas também podem ser um pouco mais complexas. Porém, nem tão complexas quanto as de classe 3 ou 4, como as duas imagens inferiores.	10
2.3	Regras caóticas do espaço elementar com evolução temporal homogênea após período transiente do mesmo tamanho do reticulado ou a metade. Para estas regras, isso acontece para uma condição de contorno periódica cujo número de células seja potência de 2.	11
2.4	Evolução temporal da regra 60 do espaço elementar, para uma condição inicial aleatória, com dinâmica homogênea a partir do tempo discreto $t = 255$	12
2.5	Representação de neurônio artificial segundo McCulloch e Pitts (1943). . .	14
2.6	Rede de camada única com a entrada \vec{x} servindo de estímulo para mais de um neurônio (RUSSELL; NORVIG, 2010).	14
2.7	Rede feedforward estimulada pela entrada \vec{x} , dando como resultado a saída \vec{y} (RUSSELL; NORVIG, 2010).	15
2.8	Processo de dropout (descarte) de conexões sináptica aleatoriamente. . . .	16
2.9	Ilustração do processo de convolução onde se tem a entrada como uma matriz binária, um filtro (que na rede feedforward faz papel de pesos sinápticos) e a saída como uma extração das características da matriz de entrada (GOODFELLOW; BENGIO; COURVILLE, 2016).	18
2.10	A função de ativação linear retificada possui comportamento quase linear, ela mantém muitas das propriedades que facilitam a otimização dos modelos lineares com métodos baseados em gradiente. Eles também preservam muitas das propriedades que fazem os modelos lineares generalizarem bem (GOODFELLOW, 2016).	24
2.11	Aplicação do algoritmo LBP em torno de um pixel com intensidade 39 em tom de cinza.	27
2.12	Extração de informação de variância do tom de cinza com relação ao pixel central e aqueles que distam uma unidade com ângulos 0° , 45° , 90° e 135°	29

3.1	Associação do padrões visuais observados nas evoluções temporais (primeira coluna), a entropia do sistema em cada momento (segunda coluna) e o histograma das configurações observadas nas evoluções temporais. Observa-se que apenas o comportamento complexo possui alta variância (WUENS-CHE, 1998).	37
3.2	Variação da entropia ao longo de 50 iterações.	38
4.1	Diagrama de blocos que representa o processamento e descreve a arquitetura do classificador, uma rede neural convolucional, utilizada no trabalho. Nele é destacado a redução da dimensão da imagem até aos 4 neurônios de saída (c1, c2, c3, c4), representando a probabilidade de a regra geradora da evolução temporal ser de cada uma das 4 classes possíveis.	45
5.1	Primitiva destacada, no interior do retângulo, do trecho de uma evolução temporal com a vizinhança contendo 3 células, observadas por 3 passos de tempo.	55
5.2	Conversão da configuração de vizinhança em valor decimal toma como referência o momento t e a coluna inicial j_0 . A conversão é feita segunda a Equação 5.1	55
5.3	Diagrama das distâncias entre os espectros das regras do espaço elementar. Cada pixel (x, y) é a informação da distância entre as regras x e y , onde as distâncias maiores são representadas por tons mais escuros e as mais baixas por tons claros.	58
5.4	Diagrama das distâncias de um conjunto contendo apenas espectros das regras equivalentes. Cada pixel (x, y) é a informação da distância entre as regras x e y , onde as distâncias maiores são representadas por cores mais escuras e por consequência, as mais baixas pelas claras.	59
5.5	Comparação dos espectros gerados para as regras dinamicamente equivalentes à regra 30, utilizando condições iniciais aleatórias para cada uma. . .	60
5.6	Comparação dos espectros gerados para as regras dinamicamente equivalentes à regra 54, utilizando condições iniciais aleatórias para cada uma. . .	60
5.7	Variação da norma do vetor f para a regra 27 do espaço elementar para reticulados com 30 células e média móvel de 250 evoluções temporais. . . .	61
5.8	Diagrama que sintetiza a classificação utilizando análise de textura. Destaca a criação do espectro médio a partir de vários espectros criados a partir de evoluções temporais de uma regra. E também, como o classificador prediz a classe de uma evolução temporal.	62
5.9	Evolução temporal da regra 1572, de raio 1,5, classificada como classe 1, sendo correta a classe 2.	72
5.10	Comparação entre o espectro da regra 224 do espaço elementar e o da regra 1.572 do espaço de raio 1,5, evidenciando a dificuldade na diferenciação entre ambos, usando a distância Euclidiana.	72

Tabelas

2.1	Exemplo de regra para um AC, função tabelada associando as vizinhanças possíveis, quando o raio $r = 1$, a estados de $\Sigma = \{0, 1\}$	7
2.2	Representação da evolução temporal (a), entrada da rede, e o filtro (b) da primeira camada convolucional.	19
2.3	Resultado da primeira camada convolucional (a) e o filtro da segunda camada (b).	19
2.4	Resultado da segunda camada convolucional (a) e o filtro da terceira camada (b).	20
2.5	Resultado da terceira camada convolucional (a) e o filtro da quarta camada (b).	20
2.6	Resultado da quarta camada convolucional e saída da rede.	21
2.7	A evolução temporal original preenchida com células nulas. Usando um filtro 3×3 , o resultado da convolução mantém a dimensão da matriz original	22
2.8	Operação maxpooling realizada em uma matriz 4×4 , extraindo as características mais marcantes como os valores mais altos, obtendo a matriz resultante com dimensão 2×2	23
2.9	Matriz de Confusão - Contagem das predições.	32
2.10	Matriz de Confusão - Índices maximizáveis.	33
2.11	Matriz de Confusão - Índices minimizáveis.	34
3.1	Resultado da classificação do espaço elementar usando a variação da entropia das evoluções temporais.	39
3.2	Matriz de confusão proveniente da classificação por entropia do espaço elementar.	39
3.3	Métricas da matriz de confusão proveniente da classificação por entropia do espaço elementar.	40
3.4	Matriz de confusão proveniente da classificação por entropia do espaço de raio 1,5.	41
3.5	Métricas da matriz de confusão proveniente da classificação por entropia do espaço elementar.	42
4.1	Resultado da classificação do espaço elementar usando rede neural convolucional.	48
4.2	Matriz de confusão proveniente da classificação do espaço elementar via rede neural convolucional.	48
4.3	Métricas da matriz de confusão proveniente da classificação do espaço elementar via rede neural convolucional.	49
4.4	Resultado da classificação do espaço raio 1,5 usando rede neural convolucional.	50

4.5	Matriz de confusão da classificação do espaço de raio 1,5 via rede neural convolucional.	50
4.6	Métricas da matriz de confusão da classificação do espaço de raio 1,5 via rede neural convolucional.	51
5.1	Resultado da classificação do espaço elementar usando usando análise de textura. Conjunto de dados baseados nas regras equivalentes e quantidade fixa de evoluções temporais.	63
5.2	Matriz de confusão do experimento 1 - Fixo.	64
5.3	Métricas da matriz de confusão do experimento 1 - Fixo.	64
5.4	Resultado da classificação do espaço elementar usando usando análise de textura. Conjunto de dados baseados nas regras equivalentes e quantidade de evoluções temporais variáveis para gerar F	65
5.5	Matriz de confusão do experimento 1 - Variável.	65
5.6	Métricas da matriz de confusão do experimento 1 - Variável.	66
5.7	Resultado da classificação do espaço elementar usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F	67
5.8	Matriz de confusão do experimento 2 - Fixo.	67
5.9	Métricas da matriz de confusão do experimento 2 - Fixo.	68
5.10	Resultado da classificação do espaço elementar usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade variável de evoluções temporais para gerar F	68
5.11	Matriz de confusão do experimento 2 - Variável.	69
5.12	Métricas da matriz de confusão do experimento 2 - Variável.	70
5.13	Resultado da classificação do espaço raio 1,5 usando usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F	73
5.14	Matriz de confusão do experimento no espaço de raio 1,5 usando conjunto espectros de referência baseados nas regras equivalentes, utilizando 500 evoluções temporais na geração do conjunto e na predição.	74
5.15	Métricas da matriz de confusão do experimento 1 - Fixas.	74
5.16	Resultado da classificação do espaço raio 1,5 usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F	75
5.17	Matriz de confusão do experimento no espaço de raio 1,5. Conjunto espectros de referência baseados nas regras equivalentes. Quantidades variáveis de evoluções temporais na geração do conjunto e 500 fixas na predição.	75
5.18	Métricas da matriz de confusão do experimento 1 - Variável.	76
5.19	Resultado da classificação do espaço raio 1,5 usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F	76
5.20	Matriz de confusão do experimento no espaço de raio 1,5. Conjunto de espectros de referência baseados nas regras do espaço elementar. Utilizando 500 evoluções temporais fixas na geração e predição.	77
5.21	Métricas da matriz de confusão do experimento 2 - Fixas.	77

5.22	Resultado da classificação do espaço raio 1,5 usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade variável para gerar F	78
5.23	Matriz de confusão do experimento no espaço de raio 1,5. Conjunto espectros de referência baseados nas regras do espaço elementar. Utilizando quantidades variáveis de evoluções temporais na geração e 500 para a predição.	78
5.24	Métricas da matriz de confusão do experimento 2 - Variáveis.	79

Sumário

1	Introdução	1
2	Referencial teórico	5
2.1	Autômato Celular	5
2.2	Classificação de Wolfram aplicada aos autômatos celulares elementares . . .	9
2.3	Redes neurais	12
2.3.1	Redes neurais convolucionais	17
2.4	Análise de textura	24
2.4.1	Algoritmo k -NN	30
2.5	Matriz de confusão	31
3	Entropia e classificação de autômatos celulares	35
3.1	Resultados da classificação utilizando entropia	39
4	Redes Neurais e classificação de autômatos celulares	43
4.1	Proposta para a classificação de autômatos celulares binários unidimensionais usando redes neurais	43
4.2	Classificação via rede neural	46
4.2.1	Classificação do espaço elementar via redes neurais convolucionais .	47
4.2.2	Classificação do espaço de raio 1,5 via redes neurais convolucional .	49
5	Análise de textura e classificação de autômatos celulares	52
5.1	Proposta para a classificação de autômatos celulares binários unidimensionais usando análise de textura	53
5.2	Classificação via análise de textura	57
5.2.1	Experimentos para as regras dinamicamente equivalentes	63
5.2.2	Experimentos para as regras do espaço elementar	66
5.2.3	Classificação do espaço de raio 1,5	70

Capítulo 1

Introdução

Proposto ainda na década de 1940 como modelo computacional que descreve a máquina auto-replicante de Neumann (1966), os Autômatos Celulares (ACs) são também modelos matemáticos aplicados em problemas variados. A estrutura do AC possui um conjunto organizado de células que podem assumir um dos estados de um conjunto finito, transitando entre eles ao longo do tempo discreto devido à ação de uma das funções de transição local, também um conjunto finito de possibilidades. O resultado é uma evolução temporal dos estados das células do reticulado (SARKAR, 2000).

A função de transição local contém um número finito de transições possíveis e a este conjunto de transições é dado o nome de regra. O conjunto de regras possíveis, dada a configuração do AC, é um espaço discreto definido pelo total de regras nele contido. Uma regra define o comportamento dinâmico do AC, e para cada uma, obtém-se suas dinâmicas típicas. Dentre as regras possíveis de cada espaço, delimitar aquelas de interesse para simulações de fenômenos, ou que apresentem comportamento dinâmico, foco de pesquisas, é uma tarefa difícil devido à quantidade de regras possíveis em cada espaço, que facilmente supera os bilhões. E por ser um sistema emergente, em geral não é possível prever o estado futuro do AC após a aplicação de uma regra por um tempo arbitrário, sem que efetivamente aplique-se a regra por este tempo (CULIK II; YU, 1988). As condições iniciais possíveis dificultam ainda mais a possibilidade de se predizer o estado futuro do sistema (WOLFRAM, 1984).

Além das regras, condições iniciais possíveis dificultam ainda mais a possibilidade de

se prever o estado futuro do AC. A condição inicial do sistema é a configuração de suas células, em qual dos estados possíveis cada célula se encontra. Por exemplo, no caso de 3 estados e 83 células, tem-se 3^{83} condições iniciais possíveis e cada uma delas resulta em uma evolução temporal com dinâmica própria, devido à aplicação de uma regra. A classificação de regras no contexto dos ACs é classificar esta dinâmica, em um conjunto de evoluções temporais geradas por uma mesma regra, definindo o comportamento típico da regra aplicada. Mas, como será discutido mais adiante, nem todas as condições iniciais levam ao comportamento típico da regra, o que pode levar a erros de classificação. A necessidade da geração de um conjunto representativo para se deduzir o comportamento dinâmico típico e a quantidade de possibilidades apresentadas, dificultam a classificação de regras do AC baseando-se nas evoluções temporais, sendo esta uma tarefa computacionalmente difícil.

O conjunto de regras largamente estudado é conhecido como espaço elementar que é um AC binário unidimensional, cuja vizinhança contém 3 células. Por esta configuração, o espaço elementar conta com 256 regras possíveis, que podem ser agrupadas em 4 classes segundo Wolfram (2002). O objetivo deste trabalho é a obtenção de meios para a classificação automática de evoluções temporais geradas pelas regras do espaço elementar e aplicar a mesma técnica em evoluções temporais de espaços maiores, com o objetivo de estabelecer uma primeira classificação para um espaço cuja classificação é desconhecida.

Dois classificadores foram desenvolvidos e avaliados: O primeiro utilizou redes neurais convolucionais tendo como entrada a própria evolução temporal como imagem binária, sendo o conjunto de treinamento o par (regra, evolução temporal). A vantagem no uso deste tipo de rede é a não necessidade de pré-processar a imagem, pois a extração de características, normalmente feita para outro tipo de rede, não se faz necessário.

O segundo classificador utilizou análise de texturas para se obter uma estratégia de classificação baseada no espectro de configurações das vizinhanças de cada célula das evoluções temporais. Construindo um conjunto de dados baseado nestes espectros, foi possível associar evoluções temporais a regras do espaço elementar, predizendo a classe de forma simples, pois a classificação do espaço elementar é conhecida.

É importante distinguir a classificação das regras de um espaço e a avaliação dos classificadores feita com o auxílio da matriz de confusão. A regra de uma classe é definida

pelo comportamento mais frequente observado em um conjunto de evoluções temporais, seu comportamento típico. Neste caso, ambos os classificadores obtiveram acurácia total ao classificar o espaço elementar. Ou seja, para cada uma das 256 regras do espaço elementar, gerado um conjunto com evoluções temporais suficientes, ambos os classificadores acertaram todas as predições ao serem avaliados. Porém, ao avaliar a qualidade de cada um utilizando a matriz de confusão, evidenciou-se as fragilidades de cada classificador. Estas fragilidades ficaram mais evidentes ao aplicá-los ao espaço raio 1,5, que contém 4 células em sua vizinhança e dois estados possíveis, cuja classificação não era conhecida.

Aplicando os classificadores na classificação do espaço de raio 1,5 observou-se grande sensibilidade às condições locais da imagem em ambos os casos. A análise de textura leva em conta apenas informações locais aos pixels, não sendo possível uma análise global das configurações. Tal fato fez com que a acurácia ao classificar o espaço de raio 1,5 fosse substancialmente inferior à obtida para o elementar, devido às configurações locais deste espaço maior, não presente ou menos frequente no espaço elementar. A mesma situação ocorre ao usar redes neurais convolucionais, pois a rede extrai automaticamente características locais da imagem. Sendo sua performance afetada da mesma forma que a análise de textura. Ainda assim, ambos os classificadores permitiram obter informações referentes ao espaço de classificação desconhecida.

Apesar do comportamento semelhante entre os dois classificadores, quanto à extração de informações da imagem, há diferenças importantes em termos computacionais. A forma como as informações foram estruturadas na análise de textura, permite adicionar espectros de forma relativamente simples ao conjunto de dados e prever corretamente a classe da regra geradora de evoluções temporais submetidas ao classificador, inicialmente predita de forma equivocada. No caso da rede neural convolucional, novos ciclos de treinamento com o objetivo de se obter melhor acurácia para o espaço de regras binárias de 4 células, por exemplo, apresentou custo computacional substancialmente maior, se comparado à análise de textura.

O próximo capítulo aborda os detalhes sobre autômatos celulares, redes neurais, análise de textura, algoritmo k -NN e matriz de confusão, pontos importantes para o desenvolvimento dos classificadores. No Capítulo 3 analisa-se a uma primeira abordagem da classificação dos autômatos celulares, partindo-se do que há na literatura, referente

ao usa da entropia. No Capítulo 4 aborda-se a classificação utilizando redes neurais. E no Capítulo 5, como a classificação pode ser feita via análise de textura. Finalmente no Capítulo 6 tem-se a conclusão.

Capítulo 2

Referencial teórico

Uma definição e delimitação mais precisa do AC abordado neste trabalho será dada na Seção 2.1 e a classificação utilizada será explicada na Seção 2.2. Serão utilizadas duas técnicas para desenvolver a classificação automática: Redes Neurais Convolucionais, abordada na Seção 2.3 e Análise de textura, explicada na Seção 2.4. A performance de ambos os classificadores será feita utilizando matriz de confusão, cujos detalhes são explicados na Seção 2.5.

2.1 Autômato Celular

Os ACs abordados aqui utilizam reticulado discreto de células com organização de uma, duas, três ou mais dimensões. Uma célula pode ser localizada no reticulado pelas coordenadas de sua posição como um vetor. Os vetores $\vec{x} = (7)$, $\vec{x} = (8, 7)$ e $\vec{x} = (3, 4, 0)$, são representações de células em reticulado linear, de duas e três dimensões, respectivamente. Representa-se um reticulado qualquer na forma \mathbb{Z}^d , onde d representa a dimensão. O conjunto discreto S contém os valores possíveis para cada célula do reticulado (KARI, 2005).

Uma função de transição local é responsável por alterar o valor de uma célula, relacionando a configuração de sua vizinhança a um estado de S , $f : S^n \rightarrow S$. A partir de uma célula usada como referência, define-se a vizinhança como um vetor de vetores que somados às coordenadas da célula de referência, resulta em deslocamentos, localizando

as células na vizinhança. Este vetor pode ser escrito na forma $\vec{N} = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)$, onde $\vec{x}_i \in \mathbb{Z}^d$. A configuração da vizinhança de uma célula relaciona elementos de \mathbb{Z}^d a estados de S , $c : \mathbb{Z}^d \rightarrow S$. Assim, a configuração de vizinhança da célula de referência no momento $t + 1$ é definida pela configuração em t , $c_{t+1}(\vec{x}) = f(c_t(\vec{x} + \vec{x}_1), c_t(\vec{x} + \vec{x}_2), \dots, c_t(\vec{x} + \vec{x}_n))$.

Uma vizinhança com células adjacentes são casos particulares. Se a vizinhança seguir uma estrutura de von Neumann, vale a relação $\|\vec{N}\|_1 \leq r$. Mas, se a vizinhança for de Moore a relação é $\|\vec{N}\|_\infty \leq r$ em relação ao vetor de vizinhança, onde o valor do raio r é a distância máxima entre a célula de referência e qualquer outra da vizinhança. A quantidade total de células ao redor da célula de referência para as vizinhanças von Neumann e Moore são respectivamente $m = dr(r + 1) + 1$ e $m = (2r + 1)^d$.

Ao descrever os deslocamentos das células na vizinhança, o vetor \vec{N} não descreve obrigatoriamente aquelas que são adjacentes, sendo este um caso particular, como as vizinhanças von Neumann e Moore. Para esta vale $\|\vec{N}\|_\infty \leq r$, enquanto para aquela vale $\|\vec{N}\|_1 \leq r$, que são as normas máxima e Manhattan respectivamente, em relação ao vetor de vizinhança, onde o valor do raio r é a distância máxima, para cada caso, entre a célula de referência e qualquer outra da vizinhança. A quantidade total de células ao redor da célula de referência para as vizinhanças von Neumann e Moore são respectivamente $m = dr(r + 1) + 1$ e $m = (2r + 1)^d$.

A vizinhança adotada influencia a dinâmica apresentada pelo AC ao longo do tempo, bem como o tipo de configuração que pode ser finita ou infinita. Dado um estado $q \in S$ definido como quiescente, a função de transição local aplicada sobre uma configuração quiescente resulta no mesmo estado q . Sendo finita, aquela configuração onde $\{\vec{x} \in \mathbb{Z}^d | c(\vec{x}) \neq q\}$ é um conjunto finito. Caso contrário, a configuração é infinita.

Em Kari (2005) também é definida uma função de deslocamento da configuração do CA. Para cada dimensão $i = 1, 2, \dots, d$ há uma função de deslocamento σ_i cuja vizinhança contém apenas vetores diretores \vec{e}_i sendo a função de transição local, a identidade. Uma translação é a combinação de funções de deslocamento. Neste sentido uma configuração periódica é aquela onde $c(\vec{x}) = c(\vec{x} + t\vec{e}_i)$, onde $t \in \mathbb{N}$ é o quanto a configuração foi deslocada no sentido \vec{e}_i .

Assim, a definição de um AC é uma tupla $A = (S, N, f, d)$. O conjunto de configurações possíveis é dado por $C = S^{\mathbb{Z}^d}$. Aplicando f sincronicamente a todas as

células do reticulado, a configuração do AC varia entre as configurações possíveis e se observa um comportamento global, representado por uma função de transição global descrita como $G : C \rightarrow C$. A dinâmica apresentada pelo AC à medida que se aplica tal função de transição é o resultado da iteração de G . A cada momento discreto t se aplica sobre o reticulado de células a função $G^t(C)$, sendo t o número de vezes em que G foi composta com ela mesma ($G(G(G(...)))$). Dessa forma pode-se definir a tupla $E = (G^1(C_0), G^2(C_1), \dots, G^t(C_{t-1}))$ como uma evolução temporal finita, a variação dos estados das células ao longo do tempo, onde C_0 é a condição inicial e $C_t = G^t(C_{t-1})$ é a configuração do AC no momento t .

Um AC Elementar (ACE) que será usado como referência para o aprendizado de máquina nos próximos capítulos é um AC binário unidimensional, onde $\vec{x} \in \mathbb{Z}$, $S = \{0, 1\}$, $N = \{(-1), (0), (1)\}$, $c : \mathbb{Z} \rightarrow S$ e $c_{t+1} = f(c_t(\vec{x} + (-1)), c_t(\vec{x} + (0)), c_t(\vec{x} + (1)))$. Para $d = 1$ e $r = 1$ o número total de células é o mesmo, tanto para a vizinhança de von Neumann quanto Moore, $dr(r + 1) + 1 = (2r + 1)^d = 3$. Assim, é possível definir uma função de transição local f , ou regra, como a representada pela Tabela 2.1 para o ACE.

Tabela 2.1: Exemplo de regra para um AC, função tabelada associando as vizinhanças possíveis, quando o raio $r = 1$, a estados de $\Sigma = \{0, 1\}$.

Vizinhança	Transição	Vizinhança	Transição
(1, 1, 1)	0	(0, 1, 1)	1
(1, 1, 0)	0	(0, 1, 0)	1
(1, 0, 1)	0	(0, 0, 1)	1
(1, 0, 0)	1	(0, 0, 0)	0

Com 3 células na vizinhança e dois estados em S , há $2^3 = 8$ valores possíveis para as vizinhanças da função de transição local e como cada uma pode variar entre os dois estados de S , a quantidade total de regras para o espaço elementar é $2^8 = 256$ possíveis, identificadas segundo os estados resultantes devido à ação da função. A regra da Tabela 2.1 pode ser representada pela sequência binária $00011110_2 = 30_{10}$ e por isso ela é denominada regra 30, o mesmo valendo para as demais 255. Para o espaço de raio 1,5 vale a mesma contagem, assumindo uma vizinhança de Moore, há $(2r + 1)^d = (2\frac{3}{2} + 1)^1 = 4$ células na vizinhança. O que resulta em $2^4 = 16$ transições de estados para a função de

transição local e $2^{16} = 65536$ regras possíveis para o espaço de raio 1,5 (MARTINEZ, 2013).

No caso dos ACEs, o resultado da função de transição global G é um reticulado de células. Logo, representa-se a evolução temporal E por uma matriz binária onde cada linha representa o estado do reticulado naquele momento discreto. Fazendo uma substituição dos estados por cores, assumindo que 0 representa branco e 1 a cor preta, obtém-se um padrão visual característico de cada regra. Comportamento emergente, uma vez que os padrões globais, como os apresentado na Figura 2.1, não se encontram na regra do AC (WOLFRAM, 2002).



Figura 2.1: Evolução temporal da regra 30 em um reticulado com 50 células, executada com a regra por 100 iterações.

Além dos elementos já citados, o padrão visual que surge ao se aplicar a regra do AC por um período de tempo, também depende do reticulado, que pode ser finito ou infinito. E aquele pode ter ainda condição de contorno periódica, ou seja, a sequência de células do reticulado não possui extremidades, tendo a geometria de anel. Ou o reticulado finito pode ter uma condição de contorno não-periódica, implicando na existência de uma célula inicial e outra final. Para cada um dos casos o padrão visual que surge pode ser

completamente diferente. Para este trabalho o reticulado considerado possui condição de contorno periódico. O espaço elementar foi largamente estudado, iniciando pelos trabalhos de Wolfram (2002) que o classificou em 4 classes distintas a depender do comportamento dinâmico apresentado na evolução temporal, classificação abordada com mais detalhes na próxima seção.

2.2 Classificação de Wolfram aplicada aos autômatos celulares elementares

Como visto, aplicar uma regra sobre um reticulado ao longo do tempo, faz com que seu estado seja alterado a cada momento. Gerando uma evolução temporal dos estados do reticulado que pode ser representada como uma imagem binária, sendo possível observar o padrão visual devido à aplicação da regra. Para um conjunto grande o suficiente de reticulados com estados diferentes entre si, as condições iniciais, aplicando a mesma regra sobre todos eles, as evoluções temporais resultantes possuirão semelhanças, o comportamento típico da regra.

Há várias formas de se classificar as regras geradoras das evoluções temporais, cada uma com um objetivo. Em Li and Packard (1990) as regras foram separadas em 5 classes. No Capítulo 3 será abordado como classificar automaticamente regras de autômatos celulares usando entropia, segundo a abordagem de Wuensche (1998) que separa as evoluções temporais em 3 classes. Mas, neste trabalho foi escolhida a classificação de Wolfram (1984) que separa as regras dos autômatos celulares em 4 classes, a saber:

1. A evolução leva a um estado homogêneo, com apenas 1s ou 0s;
2. A evolução leva a um conjunto de estruturas simples e estáveis ou periódicas;
3. A evolução leva a um padrão caótico;
4. A evolução leva a estruturas localizadas complexas, às vezes de longa duração.

A Figura 2.2 mostra exemplos para cada uma das classes. Da esquerda para a direita, as duas primeiras evoluções temporais são da classe 1, a terceira é da 2, e as duas a baixo são respectivamente das classes 3 e 4.

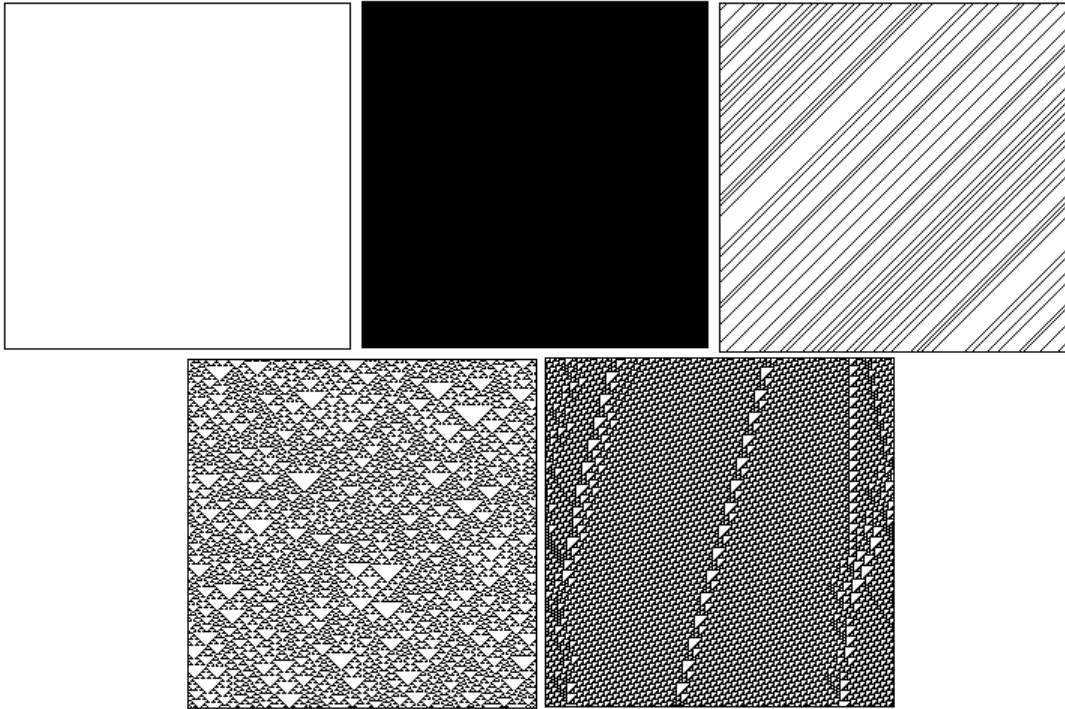


Figura 2.2: Exemplos de evoluções temporais para cada classe de Wolfram. As evoluções classificadas como homogêneas podem ser tanto aquelas onde todas as células são 0, quanto apenas 1. Aqui a estrutura simples da classe 2 é exemplificada por linhas diagonais, mas também podem ser um pouco mais complexas. Porém, nem tão complexas quanto as de classe 3 ou 4, como as duas imagens inferiores.

Algo a ser considerado relacionado à classificação de ACs binários é o fato de não ser possível classificá-los completamente, segundo Culik (CULIK II; YU, 1988). A grande variabilidade de resultados possíveis para a evolução temporal devido à condição inicial, resulta em casos de classificação equivocada da regra geradora, como exemplificado na Figura 2.3, onde as regras classe 3 (60, 90, 102, 153, 155 e 165) apresentam dinâmica homogênea após período transiente, o que levaria a se classificar estas regras como classe 1 se o período transiente não fosse levado em conta.

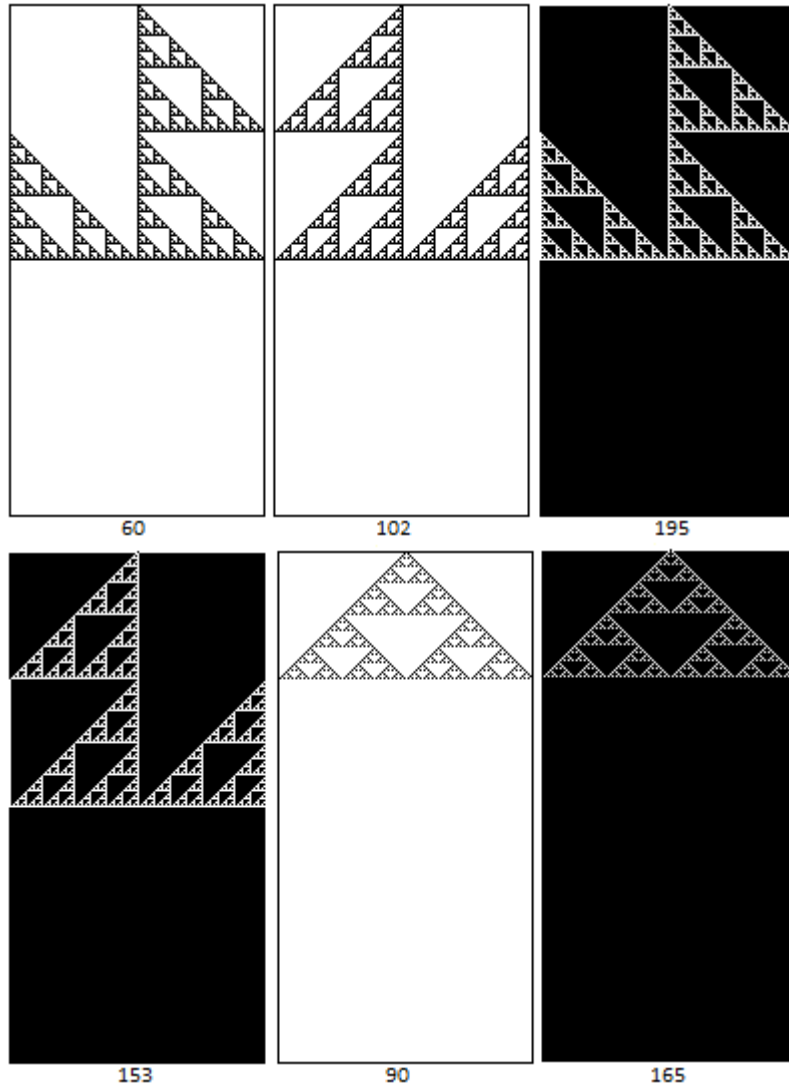


Figura 2.3: Regras caóticas do espaço elementar com evolução temporal homogênea após período transitente do mesmo tamanho do reticulado ou a metade. Para estas regras, isso acontece para uma condição de contorno periódica cujo número de células seja potência de 2.

Para estas regras isso ocorre, mesmo quando a condição inicial é aleatória, para reticulados de tamanho que seja potência de 2 e tenha condição de contorno periódica. A Figura 2.4 mostra evolução temporal da regra 60 do espaço elementar em um reticulado com 256 células e condição de contorno periódica, que apresenta configuração homogênea a partir do tempo discreto $t = 255$. O mesmo ocorre para as regras 102, 153 e 195. Enquanto que para as regras 90 e 165, ocorre configuração homogênea a partir do tempo discreto 127.

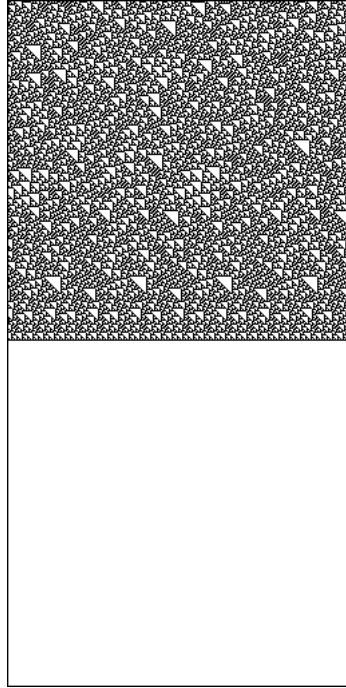


Figura 2.4: Evolução temporal da regra 60 do espaço elementar, para uma condição inicial aleatória, com dinâmica homogênea a partir do tempo discreto $t = 255$.

A quantidade de condições iniciais possíveis propicia que situações particulares possam levar a evoluções temporais não aderentes ao comportamento típico da regra geradora. A necessidade de uma ferramenta que permita classificar automaticamente uma evolução temporal vem, não somente, do grande número de condições iniciais, mas também pela quantidade de regras possíveis, sendo 256 para o ACE; porém; se o raio da vizinhança for 2, será superior a 4 bilhões de regras possíveis. Classificar este espaço é tarefa computacionalmente difícil.

2.3 Redes neurais

O clássico artigo de McCulloch e Pitts (1943) associou os conhecimentos sobre a fisiologia do neurônio com a lógica matemática, propondo a computação de qualquer função computável a partir da construção de redes neurais. E apesar de os autores indicarem que não estavam propondo que uma rede neural natural funcionasse de tal forma, o trabalho permitiu abordagem mais analítica das redes neurais artificiais e o desenvolvimento posterior na área. Este trabalho, que é considerado o início das pesquisas em redes neurais

e inteligência artificial, causou grande impacto na comunidade quando publicado (HAYKIN,1998).

Os anos seguintes foram muito profícuos, com muitos trabalhos relevantes, inclusive a publicação de Hebb (1949), o primeiro trabalho a trazer formulação de regra para aprendizagem e modificação sináptica, trazendo conceito importante de reforço ou enfraquecimento da sinapse de acordo com a frequência da ativação neural devido à conexão com outro neurônio. E apesar de seu grande impacto em outras áreas, seu livro teve pouca relevância na engenharia até o trabalho de Rochester et al. (1956) que simulou o comportamento de 69 e 512 neurônios artificiais, observando neste último, conforme previsto pela teoria, agrupamento de neurônios.

Apesar do avanço na teoria das redes neurais nos anos que se seguiram, Minsky e Papert (1969) publicaram o livro que mostrou a limitação dos perceptrons de uma camada e lançando dúvidas se tais limitações seriam superadas em perceptrons de múltiplas camadas, o que levou ao atraso nas pesquisas de redes neurais em mais de 10 anos. Na década de 80 o interesse sobre as redes neurais foi retomado devido a trabalhos com as hoje chamadas redes de Hopfield, que fazem uma associação entre o conhecimento sobre redes neurais e física estatística. A partir deste momento, devido ao grande avanço nas pesquisas envolvendo redes neurais, elas passaram a ser utilizadas para solucionar uma gama ampla de problemas, dos quais este trabalho se foca na classificação de imagens via rede neurais (HAYKIN,1998).

O modelo de redes neurais artificiais é inspirado nas redes de neurônios naturais que compõem o sistema nervoso. E sem entrar em detalhes biológicos a respeito do funcionamento interno das células neurais, cada neurônio recebe estímulos de vários outros, conectados a ele através das sinapses, respondendo apenas quando certo limiar de intensidade destes estímulos é ultrapassado. Neste ponto o neurônio receptor passa a estimular outros ou agir no controle de alguma parte do corpo. A modelagem artificial do neurônio representa tanto os estímulos, as sinapses e a resposta do neurônio como valores numéricos, onde os estímulos constituem o vetor de entrada, as sinapses, pesos atribuídos às conexões, enquanto a saída é o resultado de uma função. A representação do comportamento neural envolve a soma das entradas ponderadas pelos pesos, passada para a função que simula o limiar de ativação, oferecendo a resposta ao estímulo recebido. A computação feita por

um neurônio j pode ser resumida pela Equação 2.1, onde a saída y_j é resultado da função de ativação f que recebe a soma ponderada entre as entradas x_i e os pesos w_{ij} . Por isso a computação feita por redes neurais é dita distribuída, pois cada neurônio processa suas entradas e responde de acordo (HAYKIN,1998).

$$y_j = f \left(\sum_{i=0}^{I-1} x_i \times w_{ij} \right) \quad (2.1)$$

O neurônio artificial proposto por McCulloch e Pitts (1943) pode ser representado como o vértice de um grafo direcionado, onde as entradas (estímulos) x_i estão conectadas pelas arestas ponderadas pelos pesos w_{ij} , indicando a sinapse entre os neurônios i e j (Figura 2.5). A entrada $\vec{x} = (x_0, x_1, \dots, x_i, \dots, x_{I-1})$ pode servir de estímulo para vários neurônios, o que permite montar uma rede de camada única (Figura 2.6). Agregando mais camadas e interligando seus neurônios se obtém uma rede de múltiplas camadas, cujas mais simples são as redes feedforward, Figura 2.7 (RUSSELL; NORVIG, 2010).

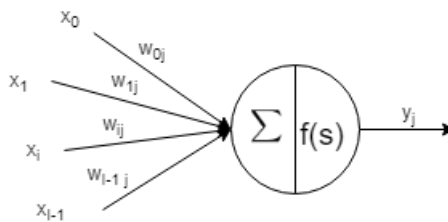


Figura 2.5: Representação de neurônio artificial segundo McCulloch e Pitts (1943).

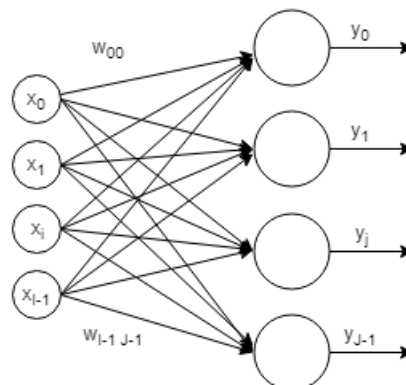


Figura 2.6: Rede de camada única com a entrada \vec{x} servindo de estímulo para mais de um neurônio (RUSSELL; NORVIG, 2010).

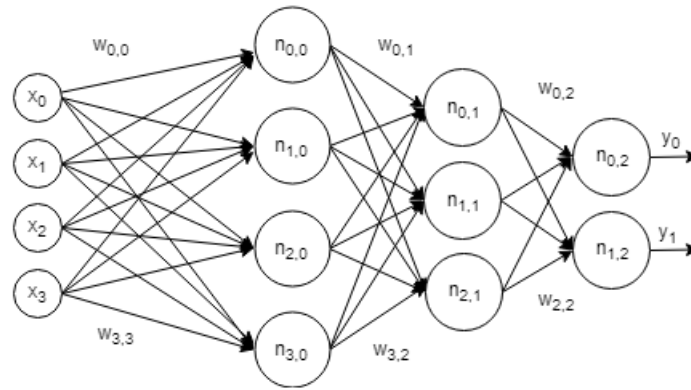


Figura 2.7: Rede feedforward estimulada pela entrada \vec{x} , dando como resultado a saída \vec{y} (RUSSELL; NORVIG, 2010).

A rede neural da Figura 2.7 possui 4 camadas com duas delas ocultas, sendo estas as camadas não estimuladas pela entrada e tão pouco geram a saída. Partindo da primeira camada, a entrada \vec{x} se propaga através das conexões dos neurônios até que a última camada gere a saída \vec{y} da rede. A instância de qualquer problema que possa ser representado por uma entrada \vec{x} poderia ser modelado por redes neurais. Os pesos da rede representam o conhecimento do problema a ser resolvido, uma vez que eles alteram a entrada até que a saída seja gerada.

Mas se a saída obtida pela rede não for a desejada, significa que o conjunto de pesos da rede não representam o conhecimento necessário sobre o problema, evidenciando a necessidade de alterá-los. Algoritmos que alteram os pesos da rede neural artificial (RNA), para adequar o conhecimento da rede àquele esperado, são chamados algoritmos de aprendizado, dos quais pode-se identificar dois grupos: os algoritmos de aprendizado supervisionado e os não-supervisionados. Para este espera-se que após a submissão de entradas à rede seus pesos sejam ajustados para que a saída se estabilize em um ponto, não indicado a priori. No caso dos algoritmos supervisionados a saída desejada está associada à entrada, sendo possível obter o erro, referente à saída da rede, que se pretende mínimo. Neste caso o algoritmo altera os pesos de modo a obter o menor erro possível após o treinamento. A aplicação de uma ou outra classe destes algoritmos está ligado ao problema a ser modelado. No escopo deste trabalho levam-se em conta apenas algoritmos supervisionados, uma vez que se aborda a classificação do comportamento dinâmico de evoluções temporais em 4 classes pré-estabelecidas, de forma que treinar a rede para reconhecê-las,

significa alterar os pesos sinápticos até que a rede possua precisão na identificação das classes (RUSSELL; NORVIG, 2010).

Além dos algoritmos de treinamento, a arquitetura da rede influencia muito em sua acurácia. Uma forma de aumentá-la é a eliminação aleatória de parte das conexões sinápticas de rede para se evitar o overfitting, que é o efeito obtido quando após o treinamento da rede ela prediz perfeitamente o conjunto de treinamento, mas se mostra incapaz de previsões precisas para qualquer outro conjunto. À eliminação de conexões sinápticas dá-se o nome de dropout (descarte), que está ilustrado na Figura 2.8. Não há como definir antecipadamente a taxa de descarte das conexões, sendo esta definida no desenvolvimento da rede neural, observando a acurácia obtida em cada fase de treinamento (SRIVASTAVA et al., 2014).

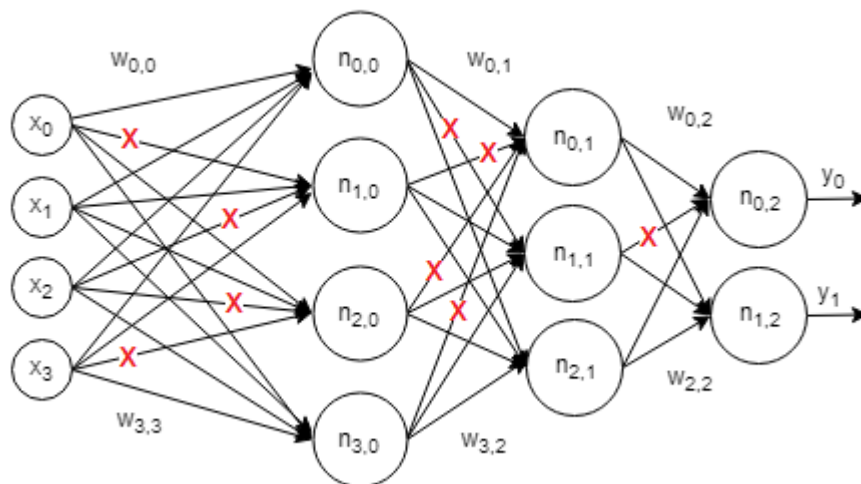


Figura 2.8: Processo de dropout (descarte) de conexões sináptica aleatoriamente.

A estrutura da rede feedforward, Figura 2.7, é uma das estruturas mais simples de redes neurais, mas há muitas outras propostas, para a solução de diversos problemas. O tipo de problema a ser tratado aqui é a classificação de imagens, tarefa para a qual as redes neurais convolucionais são mais adequadas que uma rede feedforward. A Seção 2.3.1 aborda este tipo de rede que são o tipo de rede utilizada neste trabalho para abordar a classificação dos ACs.

2.3.1 Redes neurais convolucionais

A estrutura da rede feedforward representada na Figura 2.7 é uma arquitetura simples, mas muito versátil e aplicada na solução de problemas. No entanto, na classificação de imagens ela apresenta uma limitação importante: a necessidade de pré-processamento na extração de características referente à imagem que se pretende classificar; o que é submetido à rede é um vetor destas características. Este pré-processamento é desnecessário para redes neurais convolucionais, pois estas são capazes de extrair as características da imagem automaticamente, devido à convolução empregada nas camadas convolucionais (GOODFELLOW; BENGIO; COURVILLE, 2016).

A convolução é um operador que age sobre duas funções dadas (h e g) e resulta em uma terceira ($h * g$). Definindo que g seja um sinal ou o resultado de um sistema qualquer e h seja um filtro arbitrário, $(h * g)$ é a soma do produto entre h e g , segundo o deslocamento de uma região subentendida pelo filtro h . Este processo é resumido pela equação Equação 2.2 (HAYKIN and VEEN, 1998).

$$(h * g)[n_0, n_1] = \sum_{m_0=-\infty}^{\infty} \sum_{m_1=-\infty}^{\infty} h[n_0 - m_0, n_1 - m_1]g[m_0, m_1] \quad (2.2)$$

No contexto deste trabalho, na Equação 2.2 a função g é a imagem da evolução temporal e a h é o filtro. As variáveis n_0 e n_1 são coordenadas dos pixels da imagem, enquanto m_0 e m_1 é o deslocamento do filtro sobre a imagem. A Figura 2.9 ilustra a operação. A terceira coluna da primeira linha do resultado da convolução é obtido fazendo a soma ponderada entre o filtro e a área destacada da camada da rede. Os demais valores são obtidos de forma semelhante quando o filtro se deslocou da primeira linha e primeira coluna até a terceira coluna, quando o filtro não se desloca mais. Os demais valores são obtidos da mesma forma, deslocando o filtro para as próximas linhas.

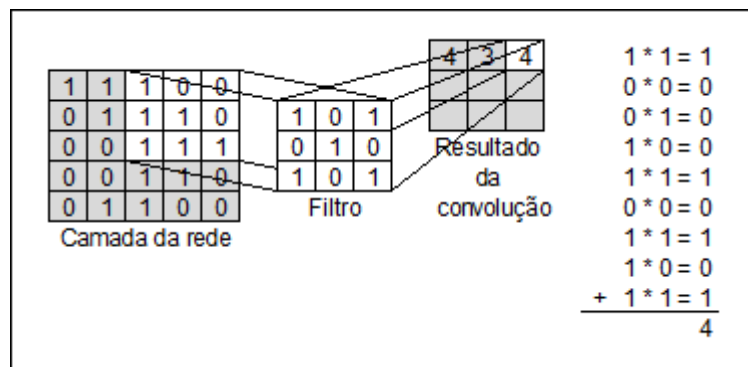


Figura 2.9: Ilustração do processo de convolução onde se tem a entrada como uma matriz binária, um filtro (que na rede feedforward faz papel de pesos sinápticos) e a saída como uma extração das características da matriz de entrada (GOODFELLOW; BENGIO; COURVILLE, 2016).

No contexto de redes neurais convolucionais o filtro $h[n_0 - m_0, n_1 - m_1]$ faz o papel dos pesos sinápticos das redes feedforward na operação de convolução bi-dimensional. Uma das diferenças entre redes feedforward e convolucionais é o fato desta última não ser totalmente conectada. Através dos filtros, grupos de neurônios pertencentes a uma camada se ligam a apenas um da próxima (HAYKIN, 1998).

Para ilustrar o processo completo de uma rede hipotética, na sequência é detalhada o funcionamento de uma rede neural convolucionacional que recebe como entrada uma evolução temporal com 10 células e 10 passos de tempo. A primeira linha é a condição inicial e as demais o estado do reticulado ao longo do tempo, organizado de cima para baixo. A Tabela 2.2a mostra a evolução temporal submetida à rede. Esta camada possui apenas um filtro, detalhado na Tabela 2.2b.

Tabela 2.2: Representação da evolução temporal (a), entrada da rede, e o filtro (b) da primeira camada convolucional.

Evolução temporal									
1	0	1	0	1	1	0	0	1	1
0	0	1	0	1	0	1	1	1	0
0	1	1	0	1	0	1	0	0	1
0	1	0	0	1	0	1	1	1	1
0	1	1	1	1	0	1	0	0	0
1	1	0	0	0	0	1	1	0	0
1	0	1	0	0	1	1	0	1	1
0	0	1	1	1	1	0	0	1	0
0	1	1	0	0	0	1	1	1	1
0	1	0	1	0	1	1	0	0	0

a

Filtro camada 1		
0.143	1.001	0.798
0.868	0.201	0.653
0.729	0.664	0.861

b

A convolução da primeira camada resulta na Tabela 2.3a e esta é a entrada da segunda camada que conta com o filtro descrito na Tabela 2.3b. Observa-se aqui a redução nas dimensões da matriz como resultado do processo de convolução.

Tabela 2.3: Resultado da primeira camada convolucional (a) e o filtro da segunda camada (b).

Resultado da 1ª convolução							
3.119	2.595	4.052	2.664	4.255	1.661	3.249	3.729
2.316	2.799	3.323	1.866	4.052	3.525	5.064	4.051
3.525	4.266	3.848	2.595	4.052	2.519	2.594	2.520
3.248	2.594	2.520	2.070	3.323	3.525	4.203	2.671
4.458	3.474	2.671	2.005	3.119	3.248	2.802	2.393
3.526	1.869	3.122	2.907	3.045	3.597	3.526	1.661
3.119	3.248	2.594	2.520	3.729	3.537	3.848	4.254
2.316	4.458	3.474	3.532	3.322	2.390	3.249	2.723

a

Filtro camada 2		
0.077	0.222	0.352
0.298	0.741	0.123
0.889	0.301	0.203

b

O resultado da segunda camada convolucional está descrito na Tabela 2.4a, onde se

pode observar a redução nas dimensões da matriz, mencionada anteriormente. O filtro desta camada está descrito na Tabela 2.4b. O processo se repete e pode ser conferido nas Tabelas 2.5 e 2.6, quando o resultado é linearizado, Tabela 2.6b, e passa por uma camada densamente conectada como nas redes feedforward, gerando o resultado da Tabela 2.6d.

Tabela 2.4: Resultado da segunda camada convolucional (a) e o filtro da terceira camada (b).

Resultado da 2ª convolução						Filtro camada 3		
10.614	11.041	10.297	9.765	11.169	10.995	-0.474	0.778	-0.923
10.834	9.537	9.202	9.926	11.139	10.745	0.055	-0.279	0.754
11.324	9.290	8.603	8.879	10.223	10.367	-0.520	0.372	0.412
10.275	7.940	8.757	9.677	10.704	10.079			
9.151	8.974	9.029	9.636	11.115	10.770			
9.547	10.786	9.882	10.740	10.812	9.728			
a						b		

Tabela 2.5: Resultado da terceira camada convolucional (a) e o filtro da quarta camada (b).

Resultado da 3ª convolução				Filtro camada 4		
10.614	11.041	10.297	9.765	-0.474	0.778	-0.923
10.834	9.537	9.202	9.926	0.055	-0.279	0.754
11.324	9.290	8.603	8.879	-0.520	0.372	0.412
10.275	7.940	8.757	9.677			
a				b		

Tabela 2.6: Resultado da quarta camada convolucional e saída da rede.

Resultado da 4ª convolução

2.051	2.878
2.287	3.255

a

Linearização

2,051	2,878	2,287	3,255
-------	-------	-------	-------

b

Camada de saída

0,234	0,200	0,324	0,243
-------	-------	-------	-------

d

Matriz de pesos

0.056	0.002	0.069	0.080
0.044	0.063	0.091	0.055
0.021	0.041	0.072	0.070
0.080	0.035	0.094	0.032

c

A Tabela 2.6a mostra o resultado da quarta camada convolucional que possui 4 células que servem como entrada para a última camada, densamente conectada. Para tal a Tabela 2.6a é transformada em 2.6b após linearização, localizando as células da matriz (c_{ij}) e a posição no vetor (v_x) fazendo $x = j + i * J$, onde $J = 4$. Então as Tabelas 2.6b e 2.6c são multiplicadas, resultando na Tabela 2.6d, que neste caso tem a saída gerada pela função softmax. Ela recebe o vetor V com K componentes que representa a saída da última camada e retorna outro vetor normalizado $V_\sigma = (\sigma(V)_0, \sigma(V)_1, \dots, \sigma(V)_K)$, segundo uma distribuição de probabilidade, conforme a Equação 2.3, onde v_i e v_j são componentes de V (GOODFELLOW; BENGIO; COURVILLE, 2016).

$$\sigma(V)_i = e^{v_i} \left(\sum_{j=1}^K e^{v_j} \right)^{-1} \quad i = 1, \dots, K \quad V = (v_1, v_2, \dots, v_K) \quad (2.3)$$

O processamento de uma rede neural artificial convolucional descrito reduz a dimensão da matriz resultante. E neste caso pode-se perder características importantes na identificação da classe referente à dinâmica apresentada pela matriz. Para se evitar isso é possível realizar uma operação de *preenchimento* que é a inclusão de pixels nulos ao redor da imagem antes da convolução, como mostra a Tabela 2.7, onde é possível observar a mesma operação da primeira camada convolucional detalhada na Tabela 2.2, mas com a evolução temporal ampliada para que o resultado da convolução mantenha a dimensão original.

Tabela 2.7: A evolução temporal original preenchida com células nulas. Usando um filtro 3×3 , o resultado da convolução mantém a dimensão da matriz original

Evolução temporal

0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	1	1	0	0	1	1	0
0	0	0	1	0	1	0	1	1	1	0	0
0	0	1	1	0	1	0	1	0	0	1	0
0	0	1	0	0	1	0	1	1	1	1	0
0	0	1	1	1	1	0	1	0	0	0	0
0	1	1	0	0	0	0	1	1	0	0	0
0	1	0	1	0	0	1	1	0	1	1	0
0	0	0	1	1	1	1	0	0	1	0	0
0	0	1	1	0	0	0	1	1	1	1	0
0	0	1	0	1	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Filtro

0.143	1.001	0.798
0.868	0.201	0.653
0.729	0.664	0.861

Resultado da convolução

0.201	2.382	0.865	3.111	1.518	2.659	2.393	2.907	2.247	1.798
1.862	3.119	2.595	4.052	2.664	4.255	1.661	3.249	3.729	2.676
1.514	2.316	2.799	3.323	1.866	4.052	3.525	5.064	4.051	1.737
2.312	3.525	4.266	3.848	2.595	4.052	2.519	2.594	2.520	2.070
2.976	3.248	2.594	2.520	2.070	3.323	3.525	4.203	2.671	1.144
2.316	4.458	3.474	2.671	2.005	3.119	3.248	2.802	2.393	1.393
2.000	3.526	1.869	3.122	2.907	3.045	3.597	3.526	1.661	1.798
1.862	3.119	3.248	2.594	2.520	3.729	3.537	3.848	4.254	3.405
1.514	2.316	4.458	3.474	3.532	3.322	2.390	3.249	2.723	1.212
1.451	2.000	2.665	0.344	1.521	1.652	2.868	2.810	1.942	1.144

Após a convolução da imagem que sofreu o processo de *preenchimento* é possível

extrair as características mais marcantes filtrando a imagem com o processo de *maxpooling*. Esta operação permite não só extrair as características mais marcantes como também a redução da dimensão da imagem. Dada uma janela de dimensão $m \times n$, será escolhido um pixel sob esta janela, com o maior valor. A Tabela 2.8 exemplifica o processo a partir de uma matriz 4×4 à esquerda, que resulta na matriz 2×2 à direita. Neste exemplo, o quadrante superior à esquerda resulta 3.119 ao passar pela operação de *maxpooling*, o mesmo acontecendo com as demais subdivisões da matriz.

Tabela 2.8: Operação maxpooling realizada em uma matriz 4×4 , extraíndo as características mais marcantes como os valores mais altos, obtendo a matriz resultante com dimensão 2×2 .

3.119	2.595	4.052	2.664			
2.316	2.799	3.323	1.866		3.119	4.052
3.525	4.266	3.848	2.595		4.266	3.848
3.248	2.594	2.520	2.070			

Além do processo maxpooling a função de ativação linear retificada (ReLU) também é uma função de ativação que processa a entrada de uma camada, resultando na ativação ou não dos neurônios se o resultado do estímulo for positiva. Este tipo de função, devido às suas características lineares, facilitam a otimização quando da utilização de algoritmos de treinamento baseado em gradientes. A Figura 2.10 descreve o comportamento da função conforme a entrada recebida.

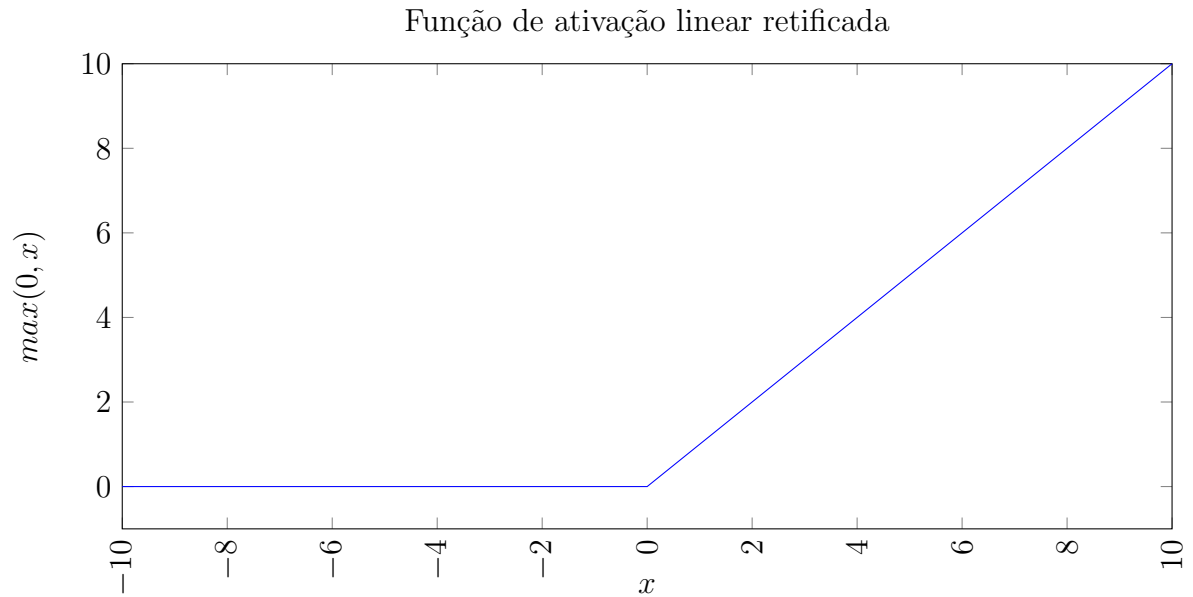


Figura 2.10: A função de ativação linear retificada possui comportamento quase linear, ela mantém muitas das propriedades que facilitam a otimização dos modelos lineares com métodos baseados em gradiente. Eles também preservam muitas das propriedades que fazem os modelos lineares generalizarem bem (GOODFELLOW, 2016).

A rede neural desenvolvida para a construção do classificador será abordada no Capítulo 4. A próxima seção aborda a técnica utilizada para a construção do segundo classificador, análise de textura, e a seguinte define como ambos os classificadores serão avaliados.

2.4 Análise de textura

Em uma imagem há variações de cores e intensidade que, ao se repetirem, formam as texturas que possuem atributos locais com estatísticas que variam lentamente ao longo da imagem ou permanecem parcialmente periódicas. A própria textura é formada por sub-padrões, com regras específicas de posicionamento na imagem, e estes sub-padrões são formados por unidades fundamentais denominadas primitivas, caracterizadas por formas geométricas ou padrões locais de pixels. As texturas fornecem informações importantes sobre o arranjo dos elementos da imagem e também caracterizam a superfície de objetos, possibilitando sua classificação, que se torna mais simples se as texturas são bem diferen-

ciadas. Por ser um elemento visual, uma textura pode ser descrita em termos qualitativos, mas a análise de texturas procura quantificar a descrição qualitativa, caracterizando mais precisamente a textura, permitindo usar sua análise em áreas variadas, desde avaliação de imagens médicas até detecção automática de elementos na imagem (ACHARYA; RAY, 2005).

Como definido na Seção 2.1, a imagem que representa a evolução temporal é obtida substituindo-se os estados por cores, que no caso dos ACs são preto e branco. Este modo de ver a evolução temporal permite usar técnicas estabelecidas para extrair características de imagens. Em He e Wang (1990), a análise de textura foi usada para se classificar imagens de satélites com 256 níveis de cinza, utilizando algoritmo denominado *texture spectrum* para extrair características da imagem. Lá foi levado em consideração uma vizinhança $V = \{V_0, V_1, \dots, V_8\}$, onde V_i é uma das 8 células em torno de uma célula central, sendo V_0 a superior à esquerda e as demais obtidas variando-se em sentido horário pela vizinhança. A unidade de textura $TU = \{E_1, E_2, \dots, E_8\}$ é obtida através da Equação 2.4, onde cada E_i corresponde ao valor transformado do pixel V_i .

$$E_i = \begin{cases} 0, & V_i < V_0 \\ 1, & V_i = V_0 \\ 2, & V_i > V_0 \end{cases} \quad (2.4)$$

Dados os 3 valores possíveis para a transformação e as 8 células da vizinhança, tem-se $3^8 = 6561$ configurações possíveis. A imagem transformada é obtida na forma de matriz, cuja dimensão é a mesma da imagem, criando correspondência entre células da matriz e pixels da imagem, Cada célula contém o valor da unidade de textura TU da vizinhança do pixel correspondente, valor este obtido segundo a Equação 2.5 (HE; WANG, 1990).

$$N_{TU} = \sum_{i=1}^8 E_i \cdot 3^{i-1} \quad (2.5)$$

Exemplificando o algoritmo, assume-se que a vizinhança de um pixel tenha a seguinte configuração $V = \{65, 253, 173, 65, 69, 56, 228, 55, 215\}$, que é convertida para a unidade de textura $TU = \{2, 2, 1, 2, 0, 2, 0, 2\}$, cujo valor decimal referente ao pixel central é 4931, conforme a Equação 2.5. Fazendo o mesmo para cada pixel é obtida uma transformação da imagem onde a intensidade de cada pixel varia de 0 a $3^8 - 1$, ainda conforme a Equação

2.5. Sendo assim, é possível construir um vetor de frequências relativas para cada valor possível N_{TU} e associar este vetor à imagem conhecida para a classificação (HE; WANG, 1990).

Mas existem outras formas de se extrair características texturais de uma imagem. Em abordagens estatísticas da análise de textura, o ponto inicial é a extração de características da imagem, informações extraídas da vizinhança ao redor do pixel. Um algoritmo, semelhante ao utilizado por He e Wang (1990), que se propõe a extrair informações sobre as texturas da imagem é o algoritmo **Local Binary Pattern** (LBP), que permite extrair informações sobre mudanças na intensidade de brilho ao redor de um pixel central, comparando os pixels em sentido horário, conforme a Equação 2.6, onde c_j^t é o pixel central e c_x^y os pixels da vizinhança (MACHICAO et al., 2018).

$$s(c_j^t - c_x^y) = \begin{cases} 0, & c_j^t - c_x^y \geq 1 \\ 1, & c_j^t - c_x^y < 1 \end{cases} \quad (2.6)$$

A Equação 2.6 destaca os pixels que são mais intensos que o pixel central, transformando a vizinhança em um padrão binário que será atribuído ao pixel central na imagem transformada, resumindo assim em um único valor a diferença de intensidade do pixel central e os pixels da vizinhança.

Figura 2.11: Aplicação do algoritmo LBP em torno de um pixel com intensidade 39 em tom de cinza.



(a)

155	239	255
101	52	0
255	192	155

(b)

-103	-187	-203
-49		52
-203	-140	-103

(c)

1	1	1
1		0
1	1	1

(d)

Exemplificando o algoritmo, a Figura 2.11(a) mostra vizinhança de um pixel central de uma imagem em tons de cinza, enquanto a Figura 2.11(b) mostra os valores numéricos dos pixels. A Figura 2.11(c) mostra as diferenças de intensidade dos pixels. E a Figura 2.11(d) mostra o padrão binário resultante da aplicação da Equação 2.6. No sentido horário, começando pelo canto superior esquerdo, obtém-se o valor $11110111_2 = 247_{10}$ que substituirá o valor do pixel central na imagem transformada.

A imagem transformada pode apresentar vários pixels com tom de cinza cujo valor se traduziu em 247_{10} , no exemplo da Figura 2.11. Este fato propicia a construção de histograma que caracterize a imagem original, por meio da Equação 2.7, onde k é o tom de cinza contabilizado. Se o raio valer 1, k poderá variar entre 0 e 2^8 , que perfazem 255 níveis de cinza. E $q(x, y) = 1$ se $x = y$ ou 0 caso contrário.

$$H(k) = \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} q(LBP_{p,r}(i, j), k) \quad (2.7)$$

O algoritmo LBP foi pensado para extrair um padrão binário de uma vizinhança cuja imagem seja em tons de cinza ou mesmo colorida, neste caso cada pixel é pensado em um

número de 24 bits, variando no intervalo $[0, 2^{24} - 1]$. No caso de a imagem já ser binária, há uma observação pertinente sobre a Equação 2.6. Se o pixel central for 1, a configuração da vizinhança não se altera. Mas se o pixel central for 0, a vizinhança conterá apenas o valor 1.

Transformar a evolução temporal em um espectro de frequência utilizando a **transformada de Fourier** (Equação 2.8) também permite a extração de características, agora por meio de uma imagem espectral da evolução temporal.

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-\frac{xu}{M} 2\pi j} e^{-\frac{yv}{N} 2\pi j} \quad (2.8)$$

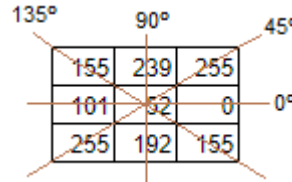
Na Equação 2.8, x e y são as coordenadas de um pixel da imagem, $f(x, y)$ é a função que retorna a intensidade do pixel, M e N são as dimensões da imagem a ser transformada. A imagem espectral que se pretende obter é representada pela função $F(u, v)$.

A **Matriz de Co-ocorrência de Nível de Cinza** é outro algoritmo usado na extração de características das texturas, baseado na variação dos níveis de cinza da imagem. A princípio o algoritmo se aplica somente a imagens que contenham apenas tons de cinza. Porém, mesmo imagens coloridas podem ser tratadas como imagens em tons de cinza. Utilizando as 3 cores básicas vermelho, verde e azul codifica-se um pixel colorido como a junção destas 3 cores através da codificação RGB (Read-Green-Blue), cada pixel contendo uma intensidade para cada cor. A intensidade na codificação RGB variando de 0 a 255 para cada cor é possível escrever (123, 201, 45) para indicar que a intensidade de vermelho é 123, de verde é 201 e 45 para azul. Pode-se usar o termo canal para indicar as três cores básicas para uma imagem colorida, sendo possível separar uma imagem colorida em 3 imagens tons de cinza, uma para cada canal.

Uma matriz de co-ocorrência é uma matriz quadrada com as dimensões dos níveis de cinza possível. Se a variação vai de 0 a 255, a matriz possui dimensão 256×256 . Cada célula (i, j) da matriz cujo valor seja p , indica a proporção de ocorrências de mudanças no tom de cinza i para j . Mudança esta medida a partir de um pixel central com determinado ângulo e distância até outro pixel. Isso significa que o algoritmo resulta em mais de uma matriz, a depender o ângulo considerado. A Figura 2.12 exemplifica a extração da variância do tom de cinza entre o pixel central de intensidade 52 e os que distam uma

unidade, em ambos os sentidos, com ângulos de 0°, 45°, 90° e 135°, resultando assim em 4 matrizes de co-ocorrência, uma para cada ângulo.

Figura 2.12: Extração de informação de variância do tom de cinza com relação ao pixel central e aqueles que distam uma unidade com ângulos 0°, 45°, 90° e 135°.



A equação 2.9 descreve como obter a proporção de ocorrência da mudança de intensidade i para j dado um ângulo, que se traduz nas variações Δx e Δy . Nela, W significa é o inverso da quantidade de pixels da imagem, desconsiderando as variações Δx e Δy . A função Q , Equação 2.11, conta a variação de i para j , definido pela Equação 2.12 (ALBREGTSEN, 2008).

$$P(i, j|\Delta x, \Delta y) = W \times Q(i, j|\Delta x, \Delta y) \quad (2.9)$$

$$W = \frac{1}{(M - \Delta x)(N - \Delta y)} \quad (2.10)$$

$$Q(i, j|\Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A \quad (2.11)$$

$$A = \begin{cases} 1, & f(x, y) = i \text{ e } f(x + \Delta x, y + \Delta y) = j \\ 0, & \text{caso contrário} \end{cases} \quad (2.12)$$

A partir dessas informações sobre a textura se pode analisar a imagem com foco em vários objetivos, onde a classificação é um deles. Na Seção 5.1 é apresentada a forma como se propõe extrair informações das texturas contidas nas imagens binárias, as evoluções temporais. E como usar estas informações na classificação dos ACs.

2.4.1 Algoritmo k -NN

O algoritmo k -Nearest Neighborhood (k -NN) é um classificador bem conhecido e de fácil aplicação. Envolve definir a distância entre dois objetos, sendo um que se busca classificar e outro pertencente a um conjunto $C \in \mathbb{R}^d$ de objetos usados como referência, uma vez que cada um deles possui rótulos a definir sua classe. A menor distância encontrada entre os dois objetos associa a classe que se pretende descobrir, àquela conhecida do conjunto de referência. A maior dificuldade do algoritmo é identificar qual medida de distância é a mais adequada para o problema de classificação que se pretende resolver (MUCHERINO; PAPAJORGJI; PARDALOS, 2009).

Sendo $d(\vec{a}, \vec{b})$ a medida de distância entre os vetores \vec{a} e $\vec{b} \in \mathbb{R}^d$, não há restrições sobre qual medida de distância pode ser usada no algoritmo, mas qualquer que seja $d(\vec{a}, \vec{b})$, deve satisfazer as Equações 2.13, 2.14 e 2.15, que caracterizam qualquer medida de distância. O parâmetro k do algoritmo refere-se a quantos objetos de C , mais próximos do objeto que se pretende classificar, devem ser considerados para se definir a classificação, sendo esta proximidade definida pela medida de distância adotada (AGGARWAL, 2014).

$$d(\vec{a}, \vec{b}) \geq 0 \tag{2.13}$$

$$d(\vec{a}, \vec{b}) = d(\vec{b}, \vec{a}) \tag{2.14}$$

$$d(\vec{a}, \vec{b}) = 0 \Leftrightarrow \vec{a} = \vec{b} \tag{2.15}$$

As informações que se podem extrair de imagens utilizando análise de textura, abordada na Seção 5.2, organizadas na forma de vetores, podem ser utilizadas na classificação destas mesmas imagens caso se tenha um conjunto C de referência, com informações sobre imagens cujas classes são previamente conhecidas. Nestas condições a aplicação do algoritmo k -NN para determinar a classe de uma imagem é aderente, caso seja definida uma medida de distância apropriada como a Euclidiana (Equação 2.16).

$$d(\vec{a}, \vec{b}) = \sqrt{\sum_n^{i=0} (a_i - b_i)^2} \quad (2.16)$$

Pela quantidade de evoluções temporais necessárias para se criar a base de conhecimento necessária à classificação, decidiu-se pela distância Euclidiana na comparação entre os vetores de características, por ser esta de simples implementação e eficiente do ponto de vista computacional, com $O(n)$ de complexidade para se calcular a distância.

2.5 Matriz de confusão

Antes de descrever o uso da matriz de confusão é importante detalhar como os classificadores decidem qual a classe predita. Devido à variação possível na dinâmica apresentada pelas evoluções temporais geradas por uma regra, para definir sua classe não basta classificar o comportamento dinâmico de uma única evolução, mas sim de um conjunto. O comportamento mais frequente, o comportamento típico, define a classe da regra. Se houver uma distribuição igualitária entre as 4 classes de Wolfram (1984), cada uma apresentará 25% de frequência no conjunto. Qualquer valor acima disso define a classe mais frequente, logo, será assumida como a classe predita para a regra.

Como será visto nos próximos capítulos, ambos os classificadores predisseram corretamente as classes de todas as regras do espaço elementar e apresentaram fragilidades ao serem aplicados no espaço maior. No entanto, tal resultado não auxilia na avaliação da performance de cada um. A **matriz de confusão** auxilia na avaliação da qualidade do classificador. A partir desta matriz é possível fazer uma contagem das predições que são **verdadeiro positivo** (VP), aquelas onde há uma correspondência entre a classe predita e a real. As **verdadeiro negativas** (VN), são aquelas corretamente preditas como não sendo de uma classe; **falso positivo** (FP) diz respeito às predições feitas para uma classe distinta da real, enquanto as **falso negativas** (FN) são aquelas preditas erroneamente como não sendo de uma classe que seria a predição correta.

A matriz de confusão C tem uma dimensão 4×4 , onde as predições estão em colunas enquanto nas linhas, as classes reais. A contagem das VP's é feita para cada classe observando os valores da diagonal principal, enquanto para as VN's soma-se aqueles fora

da linha e coluna correspondente à classe. Para os FP's conta-se valores referentes à coluna, mas que não estejam na mesma linha da classe, invertendo esta relação para os FN's. A Tabela 2.9 resume como calcular tais variáveis, onde x representa a classe que se deseja medir. Por exemplo, o valor Verdadeiro Positivo de uma classe x é o valor c da linha e coluna correspondente a classe (c_{xx}). Enquanto o valor Verdadeiro Positivo de uma classe é a contagem de toda a matriz excetuando a linha e coluna correspondente à classe.

Tabela 2.9: Matriz de Confusão - Contagem das predições.

Verdadeiro Positivo	$VP_x = c_{xx}$
Verdadeiro Negativo	$VN_x = \sum_{i=1}^4 \sum_{j=1}^4 c_{ij}, se\ i \neq x\ e\ j \neq x$
Falso Positivo	$FP_x = \sum_{i=1}^4 \sum_{j=1}^4 c_{ij}, se\ i = x\ e\ j \neq x$
Falso Negativo	$FN_x = \sum_{i=1}^4 \sum_{j=1}^4 c_{ij}, se\ i \neq x\ e\ j = x$

A partir destas contagens calcula-se a **acurácia global** (AG) do classificador, o nível de acertos no universo de casos avaliados, que é a proporção entre a diagonal principal e toda a matriz. Com relação às classes, foi calculada a **taxa de verdadeiros positivos** (TVP), também conhecida como *sensibilidade*, sendo a proporção de verdadeiros positivos corretamente preditos; a **taxa de verdadeiros negativos** (TVN), também conhecida como *especificidade*, que mede os verdadeiros negativos corretamente preditos; o **valor preditivo positivo** (VPP), também conhecida como *precisão*, que é a taxa de confirmações dentre todas as predições positivas; e o **valor preditivo negativo** (VPN), que é a taxa de negação por todos os itens preditos como não sendo da classe. Já a **acurácia** (AC) em classificar uma das classes, indica quanto o classificador acerta na predição desta. Estas taxas são individuais para cada classe, e quanto mais próximas

de 1 melhor (FAWCETT, 2006) (METZ, 1978), ou seja, deseja-se maximizá-las. Todas essas grandezas mencionadas estão resumidos na Tabela 2.10 que indica como calcular tais métricas, onde x representa a classe que se deseja medir.

Tabela 2.10: Matriz de Confusão - Índices maximizáveis.

Acurácia Global	$AG = \sum_{i=0}^4 c_{ii} \times (\sum C)^{-1}$
Taxa de verdadeiros positivos	$TVP_x = \frac{VP_x}{P_x} = \frac{VP_x}{VP_x + FN_x} = 1 - TFN_x$
Taxa de verdadeiros negativos	$TVN_x = \frac{VN_x}{N} = \frac{VN_x}{VN_x + FP_x} = 1 - TFP_x$
Valor preditivo positivo	$VPP_x = \frac{VP_x}{VP_x + FP_x}$
Valor preditivo negativo	$VPN_x = \frac{VN_x}{VN_x + FN_x}$
Acurácia	$AC_x = \frac{VP_x + VN_x}{P_x + N_x} = \frac{VP_x + VN_x}{VP_x + FN_x + VN_x + FP_x}$

Em contraste, há ainda um conjunto de grandezas que se deseja minimizar. Entre elas, a **taxa de falsos positivos** (TFP) é a proporção entre as ocorrências classificadas como positivas erroneamente e as realmente negativas, ou seja, pode-se dizer que é a proporção de alarmes falsos que o classificador pode emitir. De forma semelhante a **taxa de falsos negativos** (TFN) é a taxa de classificações negativas feitas erroneamente, é a proporção de falsos negativos e as ocorrências positivas da classe na população, ela é a proporção com que o classificador comete equívocos e não aponta uma ocorrência verdadeira. A **taxa de falsas descobertas** (TFD) é a proporção com que o classificador aponta erroneamente uma descoberta, ou seja, a proporção entre os falsos positivos e todas as ocorrências

preditas como positivo, quer sejam ou não. A **taxa de falsas omissões** (TFO) é a taxa com que o classificador omite uma ocorrência positiva, é a proporção entre os falsos negativos e todas as ocorrências preditas como negativas. A precisão do classificador é indicada quanto mais próximo a 0 estas taxas estiverem (BANERJEE; BHADURY, 2009) (STOREY, 2002)(PATIL; NEMADE; SONI, 2018). A Tabela 2.11 resume como calcular tais variáveis, onde x representa a classe que se deseja medir.

Tabela 2.11: Matriz de Confusão - Índices minimizáveis.

Taxa de falsos positivos	$TFP_x = \frac{FP_x}{N_x} = \frac{FP_x}{FP_x + VN_x} = 1 - TVN_x$
Taxa de falsos negativos	$TFN_x = \frac{FN_x}{P_x} = \frac{FN_x}{FN_x + VP_x} = 1 - TVP_x$
Taxa de falsas descobertas	$TFD_x = \frac{VP_x}{FP_x + VP_x} = 1 - VPP_x$
Taxa de falsas omissões	$TFO_x = \frac{FN_x}{FN_x + VN_x} = 1 - VPN_x$

A matriz de confusão e as métricas derivadas permitem definir a qualidade global e particular para cada classe do classificador. Elas foram usadas nos Capítulos 4 e 5, gerando informações sobre os métodos de classificação. Antes, os Capítulos 4 e 5 detalham como cada método de classificação foi aplicado.

Capítulo 3

Entropia e classificação de autômatos celulares

Antes de apresentar os dois classificadores desenvolvidos, será apresentado o resultado da classificação do espaço elementar segundo o trabalho pioneiro de Wuensche (1998). A adaptação de sua proposta passou pela definição da quantidade mínima de condições iniciais necessárias para se observar o comportamento típico das regras do espaço elementar, definida empiricamente no presente trabalho em 500, geradas de forma aleatória no momento de se avaliar cada regra, pois foi observado que conjuntos aleatórios e diferentes de condições iniciais para cada regra oferece melhores resultados no aprendizado dos padrões. Foram usadas condições iniciais padronizadas contendo 100 células, executando a regra avaliada por 300 passos com transiente de 200, resultando em evoluções temporais, matriz binária de dimensões 100×100 células.

A definição do comportamento típico de cada evolução temporal permite definir a frequência observada para cada uma das classes de Wolfram (2002), sendo que a proporção mais alta entre os 4 valores, prediz a classe da regra geradora. Por ter sua classificação conhecida, o espaço elementar foi usado como referência para definir a correta classificação das regras. No entanto, o espaço de raio 1,5 não possui classificação prévia que permita avaliar a acurácia de um classificador neste espaço. Por isso foi feita uma classificação visual do espaço de raio 1,5 que passou a servir de referência ao avaliar a acurácia neste caso. Pela avaliação visual, 3,83% das regras são classe 1, 72,31% classe 2, 16,44% classe 3 e 7,42% classe 4. A classificação completa está disponível em <https://bit.ly/2VwVnhV>.

Wuensche (1998) definiu 3 classes possíveis em seu trabalho e propõe o uso da variação média da entropia ao longo da evolução temporal para determinar seu comportamento dinâmico, definindo-o como uma das classes *ordenada*, *complexa* e *caótica*. A Equação 3.1 define a entropia utilizada. Nela p_i é a frequência, proporção ou probabilidade de ocorrência de um determinado símbolo entre I símbolos possíveis. A entropia H é calculada fazendo a soma dos p_i com seu logaritmo na base dois.

$$H = \sum_{i=0}^{I-1} p_i \log_2(p_i) \quad (3.1)$$

Contudo, as imagens que se pretende classificar são binárias, havendo apenas 2 símbolos possíveis, 0 e 1. Assim, pode-se expandir a Equação 3.1 na Equação 3.2, que descreve como calcular a entropia utilizando a proporção de zeros (p_0) e uns (p_1) do reticulado.

$$H = p_0 \log_2(p_0) + p_1 \log_2(p_1) \quad (3.2)$$

A variação da entropia do reticulado ao longo do tempo para as classes ordenada, complexa e caótica, está destacada na Figura 3.1. O padrão de variação observada para cada uma é utilizado para classificar a regra geradora. As classes 1 e 2 do esquema de Wolfram (1984) são equivalentes à classe ordenada, enquanto as classes 3 e 4 são respectivamente caótica e complexa. A Figura 3.2 apresenta a variação de entropia ao longo do tempo para as regras 168, 79, 60 e 124 do ACE, respectivamente regras classe 1, 2, 3 e 4. A variação apresentada é compatível com as apresentadas por Wuensche (1998) para as regras 168 e 79. Porém, não para as regras 60 e 124 que apresentam comportamento muito semelhante, indicando certa dificuldade em distingui-las.

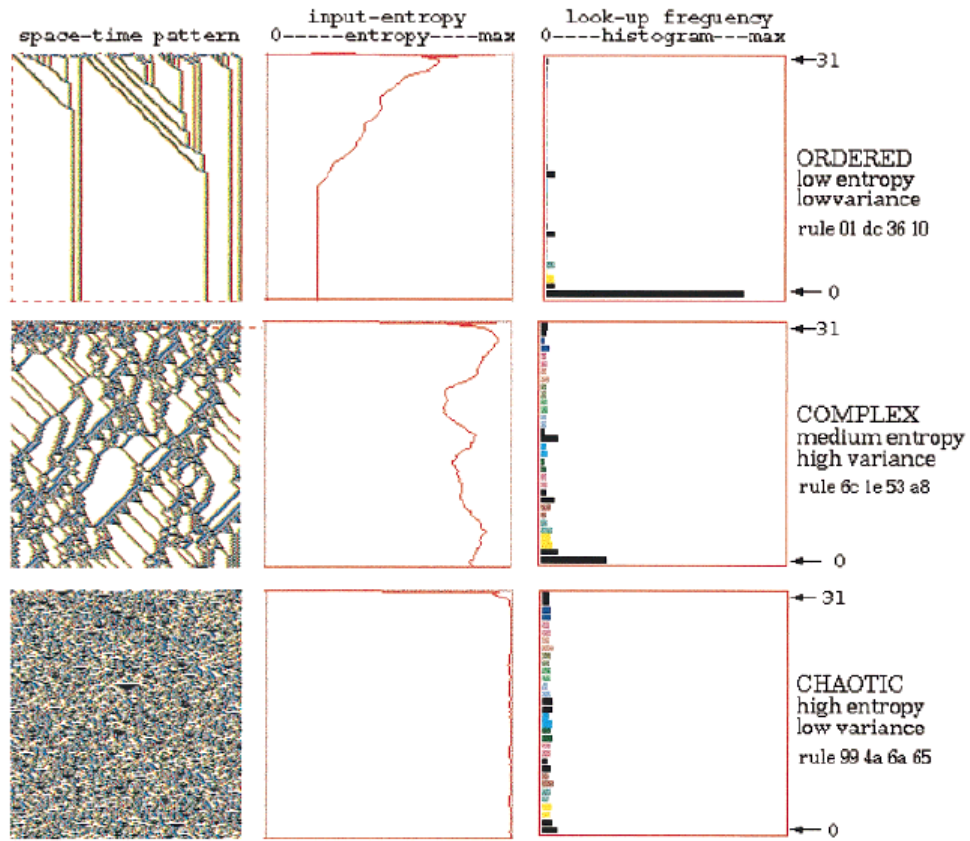


Figura 3.1: Associação do padrões visuais observados nas evoluções temporais (primeira coluna), a entropia do sistema em cada momento (segunda coluna) e o histograma das configurações observadas nas evoluções temporais. Observa-se que apenas o comportamento complexo possui alta variância (WUENSCHÉ, 1998).

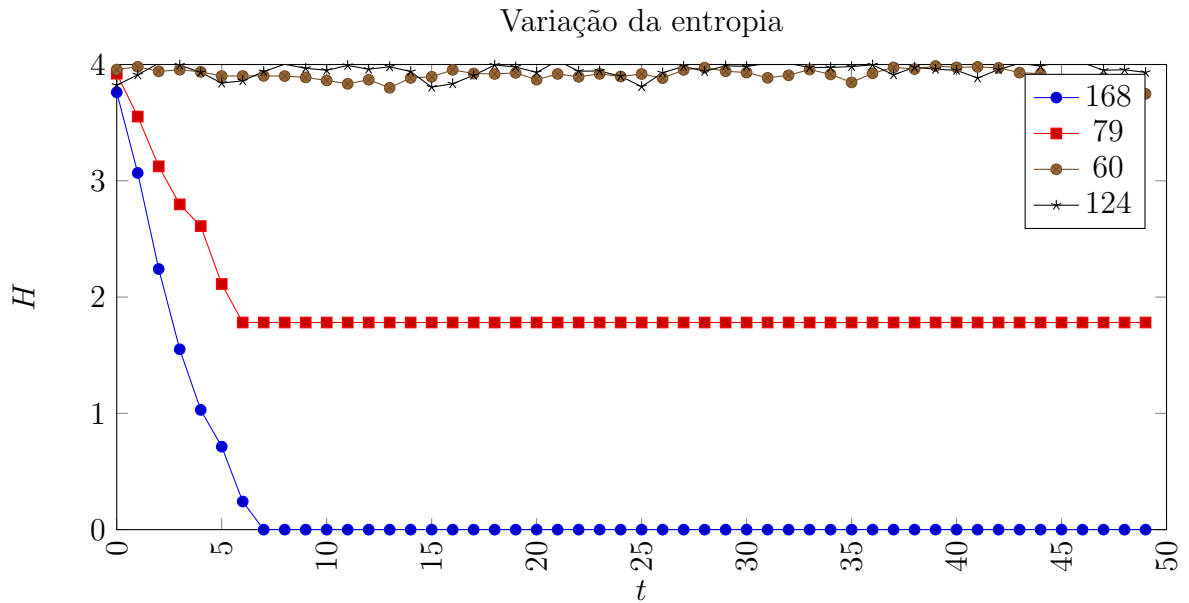


Figura 3.2: Variação da entropia ao longo de 50 iterações.

Wuensche (1998) não oferece uma avaliação da acurácia obtida ao classificar o espaço elementar usando a variação da entropia. Este capítulo apresenta uma adaptação da classificação proposta em seu trabalho, onde a variação da entropia é usada para criar um conjunto de dados a ser usado na classificação.

Dado o conjunto mínimo de 500 evoluções temporais e a variação da entropia de cada uma, foi criado um conjunto com a variação média da entropia para cada regra, servindo este como base de conhecimento a ser usado na classificação do espaço elementar e do raio 1,5. Este conjunto de dados possui vetores de 100 componentes, onde a entropia de cada momento discreto das evoluções temporais corresponde a um componente do vetor. A ele estão associadas a regra, classe e quantidade de evoluções temporais utilizadas, formando a tupla $e = (n, c, q, \vec{H})$, onde n é a regra, c a classe, q a quantidade de evoluções temporais usadas e \vec{H} é o vetor contendo a variação da entropia ao longo do tempo.

A variação média da entropia, para cada regra, ao longo das 500 evoluções temporais, definidas como mínimas para se observar o comportamento típico, serviu para a classificação que utilizou o algoritmo k -NN na predição, com $k = 1$ neste caso. O vetor de cada uma das 500 evoluções temporais é comparado com os vetores da base de conhecimento, sendo o vetor com a menor distância o escolhido.

3.1 Resultados da classificação utilizando entropia

A classificação é feita sobre um conjunto de evoluções temporais, assim, a Tabela 3.1 resume o resultado da classificação do espaço elementar. Ela indica a dificuldade do modelo em classificar corretamente regras classe 4 deste espaço. Pois, todas elas e mais uma regra classe 2 foram classificadas como classe 3, que ficou com 38 regras preditas, quando o esperado era 23.

Tabela 3.1: Resultado da classificação do espaço elementar usando a variação da entropia das evoluções temporais.

Classes	Predições	Real	Percentual
1	24	24	100,00%
2	194	195	99,49%
3	38	26	146,15%
4	0	11	0,00%

A Tabela 3.2 resume o resultado da avaliação da qualidade do classificador. Em comparação com os resultados a serem apresentados nas próximas seções, a acurácia obtida para o espaço elementar foi mais baixa, apesar de ter superado os 90%. Ainda assim, foi uma má classificação para as classes 3 e 4, podendo se justificar a alta acurácia pela quantidade reduzida de regras classe 4 neste espaço.

Tabela 3.2: Matriz de confusão proveniente da classificação por entropia do espaço elementar.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	12000	200	0	0
2	0	94480	1970	1670
3	0	2110	8750	2950
4	0	710	2280	880
Acurácia global: 0.9071				

A Tabela 3.3 mostra os detalhes da classificação obtida usando a entropia. Observa-se

que a variação da entropia apresenta boa capacidade de classificação do ACE, apesar da deficiência em algumas de suas métricas, como a apresentada pela acurácia da classe 3, de 0.9322 e a taxa de verdadeiros positivos para a classe 4 em 0,16, valor distante do desejado que é 1,00.

Tabela 3.3: Métricas da matriz de confusão proveniente da classificação por entropia do espaço elementar.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	12000	94480	8750	880
Verdadeiro negativo	116000	33520	119250	127120
Falso positivo	200	3640	5060	2990
Falso negativo	0	3020	4250	4620
Taxa de verdadeiros positivos	1,0000	0,9690	0,6731	0,1600
Taxa de verdadeiros negativos	0,9983	0,9020	0,9593	0,9770
Valor preditivo positivo	0,9836	0,9629	0,6336	0,2274
Valor preditivo negativo	1,0000	0,9174	0,9656	0,9649
Acurácia	0,9984	0,9505	0,9322	0,9439
Taxa de falsos positivos	0,0017	0,0980	0,0407	0,0230
Taxa de falsos negativos	0,0000	0,0310	0,3269	0,8400
Taxa de falsas descobertas	0,0164	0,0371	0,3664	0,7726
Taxa de falsas omissões	0,0000	0,0826	0,0344	0,0351

Foram observadas falhas na classificação do ACE, pois o comportamento típico de algumas regras não fora definido corretamente. A classe com maior frequência para a regra 26 foi a 3, predita em 46% das vezes, sendo ela da classe 2. Já as regras 54 e 106 foram classificadas como classes 2 e 3 em 42% e 68% das vezes, respectivamente. porém ambas são classe 4. Os valores apresentados nas Tabelas 3.2 e 3.3 indicam boa capacidade da variação de entropia para classificar o ACE, porém apresenta limitações importantes.

Já para o espaço de raio 1,5, a classificação via entropia apresenta diminuição drástica da acurácia ao se classificar um espaço maior do que o utilizado para o treinamento. Ainda assim, a acurácia de 0,5321 confirma a capacidade de extrair informações de um espaço

desconhecido utilizando a métrica.

Tabela 3.4: Matriz de confusão proveniente da classificação por entropia do espaço de raio 1,5.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	776248	4398333	38797	44910
2	35766	13457133	1168160	623686
3	960	5245507	2917926	932396
4	354	1990934	851064	285826
Acurácia global: 0,5321				

As métricas da Tabela 3.5 demonstram classificador a possibilidade de usar a entropia para se obter uma primeira classificação do espaço raio 1,5. A acurácia está acima de 0,70, chegando a 0,8806 para a classe 4 e outras métricas apresentam valores próximos do desejado. Porém, outras métricas apresentam valores longe do desejado. A taxa de verdadeiros positivos de 0,1515 e falsa descoberta de 0,9086 para classe 4 são exemplos. Em conjunto, as métricas indicam a possibilidade de se extrair alguma informação do espaço de regras, mas com pouca confiabilidade, resumida na acurácia global.

Tabela 3.5: Métricas da matriz de confusão proveniente da classificação por entropia do espaço elementar.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	776248	13457133	2917926	285826
Verdadeiro negativo	31991752	19310867	29850074	32482174
Falso positivo	4482040	1827612	6178863	2842352
Falso negativo	37080	11634774	2058021	1600992
Taxa de verdadeiros positivos	0,9544	0,5363	0,5864	0,1515
Taxa de verdadeiros negativos	0,8771	0,9135	0,8285	0,9195
Valor preditivo positivo	0,1476	0,8804	0,3208	0,0914
Valor preditivo negativo	0,9988	0,6240	0,9355	0,9530
Acurácia	0,8788	0,7088	0,7991	0,8806
Taxa de falsos positivos	0,1229	0,0865	0,1715	0,0805
Taxa de falsos negativos	0,0456	0,4637	0,4136	0,8485
Taxa de falsas descobertas	0,8524	0,1196	0,6792	0,9086
Taxa de falsas omissões	0,0012	0,3760	0,0645	0,0470

Os Capítulos 4 e 5 que se seguem detalham os resultados obtidos utilizando rede neural convolucional e análise de textura na classificação do espaço elementar e do espaço de raio 1,5, e detalha a acurácia de cada método, tanto global quanto específico para cada classe.

Capítulo 4

Redes Neurais e classificação de autômatos celulares

Na Seção 2.3 foi descrita a capacidade das redes neurais convolucionais em extrair características da imagem automaticamente, dispensando algum pré-processamento sobre ela. Ao encarar evoluções temporais como imagens binárias, possibilita-se o uso destas redes na classificação automática de autômatos celulares.

As evoluções temporais usadas para treinar a rede neural possuem a mesma dimensão usada no Capítulo 3. Porém, as quantidades envolvidas no treino da rede neural é diferente, pois submete-se à rede tantas evoluções temporais quantas necessárias até que a resposta da rede apresente acurácia esperada. A próxima seção detalha a arquitetura da rede desenvolvida neste trabalho e seu treinamento, enquanto as próximas descrevem os resultados obtidos.

4.1 Proposta para a classificação de autômatos celulares binários unidimensionais usando redes neurais

O aprendizado profundo é uma realidade nas pesquisas de várias áreas, hoje em dia. Destaca-se neste campo o uso das redes neurais convolucionais como ferramenta de

pesquisa, obtendo bons resultados na classificação de imagens. Em Panigrahi et al. (2018) foi feita uma avaliação da performance de uma rede convolucional usada para imagens como sendo de cachorros e gatos, alcançando acurácia de 88.31%. A rede lá utilizada é semelhante à utilizada aqui, descrita na Figura 4.1, inclusive com a mesma função de ativação e aplicação do processo de maxpooling, conforme detalhado na Seção 2.3.1, para a redução da dimensão da imagem. Vale destacar a inexistência, até o momento, de algoritmo ou meio analítico para a definição da arquitetura de redes neurais em geral para a solução de um problema específico. Por isso a escolha de replicar aqui, com as devidas adaptações, a estrutura encontrada na literatura ao usar redes convolucionais

As evoluções temporais usadas como entrada da rede possuem dimensão de 100 linhas por 100 colunas, resultando em imagem binária 100×100 pixels e um canal. A Figura 4.1 mostra um diagrama de blocos que representa a rede desenvolvida para o classificador e detalha as camadas, seus tipos e dimensões usadas. A estratégia adotada é que cada camada convolucional mantenha a dimensão da entrada fornecida (processo de preenchimento) e tenha a função retificadora (ReLU) na ativação do neurônio (GOODFELLOW; BENGIO; COURVILLE, 2016). Para a última camada interna e a de saída adotaram-se camadas densamente conectadas, sendo que a primeira também usa a função ReLU na ativação, enquanto para a saída é usada softmax. Para evitar sobre-treino, as ligações entre a última camada interna e a saída adotou-se dropout de 30% (SRIVASTAVA et al., 2014).

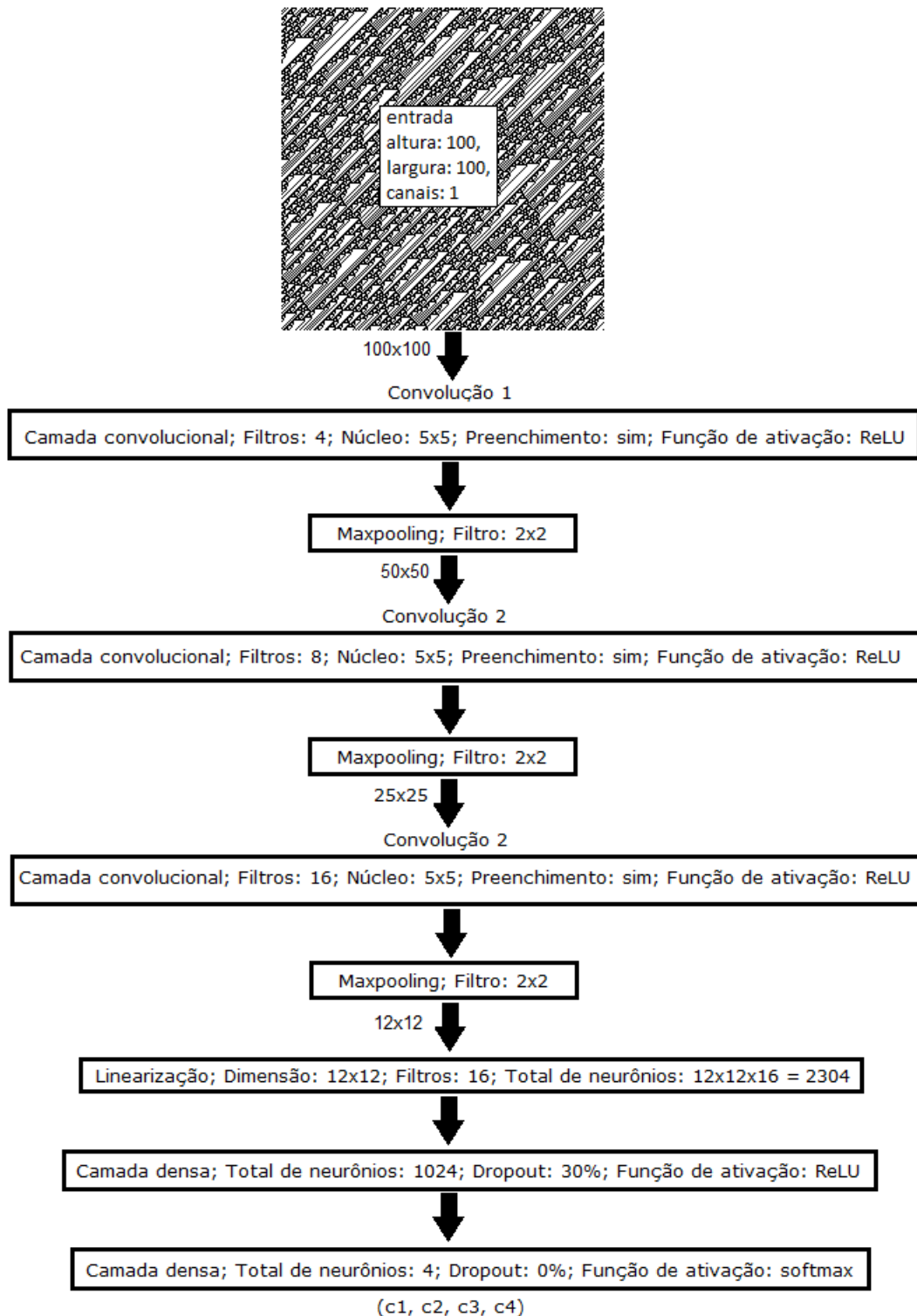


Figura 4.1: Diagrama de blocos que representa o processamento e descreve a arquitetura do classificador, uma rede neural convolucional, utilizada no trabalho. Nele é destacado a redução da dimensão da imagem até aos 4 neurônios de saída (c1, c2, c3, c4), representando a probabilidade de a regra geradora da evolução temporal ser de cada uma das 4 classes possíveis.

O processamento da rede começa com a submissão da imagem binária com apenas um canal, significando que a entrada possui uma única matriz de valores, totalizando 10000 neurônios, que após a primeira camada convolucional com 4 filtros, se obtém 4 matrizes 100×100 . Suas dimensões são então reduzidas pelo processo maxpooling com janela 2×2 , resultando em 4 matrizes 50×50 .

A segunda camada convolucional recebe as 4 matrizes 50×50 da anterior e com seus 8 filtros resulta na saída com 32 matrizes 50×50 , passando pelo maxpooling, reduzindo as dimensões para 25×25 . O mesmo processo se repete para a terceira camada convolucional que resulta em 512 matrizes 25×25 , reduzindo a dimensão para 12×12 .

Esta é a última camada convolucional e tem a dimensão de sua saída reduzida e lineariza, com total de 73.728 neurônios e ligada a uma camada densamente conectada, possuindo um total de 1024 neurônios. Esta última camada interna se liga à de saída com 4 neurônios, sofrendo dropout de 30% em suas ligações. A ativação da camada de saída é a função softmax que oferece a probabilidade para cada classe considerada.

O algoritmo gradiente descendente foi o utilizado para treinar a rede. Neste processo de treinamento o objetivo é minimizar o erro entre o resultado obtido rede e o esperado. Nas camadas densamente conectadas, o algoritmo altera os valores da matriz de pesos sinápticos, enquanto nas camadas convolucionais são os valores dos filtros que são alterados. Um subconjunto do conjunto de treinamento é submetido à rede e o algoritmos é executado, sendo cada uma destas iterações uma época de treinamento. Cada vez que o erro diminui, a rede se aproxima do comportamento esperado que é a correta classificação da regra que gerou a evolução temporal submetida à rede (HAYKIN,1998).

A mesma estrutura de rede descrita aqui e representada na Figura 4.1 foi treinada para diferentes conjuntos de treinamento com o objetivo de definir qual deles traria melhor resultados. A performance de cada um está detalhada no Seção 4.2

4.2 Classificação via rede neural

Usar redes neurais para classificação pode ser feito fornecendo à rede informações extraídas dos dados que se pretende classificar, submetendo os próprios dados alvos da

classificação e também convertendo ou tratando os dados de outra forma. Esta última foi a estratégia escolhida, fazendo com que as evoluções temporais sejam encaradas como imagens binárias e assim possam servir de entrada para redes neurais convolucionais. As Seções 4.2.1 e 4.2.2 abordam respectivamente a classificação dos espaços elementar e raio 1,5, descrevendo os resultados obtidos.

4.2.1 Classificação do espaço elementar via redes neurais convolucionais

Com relação à geração do conjunto de teste é importante frisar que há um desequilíbrio na distribuição das 4 classes, abordadas aqui, para o espaço elementar. A quantidade de regras classe 2 é muito maior que as demais, por exemplo. Para que cada classe tenha a mesma chance de ser aprendida pela rede neural, faz-se necessário que o conjunto de treinamento tenha as mesmas proporções para cada classe. Por isso, o conjunto de treinamento foi formado selecionando de forma aleatória a classe para então selecionar uma regra deste grupo ((LAURIKKALA, 2001), (BATISTA; PRATI; MONARD, 2005)).

Neste experimento tanto regras quanto condições iniciais aleatórias foram usadas para se obter um conjunto de evoluções temporais a ser usado no treinamento da rede, pois foi observado que a geração de blocos de evoluções temporais para cada regra com posterior submissão dos mesmos, em sequência, para o treinamento, não ofereceu resultados satisfatórios. A rede neural foi treinada com um total de 89.501 épocas de treinamento com 200 evoluções temporais cada uma. O fim do treinamento indicou acurácia de 91% e um patamar elevado para o erro, estabilizando em 32,85%.

A classificação do espaço elementar foi feita utilizando blocos de 500 evoluções temporais geradas a partir de condições iniciais aleatórias para cada regra do espaço. A predição da classe para a regra avaliada é definida pela resposta mais frequente obtida pela rede, a resposta mais frequente define a classe predita para a regra. Neste caso, a acurácia obtida na classificação foi total, como mostra a Tabela 4.1, onde se vê que todas as regras de cada classe foram classificadas corretamente.

Tabela 4.1: Resultado da classificação do espaço elementar usando rede neural convolucional.

Classes	Predições	Real	Percentual
1	24	24	100,00%
2	195	195	100,00%
3	26	26	100,00%
4	11	11	100,00%

A qualidade do classificador foi avaliada usando a matriz de confusão da Tabela 4.2 que indica acurácia global de 0,9902. A diferença de entre a acurácia global e a classificação do espaço ocorre porque a classificação é definida sobre um conjunto de evoluções temporais, se após avaliar o conjunto, a regra foi classificada corretamente. Já a acurácia deriva da avaliação feita para cada evolução temporal individual, neste caso cada falha individual implica na diminuição da acurácia global.

Tabela 4.2: Matriz de confusão proveniente da classificação do espaço elementar via rede neural convolucional.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	11477	392	31	15
2	0	96347	36	16
3	0	301	12699	71
4	0	211	176	5375
Acurácia global: 0,9902				

O resultado mostra que apesar da capacidade de classificação correta no espaço elementar, há imprecisões no processo, relatados na Tabela 4.3. Nela, as taxas falso positivo e negativo idealmente precisam ser 0. O mesmo acontecendo para as taxas de falsos positivo, negativo, descoberta e omissão. Já para as taxas verdadeiros positivo e negativo, valores preditivos positivo e negativo, assim como a acurácia, idealmente devem ser 1.

Tabela 4.3: Métricas da matriz de confusão proveniente da classificação do espaço elementar via rede neural convolucional.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	11554	96993	12784	5411
Verdadeiro negativo	116446	31007	115216	122589
Falso positivo	441	52	374	390
Falso negativo	0	910	245	103
Taxa de verdadeiros positivos	1,0000	0,9907	0,9812	0,9814
Taxa de verdadeiros negativos	0,9962	0,9983	0,9968	0,9968
Valor preditivo positivo	0,9632	0,9995	0,9715	0,9328
Valor preditivo negativo	1,0000	0,9715	0,9979	0,9992
Acurácia	0,9966	0,9925	0,9952	0,9962
Taxa de falsos positivos	0,0038	0,0017	0,0032	0,0032
Taxa de falsos negativos	0,0000	0,0093	0,0188	0,0186
Taxa de falsas descobertas	0,0368	0,0005	0,0285	0,0672
Taxa de falsas omissões	0,0000	0,0285	0,0021	0,0008

As Tabelas 4.2 e 4.3 mostram que o valor da acurácia global ficou pouco abaixo das acurácias para cada classe. Apesar disso a Tabela 4.3 mostra grande qualidade ao classificar o espaço elementar, pois cada métrica apresentada ficou próxima do valor desejado. Devido a esta qualidade, na próxima seção repete-se o processo, porém para o espaço raio 1,5, avaliando-se a qualidade do classificador neste caso.

4.2.2 Classificação do espaço de raio 1,5 via redes neurais convolucional

A classificação do espaço raio 1,5 não alcançou o mesmo nível de acerto observado na classificação do espaço elementar. Tal fato se compreende pelo espaço raio 1,5 possuir padrões de comportamento para os quais a rede não foi treinada. A Tabela 4.4 mostra que muitas regras foram classificadas como classe 1 de forma equivocada, já que as regras preditas são 4 vezes maiores que as regras realmente homogêneas. As classes 3 e 4 também

tiveram regras preditas erroneamente, apresentando valores superiores ao esperado, o que impactou a classe 2 que ficou abaixo da quantidade real.

Tabela 4.4: Resultado da classificação do espaço raio 1,5 usando rede neural convolucional.

Classes	Predições	Real	Percentual
1	10103	2512	402.19%
2	30340	47384	64.03%
3	19194	10771	178.20%
4	5896	4866	121.17%

A avaliação do classificador via matriz de confusão mostra a fragilidade do classificador ao ser aplicado no espaço raio 1,5. A acurácia global alcançou apenas 0,5468, como se pode observar pelos resultados na Tabela 4.5.

Tabela 4.5: Matriz de confusão da classificação do espaço de raio 1,5 via rede neural convolucional.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	1170031	3639590	45221	59875
2	86270	13253987	1332052	754487
3	375	4990019	3045156	1167729
4	705	1815533	957917	449055
Acurácia global: 0.5468				

As métricas derivadas da matriz de confusão refletem a baixa acurácia do resultado, como pode ser visto na Tabela 4.6.

Tabela 4.6: Métricas da matriz de confusão da classificação do espaço de raio 1,5 via rede neural convolucional.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	1170031	13253987	3045156	449055
Verdadeiro negativo	31597969	19514013	29722844	32318945
Falso positivo	3744685	2172809	6158123	2774155
Falso negativo	87351	10445141	2335190	1982091
Taxa de verdadeiros positivos	0,9305	0,5593	0,5660	0,1847
Taxa de verdadeiros negativos	0,8940	0,8998	0,8284	0,9209
Valor preditivo positivo	0,2381	0,8592	0,3309	0,1393
Valor preditivo negativo	0,9972	0,6514	0,9272	0,9422
Acurácia	0,8953	0,7220	0,7942	0,8732
Taxa de falsos positivos	0,1060	0,1002	0,1716	0,0791
Taxa de falsos negativos	0,0695	0,4407	0,4340	0,8153
Taxa de falsas descobertas	0,7619	0,1408	0,6691	0,8607
Taxa de falsas omissões	0,0028	0,3486	0,0728	0,0578

Apesar da correta classificação do espaço elementar evidenciado na Tabela 4.1, os resultados da avaliação da qualidade do classificador na Tabela 4.3 mostra imprecisões do classificador que afetou tanto a acurácia global quanto a de cada classe, ao classificar o espaço raio 1,5. As Tabelas 4.5 e 4.6 resumem os resultados obtidos. Para tal situação ocorrer, pode ser que os padrões visuais do espaço 1,5 que não estão presentes no espaço elementar ou as proporções destes padrões são diferentes entre o espaço de raio 1,5 e o elementar, portanto não aprendidos pela rede, contribuindo assim para os resultados.

Capítulo 5

Análise de textura e classificação de autômatos celulares

Textura está ligada à percepção de elementos gráficos ao observar uma imagem qualquer. Ao observar a fotografia de uma paisagem pode-se perceber linha, retângulos, círculos ou regiões não regulares. Isolando uma região da imagem e examinando-a mais de perto, talvez aplicando-lhe um zoom, ainda se percebe tais elementos, porém o que antes era uma região bem delimitada, passa-se a observar outros elementos mais simples constituintes da textura.

No limite, a vizinhança de um pixel, haverá um número fixo de possibilidades, as configurações possíveis da vizinhança do pixel. Sobre este conjunto de possibilidades é possível realizar uma série de análises e extrair informações das mais variadas com propósitos igualmente variados, como por exemplo melhorar a qualidade da imagem, aplicar filtros ou classificá-las. Este trabalho foca neste último objetivo.

A abordagem da análise de textura é semelhante à utilizada no Capítulo 3. Porém, enquanto lá foi extraída informações das evoluções temporais usando a entropia, neste capítulo as informações serão extraídas a partir da textura das imagens das evoluções temporais. Além disso, a quantidade de 500 evoluções temporais também foi usada. Mas, também quantidades variáveis mediante avaliação de erro definido para o acúmulo das informações, descrito mais a diante.

Por exemplo, em He e Wang (1990), imagens de satélite foram classificadas, associando

a elas o vetor de frequências relativas às configurações da vizinhança de um pixel. Com o algoritmo lá exposto foi possível obter acurácia média de 98,4% e 99,6% na classificação, a depender do grupo de imagens utilizadas.

A partir de uma base contendo evoluções temporais associadas às suas classes, (SILVA et al., 2016) propõe a extração de características das evoluções temporais utilizando uma variação do algoritmo LBP (GUO; ZHANG; ZHANG, 2010), o Local Binary Pattern Variance (LBPV), criando um vetor de característica com 10 componentes. Também é proposto usar o espectro de Fourier (SILVA et al., 2016) na extração de características, a partir das quais foi possível a criação de vetores com 64 componentes. A classificação então é feita criando uma série de vetores de características, que serão utilizados como base de comparação para o algoritmo k-Nearest Neighbor (k -NN) (COVER; HART, 2006).

Considerando apenas o primeiro mais próximo (1-NN), o espaço elementar foi classificado utilizando LBPV e o espectro de Fourier, levando a uma acurácia de 96,35% quando o LBPV foi utilizado, e 99,42% com a transformada de Fourier.

Com as evoluções temporais dos ACs aqui abordados já apresentam configuração binária na vizinhança de um pixel, é possível afirmar que ao levar em consideração os níveis de cores uma evolução temporal binária é muito mais simples que as imagens de satélite utilizadas em (HE; WANG, 1990), não sendo necessária a extração de padrões binários, pois estes já se apresentam na vizinhança do pixel. Esta percepção é a base da proposta de classificação descrita na Seção 5.1.

5.1 Proposta para a classificação de autômatos celulares binários unidimensionais usando análise de textura

Na Seção 2.4 foi abordada a análise de textura em imagens e o papel que as primitivas têm na formação das texturas, pois são suas unidades constituintes, informações obtidas em torno de um pixel. Aqui se propõe extrair informações da vizinhança do pixel (célula) de forma semelhante ao algoritmo LBP, no entanto sem a necessidade de extrair padrões binários, já que a evolução temporal é binária. Aqui a conversão para valores decimais

será feita com todas as células da vizinhança, inclusive a central, esse valor decimal representará a configuração da vizinhança. Calcular a proporção de cada configuração possível na imagem, ordenando-as em ordem crescente de seus valores decimais, permite construir um **espectro das configurações de vizinhanças**, associando a cada configuração, valor decimal, sua frequência na imagem. Construindo o espectro para cada evolução temporal de um conjunto, todas geradas a partir da mesma regra, é possível calcular o espectro médio da regra. Gerando-se um vetor para cada uma das 256 regras, tem-se o conjunto de informações utilizadas para predizer a classe da regra geradora das evoluções temporais submetidas ao classificador (NOGUEIRA; OLIVEIRA, 2019).

Uma observação se faz necessário. O algoritmo que será descrito nesta seção não se aplica às redes neurais, pois o treinamento da rede se dá de outra forma. Mas, poderia ser aplicado na classificação baseado na entropia, feita as devidas adaptações. No entanto, a abordagem usando entropia serviu apenas para avaliar o método proposto por Wuensche (1998) e verificar se o método desenvolvido aqui melhoraria a acurácia.

Detalhando o algoritmo proposto, pode-se dizer que a evolução temporal também possui texturas e estas são formadas por primitivas: as configurações da vizinhança de uma célula ao longo do tempo. No caso do AC elementar, definindo 2 passos de tempo adiante, tem-se informações da vizinhança de qualquer célula em t , $t + 1$ e $t + 2$, obtendo-se uma região cujo raio em torno da célula vale tanto para o reticulado quanto para a evolução no tempo. A Figura 5.1 exemplifica esse ponto, onde a primitiva definida para a análise das evoluções temporais é uma região quadrada da evolução temporal contendo 9 células.

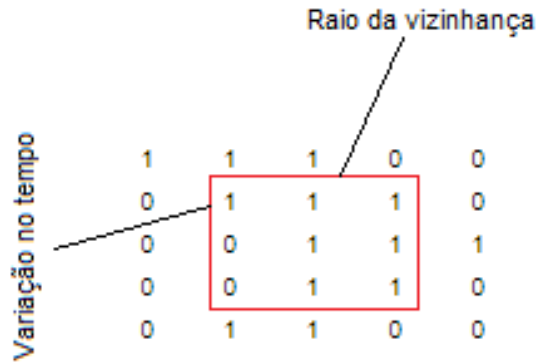


Figura 5.1: Primitiva destacada, no interior do retângulo, do trecho de uma evolução temporal com a vizinhança contendo 3 células, observadas por 3 passos de tempo.

A primitiva pode ser então convertida em um número inteiro, conforme a Equação 5.1, onde c_j^t representa as células da primitiva, i.e., a vizinhança ao redor da célula central ao longo dos tempos t , $t + 1$ e $t + 2$. Os limites j_0 e t_0 , relativos à vizinhança analisada, representam respectivamente a célula j_0 da vizinhança no momento $t_0 \in [t, t + 1, t + 2]$. No caso do espaço elementar são obtidos 512 valores distintos entre 0 e $2^9 - 1 = 511$. Assim, como outros algoritmos de extração de características, gera-se uma imagem onde o valor de cada pixel original é substituído pelo valor extraído da vizinhança, na abordagem proposta, o estado da célula é substituído pelo valor decimal referente à configuração binária da vizinhança.

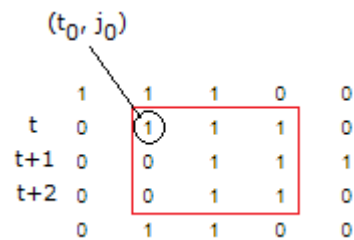


Figura 5.2: Conversão da configuração de vizinhança em valor decimal toma como referência o momento t e a coluna inicial j_0 . A conversão é feita segunda a Equação 5.1

Utilizando a Equação 5.1 define-se o vetor $\vec{s}_z = (s_0, s_1, s_2, \dots, s_{511})$ que representa o espectro referente à evolução temporal z . Por exemplo, definindo quantidades I e J quaisquer, executando a regra 0 do espaço elementar por $3I$ passos de tempo, com

$2I$ transientes, sobre um reticulado de tamanho J , produz uma evolução temporal de dimensões IJ contendo apenas a primitiva 0, como resultado $\vec{s}_z = (IJ, 0, 0, \dots, 0)$. Para outras regras, a soma para cada componente em \vec{s}_z será distinta.

$$s = \sum_{t=t_0}^{t_0+2r} \sum_{j=j_0}^{j_0+2r} c_j^t \cdot 2^{(j-j_0)+(t-t_0)\cdot(2r+1)} \quad (5.1)$$

Define-se o vetor \vec{t}_n , conforme a Equação 5.2, como sendo a soma dos espectros \vec{s}_z . Este vetor pode ser normalizado dividindo seus componentes por T que é calculado segundo a Equação 5.3. O resultado é o vetor de frequências associado à regra geradora n das Z evoluções temporais, conforme Equação 5.4.

$$\vec{t}_n = \sum_{z=0}^{Z-1} \vec{s}_z \quad (5.2)$$

$$T = \sum_{i=0}^{511} t_{ni} \quad (5.3)$$

$$\vec{f}_n = \frac{1}{T} \cdot \vec{t}_n \quad (5.4)$$

O conjunto $F = \{\vec{f}_0, \vec{f}_1, \vec{f}_2, \dots, \vec{f}_{255}\}$ é formado pelos espectros associados a cada regra do espaço elementar, a ser usado na classificação das evoluções temporais. No entanto o número de condições iniciais Z necessário para que \vec{f}_n tenha uma configuração representativa da regra n é problema que se impõe. De fato, o número total de condições iniciais possíveis é um problema computacional de difícil tratamento, devido ao tamanho de uma amostra relevante. Um reticulado com 100 células possibilita 2^{100} condições iniciais distintas; mesmo 1% deste valor significa um total de condições iniciais da ordem de 10^{28} .

Se fixar quantidade de amostras relevantes implica em um custo computacional muito alto, é possível estabelecer condição de parada para a geração dos espectros das evoluções temporais. A contribuição do vetor \vec{s}_z conforme ele é acrescido ao vetor \vec{t}_n , diminui a cada iteração; assim, esta variação negativa pode ser usada para definir que contribuições adicionais de \vec{s}_z são irrelevantes, de forma a se poder suspender o processo. Sendo assim, um vetor \vec{p} , Equação 5.5, formado pela proporção entre as componentes dos vetores \vec{s}_z

e \vec{t}_n , tem sua norma cada vez menor conforme \vec{s}_z é acrescido a \vec{t}_n . Uma vez que as condições iniciais são aleatórias, não é possível prever como cada componente varia ao longo do tempo, no entanto, é possível afirmar que a norma de \vec{p} deve diminuir ao longo do tempo. Estabelecendo um erro (e) mínimo desejado, o processo de soma dos vetores é interrompido quando $|\vec{p}| \leq e$.

$$\vec{p} = p(\vec{s}_z, \vec{t}_n) = \left(\frac{s_{0,z}}{t_{0,n}}, \frac{s_{1,z}}{t_{1,n}}, \dots, \frac{s_{511,z}}{t_{511,n}} \right) \quad (5.5)$$

No entanto é necessário observar que ao utilizar condições iniciais aleatórias para se gerar evoluções temporais de uma regra, pode-se obter uma norma menor ou igual ao erro com uma amostra muito reduzida; por isso é necessário não só garantir que uma amostra de tamanho mínimo arbitrário seja escolhida para todas as regras, como também que a convergência da norma seja suave. Para resolver ambos os pontos, foi adotado não a norma de uma única instância de \vec{p} , mas uma média móvel da quantidade mínima arbitrária de condições iniciais, interrompendo-se o processo quando a média móvel fosse menor ou igual ao erro e .

5.2 Classificação via análise de textura

é a base para classificar evoluções temporais usando primitivas de texturas neste trabalho, e saber quão distintos são os elementos de F para que possam ser bem diferenciados, é uma questão importante. A Figura 5.3 mostra um diagrama para as regras do ACE, onde os tons mais escuros indicam grande distância entre dois espectros, enquanto as mais claras indicam menor distância, sendo a coordenada de cada pixel o par ordenado (n_i, n_j) das regras do espaço elementar, fazendo com que a imagem seja quadrada (256×256). É possível observar a devida simetria na imagem, pois $d(x, y) = d(y, x)$, onde x e y são espectros do conjunto, e como esperado a distância na diagonal principal é zero. Os tons de cinza variados do diagrama indicam a possibilidade de se obter baixa ambiguidade ao comparar espectros das evoluções temporais utilizando a distância euclidiana, métrica usada para se definir a qual espectro do conjunto, aquele submetido está mais próximo.

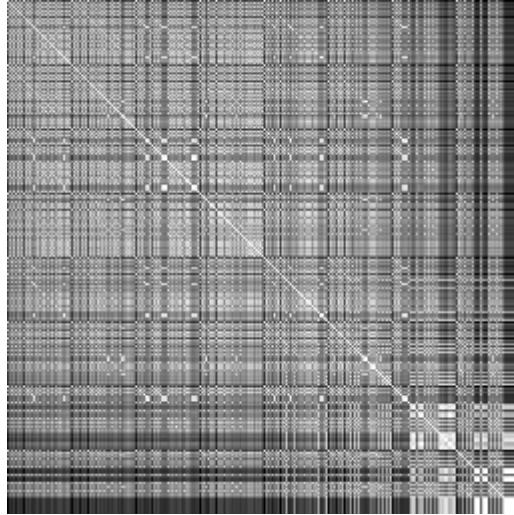


Figura 5.3: Diagrama das distâncias entre os espectros das regras do espaço elementar. Cada pixel (x, y) é a informação da distância entre as regras x e y , onde as distâncias maiores são representadas por tons mais escuros e as mais baixas por tons claros.

Reduzir o esforço computacional para criar o conjunto F utilizando apenas as regras equivalentes do espaço elementar é uma boa abordagem, pois reduz a quantidade de comparações necessárias para a predição. Neste caso, a forma de utilizar o conjunto e avaliação das evoluções temporais não muda e a distinção entre as regras utilizando a distância para diferenciá-las permanece. Ao gerar o conjunto de espectros apenas com as regras dinamicamente equivalentes obtêm-se um diagrama semelhante ao anterior, como se pode ver pela Figura 5.4, que possui configuração semelhante à 5.3, com diferença apenas na quantidade de regras utilizadas, indicando que utilizar as regras dinamicamente equivalentes também é possível criar o conjunto F com espectros distintos.

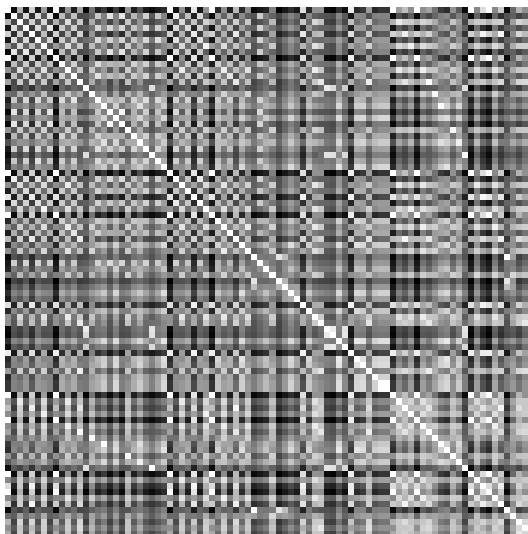


Figura 5.4: Diagrama das distâncias de um conjunto contendo apenas espectros das regras equivalentes. Cada pixel (x, y) é a informação da distância entre as regras x e y , onde as distâncias maiores são representadas por cores mais escuras e por consequência, as mais baixas pelas claras.

Os diagramas das Figuras 5.3 e 5.4 apresentam menos pontos brancos devido ao uso de condições iniciais aleatórias, fazendo com que o conjunto de evoluções temporais utilizadas para gerar os espectros de cada regra seja diferente. Um conjunto finito com as mesmas condições iniciais para gerar os espectros das regras não se mostrou eficaz para a classificação. Poderia-se aventar a possibilidade de que uma quantidade representativa de condições iniciais pudesse ser usada, mas isso implicaria em problema computacional de difícil solução devido ao número de condições iniciais possíveis para um reticulado com 100 células.

Espectros gerados a partir de um único conjunto de condições iniciais, fariam com que a distância dos espectros das regras equivalentes fosse nula, já que seriam o mesmo. Mas utilizar condições iniciais distintas mostrou-se mais eficaz, pois mesmo para as regras equivalentes os espectros se mostraram distintos. As Figuras 5.5 e 5.6, referente as regras equivalentes a 30 e 54 respectivamente, compara seus espectros evidenciando suas diferenças e esta variabilidade foi útil na correta predição das classes no espaço elementar, como será mostrada na Seção 5.2.1.

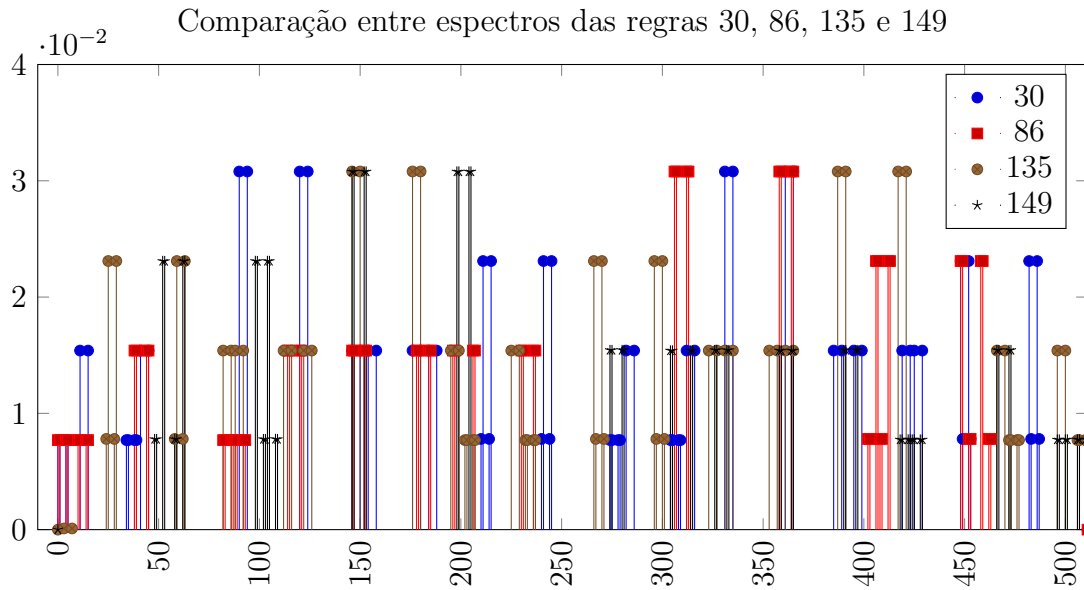


Figura 5.5: Comparação dos espectros gerados para as regras dinamicamente equivalentes à regra 30, utilizando condições iniciais aleatórias para cada uma.

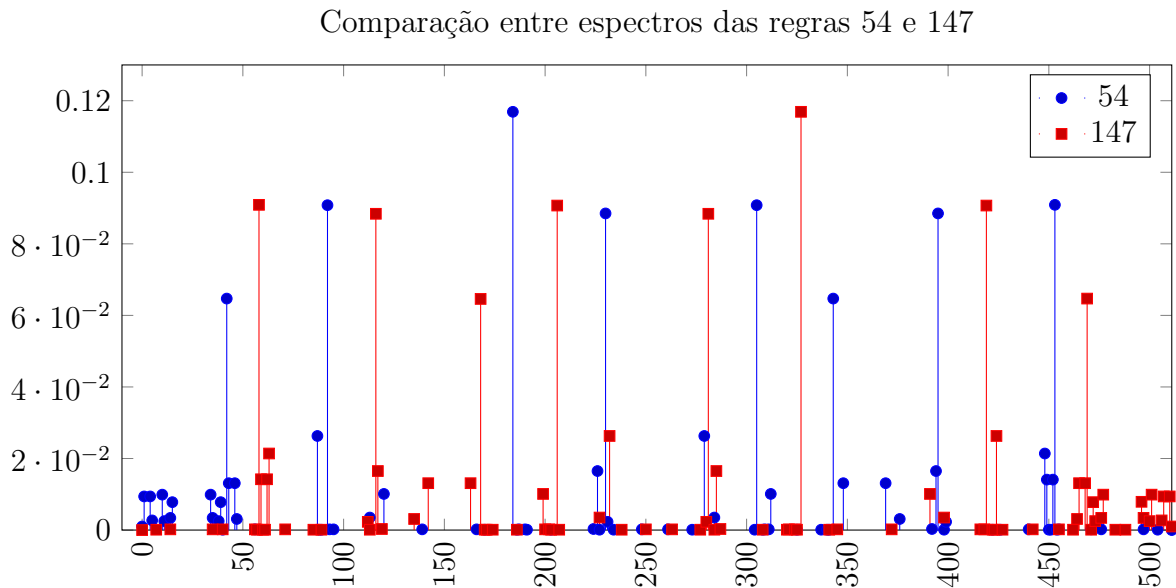


Figura 5.6: Comparação dos espectros gerados para as regras dinamicamente equivalentes à regra 54, utilizando condições iniciais aleatórias para cada uma.

Ao gerar o conjunto F sem uma quantidade fixa de evoluções temporais, mas sim esperando que se atinja um erro mínimo, observou-se uma queda acentuada no início do processo na norma de \vec{p} e ao longo do tempo esta queda se tornou muito pequena, dificultando alcançar o erro desejado. A Figura 5.7 mostra como a norma diminuiu,

conforme os valores dos espectros das evoluções temporais da regra 27 do espaço elementar foram somados, continuando o processo até se obter 27.212 evoluções temporais, para reticulados com 30 células e média móvel de 250. Ao aumentar o tamanho do reticulado para 100 células com média móvel de 500 espectros, bastaram 694 evoluções temporais para que o erro fosse menor ou igual a 10^{-5} .

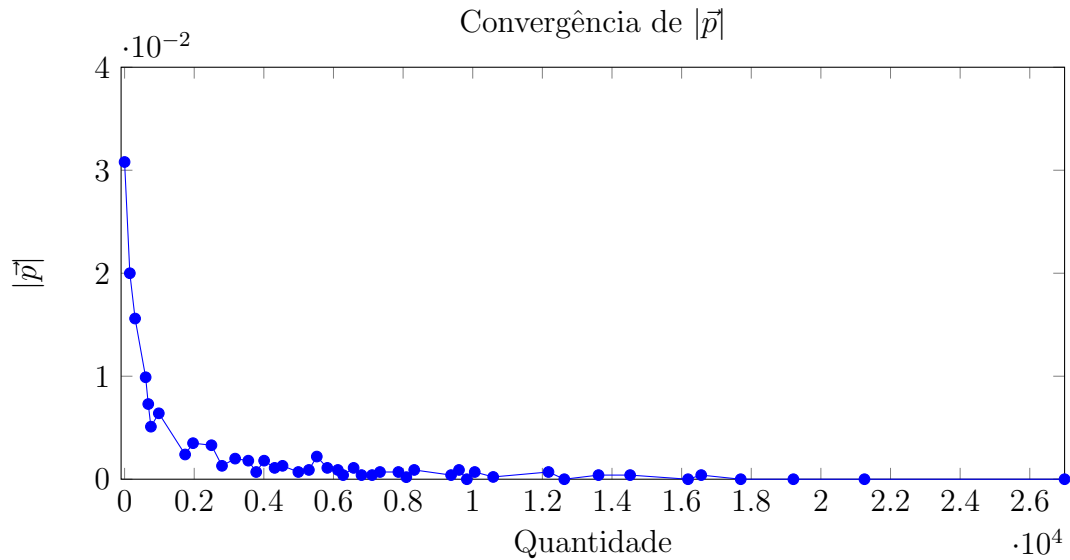
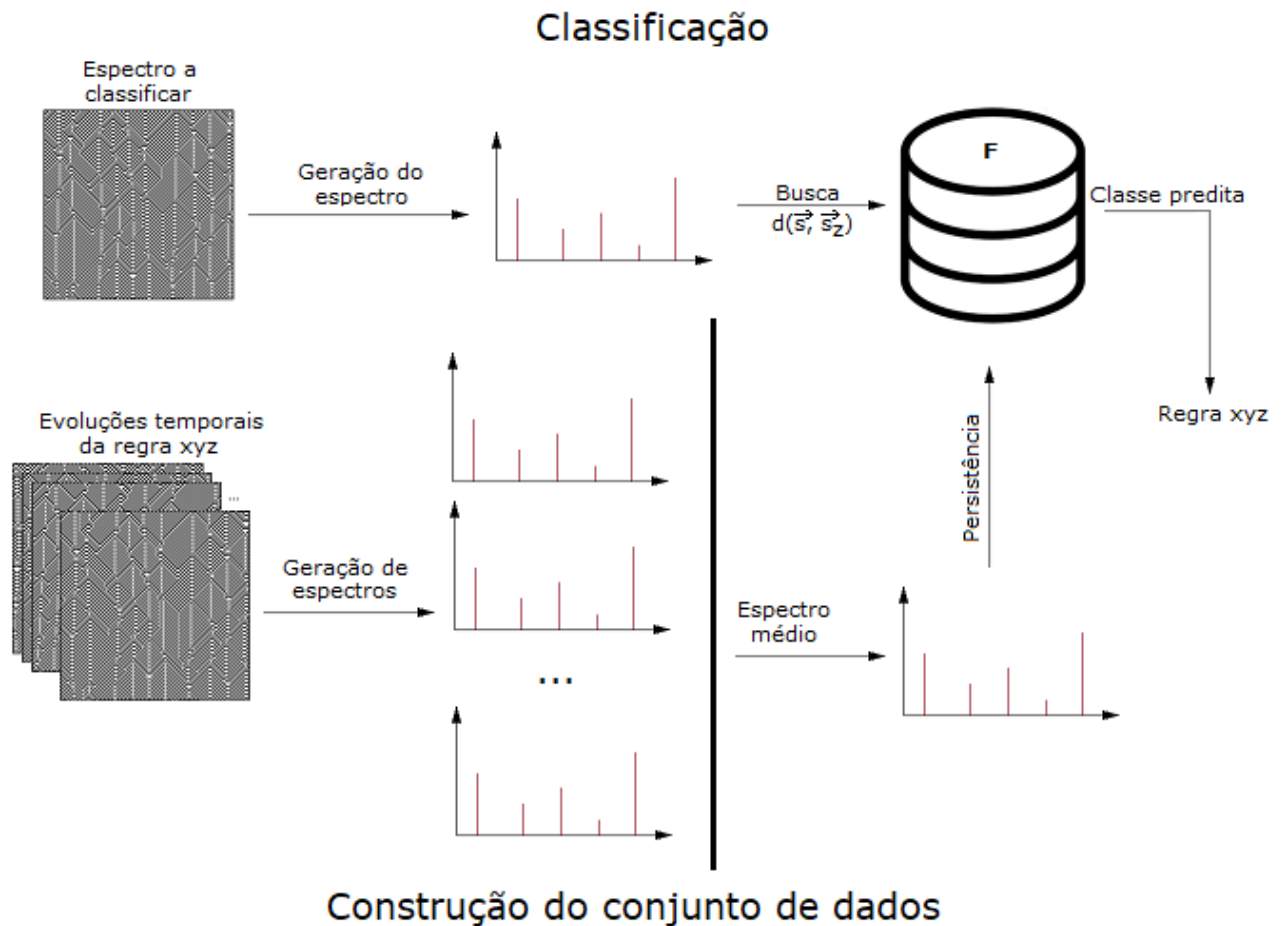


Figura 5.7: Variação da norma do vetor f para a regra 27 do espaço elementar para reticulados com 30 células e média móvel de 250 evoluções temporais.

A convergência da norma de \vec{p} mostrou ser uma boa condição de parada no acúmulo de espectros, proporcionando um conjunto que levou a bons resultados na classificação das evoluções temporais, como será descrito nesta Seções 5.2.1 e 5.2.2. Como mencionado, o número de evoluções temporais necessárias para a regra 27 foi de apenas 694, quantidade muito inferior às 2^{100} evoluções temporais possíveis.

Construído F , classifica-se uma evolução temporal, e por consequência a regra geradora, gerando seu espectro (\vec{s}_z) e calculando a distância euclidiana entre este e os elementos de F ; aquele que tiver a menor distância, é escolhido. E como cada elemento de F está associado a uma regra do espaço elementar, prever a classe é trivial. A vantagem deste processo está no fato de que a evolução temporal submetida à classificação precisa apenas ser binária, sendo irrelevante a regra geradora. O conjunto F permite associar qualquer evolução temporal de um AC binário unidimensional a uma das regras do espaço elementar, e desta associação, prever sua classe.

Figura 5.8: Diagrama que sintetiza a classificação utilizando análise de textura. Destaca a criação do espectro médio a partir de vários espectros criados a partir de evoluções temporais de uma regra. E também, como o classificador prediz a classe de uma evolução temporal.



Nas subseções seguintes os resultados foram separados em dois experimentos, no primeiro o conjunto F foi gerado utilizando apenas as regras dinamicamente equivalentes, enquanto no segundo todas foram utilizadas. Os experimentos foram subdivididos entre dois conjuntos de espectros, um onde o número de condições iniciais foi fixa em 500 e outro onde elas eram geradas, com média móvel de 500, até que o erro fosse menor a um limiar fixado. O número 500 foi definido empiricamente, quando se observou que esta quantidade oferecia melhores resultados, mas apesar disso, estabelecer um limiar para o erro permite que se busque a quantidade adequada de evoluções temporais que faça o espectro resultante ser representativo de uma regra. O tamanho do reticulado para todos os casos foi fixado em 100 células para a construção do conjunto F e a avaliação.

5.2.1 Experimentos para as regras dinamicamente equivalentes

Utilizando apenas as regras dinamicamente equivalentes do espaço elementar o esforço computacional é menor. Com o experimento busca-se compreender se haverá informações suficientes para que o classificador obtenha boa performance, apesar do número menor de regras. O experimento foi subdividido em duas abordagens, onde a primeira utilizou 500 condições iniciais fixas para cada regra e na segunda, foram acumuladas condições iniciais até que $|\vec{p}| < 10^{-5}$.

Fixando em 500 a quantidade de condições iniciais utilizadas para gerar o conjunto F , para as regras dinamicamente equivalentes, foi possível classificar corretamente todas as regras do espaço elementar. A Tabela 5.1 mostra que as regras de cada classe foram corretamente classificadas, com acurácia máxima.

Tabela 5.1: Resultado da classificação do espaço elementar usando usando análise de textura. Conjunto de dados baseados nas regras equivalentes e quantidade fixa de evoluções temporais.

Classes	Predições	Real	Percentual
1	24	24	100.0%
2	196	196	100.0%
3	26	26	100.0%
4	10	10	100.0%

Aqui também a qualidade do classificador foi avaliada através da matriz de confusão da Tabela 5.2, com as classes preditas e reais, respectivamente em linhas e colunas. A Tabela 5.3 contém as métricas, derivadas da matriz de confusão, usadas na avaliação do classificador.

Tabela 5.2: Matriz de confusão do experimento 1 - Fixo.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	11638	0	0	0
2	991	94986	23	0
3	0	0	16000	0
4	0	64	0	4297
Acurácia global: 0,9916				

A **acurácia global** de 0,9916 e as métricas da Tabela 5.3, confirmam a boa performance do classificador apresentada ao classificar o espaço elementar.

Tabela 5.3: Métricas da matriz de confusão do experimento 1 - Fixo.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	11638	94986	16000	4297
Verdadeiro negativo	116362	33014	112000	123703
Falso positivo	0	1014	0	64
Falso negativo	991	64	23	0
Taxa de verdadeiros positivos	0,9215	0,9993	0,9986	1,0000
Taxa de verdadeiros negativos	1,0000	0,9702	1,0000	0,9995
Valor preditivo positivo	1,0000	0,9894	1,0000	0,9853
Valor preditivo negativo	0,9916	0,9981	0,9998	1,0000
Acurácia	0,9923	0,9916	0,9998	0,9995
Taxa de falsos positivos	0,0000	0,0298	0,0000	0,0005
Taxa de falsos negativos	0,0785	0,0007	0,0014	0,0000
Taxa de falsas descobertas	0,0000	0,0106	0,0000	0,0147
Taxa de falsas omissões	0,0084	0,0019	0,0002	0,0000

Apesar da boa performance do classificador, ao considerar o valor desejado de 1,0000 para o **valor preditivo positivo** e de 0,0000 para a **taxa de falsa descoberta**, percebe-

se altos para este e baixos para aquele, com mínimos respectivos de 0,9853 e 0,0147.

No experimento com quantidades variáveis de condições iniciais, para gerar o conjunto F , levando em conta apenas as regras dinamicamente equivalente, foi adotado 10^{-5} como erro desejado. Assim como o classificador, cujos resultados estão na Tabela 5.1, este também acertou a classe de todas as regras, como se pode ver pela Tabela 5.4.

Tabela 5.4: Resultado da classificação do espaço elementar usando usando análise de textura. Conjunto de dados baseados nas regras equivalentes e quantidade de evoluções temporais variáveis para gerar F

Classes	Predições	Real	Percentual
1	24	24	100.0%
2	196	196	100.0%
3	26	26	100.0%
4	10	10	100.0%

A Tabela 5.5 contém a matriz de confusão referente ao classificador, com as classes preditas em linha e reais nas colunas. Já a Tabela 5.6 contém as métricas, derivadas da matriz de confusão, usadas na avaliação do classificador.

Tabela 5.5: Matriz de confusão do experimento 1 - Variável.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	9902	0	0	0
2	336	88869	33	0
3	0	0	23003	0
4	0	83	0	5774
Acurácia global: 0,9965				

Evidencia-se um incremento no valor da **acurácia global**, agora de a 0,9965, permitindo dizer que a estratégia adotada melhorou a efetividade do classificador no espaço elementar.

Tabela 5.6: Métricas da matriz de confusão do experimento 1 - Variável.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	9902	88869	23003	5774
Verdadeiro negativo	118098	39131	104997	122226
Falso positivo	0	369	0	83
Falso negativo	336	83	33	0
Taxa de verdadeiros positivos	0,9672	0,9991	0,9986	1,0000
Taxa de verdadeiros negativos	1,0000	0,9907	1,0000	0,9993
Valor preditivo positivo	1,0000	0,9959	1,0000	0,9858
Valor preditivo negativo	0,9972	0,9979	0,9997	1,0000
Acurácia	0,9974	0,9965	0,9997	0,9994
Taxa de falsos positivos	0,0000	0,0093	0,0000	0,0007
Taxa de falsos negativos	0,0328	0,0009	0,0014	0,0000
Taxa de falsas descobertas	0,0000	0,0041	0,0000	0,0142
Taxa de falsas omissões	0,0028	0,0021	0,0003	0,0000

As taxas individuais mostram melhora generalizada na performance do classificador para o espaço elementar, apesar de a **Taxa de verdadeiros positivos** para a classe 4 ser de 0,9858 e a **Taxa de falsas descobertas** para a classe 1 ser de 0,0328.

5.2.2 Experimentos para as regras do espaço elementar

Neste experimento o conjunto F foi gerado utilizando todas as regras do espaço elementar. O experimento foi sub-dividido em duas abordagens, onde a primeira utilizou 500 condições iniciais fixas para cada regra e, na segunda, foram acumuladas condições iniciais até que $|\vec{p}| < 10^{-5}$.

Fixando em 500 a quantidade de condições iniciais utilizadas para gerar o conjunto F , para cada uma das regras do espaço elementar, obteve-se a Tabela 5.7 como resultado da classificação.

Tabela 5.7: Resultado da classificação do espaço elementar usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F .

Classes	Predições	Real	Percentual
1	24	24	100,0%
2	196	196	100,0%
3	26	26	100,0%
4	10	10	100,0%

Relata-se a qualidade do classificador na Tabela 5.8 (matriz de confusão do experimento), com as classes previstas e reais, respectivamente em linhas e colunas. A Tabela 5.9 contém as métricas, derivadas da matriz de confusão, usadas na avaliação do classificador.

Tabela 5.8: Matriz de confusão do experimento 2 - Fixo.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	12002	0	0	0
2	627	96843	30	4
3	0	1	13000	0
4	0	91	0	5403
Acurácia global: 0,9941				

A **acurácia global** apresenta redução no seu valor, pouco menor que o experimento anterior, chegando a 0,9941, indicando que numericamente, esta estratégia diminuiu a capacidade preditiva do classificador, com relação ao experimento anterior.

Tabela 5.9: Métricas da matriz de confusão do experimento 2 - Fixo.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	12002	96843	13000	5403
Verdadeiro negativo	115998	31157	115000	122597
Falso positivo	0	661	1	91
Falso negativo	627	92	30	4
Taxa de verdadeiros positivos	0,9504	0,9991	0,9977	0,9993
Taxa de verdadeiros negativos	1,0000	0,9792	1,0000	0,9993
Valor preditivo positivo	1,0000	0,9932	0,9999	0,9835
Valor preditivo negativo	0,9946	0,9971	0,9997	1,0000
Acurácia	0,9951	0,9942	0,9998	0,9993
Taxa de falsos positivos	0,0000	0,0208	0,0000	0,0007
Taxa de falsos negativos	0,0496	0,0009	0,0023	0,0007
Taxa de falsas descobertas	0,0000	0,0068	0,0001	0,0165
Taxa de falsas omissões	0,0054	0,0029	0,0003	0,0000

Neste caso as métricas da matriz de confusão se afastaram de seu ponto ótimo. Evidenciando que esta estratégia diminuiu a qualidade do classificador, com relação à anterior.

No experimento com quantidades variáveis de condições iniciais para as regras do espaço elementar, foi adotado 10^{-5} como erro desejado. Obteve-se a classificação descrita na Tabela 5.10.

Tabela 5.10: Resultado da classificação do espaço elementar usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade variável de evoluções temporais para gerar F .

Classes	Predições	Real	Percentual
1	24	24	100,0%
2	196	196	100,0%
3	26	26	100,0%
4	10	10	100,0%

A avaliação da qualidade do classificador resultou na matriz de confusão da Tabela 5.11, com as classes preditas e reais, respectivamente em linhas e colunas. A Tabela 5.12, contém as métricas, derivadas da matriz de confusão, usadas na avaliação do classificador.

Tabela 5.11: Matriz de confusão do experimento 2 - Variável.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	10268	0	0	0
2	228	91404	23	2
3	0	2	18657	0
4	0	121	0	7294
Acurácia global: 0,9971				

Esta foi a melhor marca alcançada pela **acurácia global** dentre os testes realizados, mostrando a efetividade do classificador para o espaço elementar, o que permite dizer ser esta a melhor configuração do classificador.

Tabela 5.12: Métricas da matriz de confusão do experimento 2 - Variável.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	10268	91404	18657	7294
Verdadeiro negativo	117732	36596	109343	120706
Falso positivo	0	253	2	121
Falso negativo	228	124	23	2
Taxa de verdadeiros positivos	0,9783	0,9986	0,9988	0,9997
Taxa de verdadeiros negativos	1,0000	0,9931	1,0000	0,9990
Valor preditivo positivo	1,0000	0,9972	0,9999	0,9837
Valor preditivo negativo	0,9981	0,9966	0,9998	1,0000
Acurácia	0,9982	0,9971	0,9998	0,9990
Taxa de falsos positivos	0,0000	0,0069	0,0000	0,0010
Taxa de falsos negativos	0,0217	0,0014	0,0012	0,0003
Taxa de falsas descobertas	0,0000	0,0028	0,0001	0,0163
Taxa de falsas omissões	0,0019	0,0034	0,0002	0,0000

Evidencia-se este experimento como aquele que levou ao melhor resultado em classificar automaticamente regras geradoras de evoluções temporais, do espaço elementar. Apesar de já se conhecer a classificação deste espaço, o processo pode auxiliar em definir a classe da regra desconhecida, geradora de uma evolução temporal binária submetida ao classificador. Sendo assim, o espaço elementar não é um limite, já que, em última análise, qualquer matriz binária pode ser submetida ao classificador, mesmo não tendo sido gerada por um AC qualquer. Este fato nos permite submeter ao classificador evoluções temporais geradas por regras de espaços maiores que o elementar. Feito isso, na Subseção 5.2.3 avalia-se a submissão de evoluções temporais geradas por regras do espaço 1,5.

5.2.3 Classificação do espaço de raio 1,5

Observada a eficácia do classificador na predição das regras elementares, busca-se aplicá-lo no espaço de raio 1,5 e avaliar quão escalável o classificador seria. Para tal, foi feita uma classificação visual do espaço 1,5, que passou a servir de referência para se avaliar

a acurácia neste caso. A classificação completa está disponível em <https://bit.ly/2VwVnhV>.

Como feito para o espaço elementar, geraram-se 500 espectros para cada uma das 65.536 regras deste espaço, e calculou-se o percentual de predição para cada uma das 4 classes, sendo a de maior taxa, a classe predita. Assim, se houver 25% de predições para cada uma, significa situação cuja predição é indefinida. Porém, tal indefinição para as quatro classes não foi observada, mas sim para pares ou trincas de classes, impactando a qualidade do classificador.

Para a predição, os espectros das evoluções temporais foram comparados tanto com um conjunto F de espectros construídos a partir de 500 evoluções temporais para cada regra do espaço elementar, quanto com o conjunto onde os espectros foram acumulados até que um erro máximo fosse alcançado. Enquanto que para o espaço elementar ambas as estratégias apresentaram altas taxas de acurácia, tal fato não se observou para o espaço de raio 1,5 devido às indefinições citadas.

Analisando as indefinições observadas, percebeu-se que o cálculo de distância entre os espectros é muito influenciado pela frequência com que ocorre a configuração da vizinhança na evolução temporal, fazendo com que evoluções temporais de regras classe 2 fossem classificadas como 1, por ser pequena a ocorrência de configurações relacionadas à classe 2 da evolução temporal analisada. Isso quer dizer que, a depender do espectro gerado por uma evolução temporal, sua classificação poderá ser distinta da classe geradora devido à frequência da configuração de vizinhança, podendo uma evolução temporal gerada por uma regra periódica, classe 2, ser classificada como 3 ou 4 se os elementos que se repetem fazem com que seu espectro seja mais próximo de espectros dessas classes.

Exemplificando, a regra 1.572 do espaço de raio 1,5 teve sua classe predita corretamente como 2, mas 25% das vezes foi classificada com classe 1, devido à escassa quantidades de linhas verticais em algumas evoluções temporais da regra. A Figura 5.9 mostra uma evolução temporal da regra 1.572 com apenas duas linhas verticais, sendo o espectro mais próximo o da regra 224 do espaço elementar, regra homogênea, distante 0,0346 do espectro da evolução temporal referente à Figura 5.9. Construindo um par ordenado (componente; frequência) do espectro da evolução temporal, os únicos que não possuem frequências nulas são: (0; 0,97), (73; 0,01), (146; 0,01) e (292; 0,01). Dada esta configuração, este espectro está muito mais próximo daqueles gerados por evoluções temporais

homogêneas que de outros.

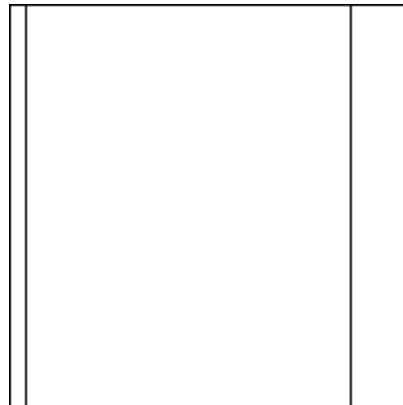


Figura 5.9: Evolução temporal da regra 1572, de raio 1,5, classificada como classe 1, sendo correta a classe 2.

A Figura 5.10 mostra uma comparação entre os espectros da regra 224 do espaço elementar e o da evolução temporal referente à regra 1.572. É possível perceber que neste caso há uma impossibilidade de diferenciar os dois espectros, pois as contribuições das componentes além do 0 são muito pequenas, fazendo com que o espectro seja considerado, devido à proximidade, o de uma evolução temporal homogênea.

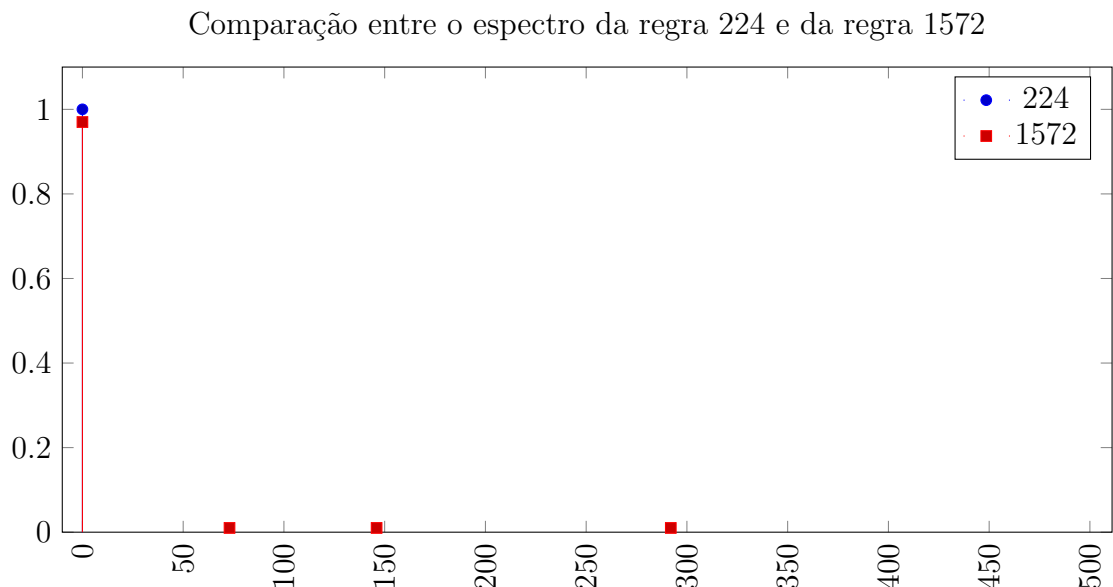


Figura 5.10: Comparação entre o espectro da regra 224 do espaço elementar e o da regra 1.572 do espaço de raio 1,5, evidenciando a dificuldade na diferenciação entre ambos, usando a distância Euclidiana.

A situação particular da regra 1.572 se repetiu para outras periódicas classificadas como classe 1, 3 ou 4. Mas não apenas as periódicas, todas as classes apresentaram falsos positivos devido à proporção das configurações de vizinhança levando a conclusões equivocadas, afetando a acurácia da classificação.

Os mesmos experimentos feitos para o espaço elementar, foram repetidos no raio 1,5, separando-os em dois grupos: 1) O conjunto F de referência formado por espectros gerados a partir das regras equivalentes dinamicamente, para 500 evoluções temporais fixas e variáveis, mas também assumindo 500 como uma média móvel na formação do conjunto, até se obter um erro abaixo de um limiar desejado. 2) O conjunto F de referência sendo formado por espectros gerados a partir das regras do espaço elementar, para 500 evoluções temporais fixas, mas também variáveis segundo a média móvel.

Experimento 1 - Conjunto F gerado com quantidades fixas de evoluções temporais: A classificação neste caso não obteve a mesma performance ao classificar o espaço elementar, conforme descrito na Tabela 5.13.

Tabela 5.13: Resultado da classificação do espaço raio 1,5 usando usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F .

Classes	Predições	Real	Percentual
1	1352	2512	53,82%
2	31011	47387	65,44%
3	32492	10771	301,66%
4	681	4866	14,00%

As Tabelas 5.14 e 5.15, apresentam respectivamente a matriz de confusão e os valores obtidos para as taxa avaliadas.

Tabela 5.14: Matriz de confusão do experimento no espaço de raio 1,5 usando conjunto espectros de referência baseados nas regras equivalentes, utilizando 500 evoluções temporais na geração do conjunto e na predição.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	627500	48500	0	0
2	627500	14393500	170000	314500
3	1000	9016000	5183000	2046000
4	0	235500	32500	72500
Acurácia global: 0,6188				

Tabela 5.15: Métricas da matriz de confusão do experimento 1 - Fixas.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	627500	14393500	5183000	72500
Verdadeiro negativo	32140500	18374500	27585000	32695500
Falso positivo	48500	1112000	11063000	268000
Falso negativo	628500	9300000	202500	2360500
Taxa de verdadeiros positivos	0,4996	0,6075	0,9624	0,0298
Taxa de verdadeiros negativos	0,9985	0,9429	0,7137	0,9919
Valor preditivo positivo	0,9283	0,9283	0,3190	0,2129
Valor preditivo negativo	0,9808	0,6640	0,9927	0,9327
Acurácia	0,9798	0,7589	0,7442	0,9257
Taxa de falsos positivos	0,0015	0,0571	0,2863	0,0081
Taxa de falsos negativos	0,5004	0,3925	0,0376	0,9702
Taxa de falsas descobertas	0,0717	0,0717	0,6810	0,7871
Taxa de falsas omissões	0,0192	0,3360	0,0073	0,0673

Experimento 1 - Conjunto de F gerado com quantidades variáveis de evoluções temporais: A classificação obteve o resultado descrito na Tabela 5.16.

Tabela 5.16: Resultado da classificação do espaço raio 1,5 usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F .

Classes	Predições	Real	Percentual
1	1352	2512	53,82%
2	31011	47387	65,44%
3	32492	10771	301,66%
4	681	4866	14,00%

As Tabelas 5.17 e 5.18, apresentam respectivamente a matriz de confusão e os valores obtidos para as taxas avaliadas.

Tabela 5.17: Matriz de confusão do experimento no espaço de raio 1,5. Conjunto espectros de referência baseados nas regras equivalentes. Quantidades variáveis de evoluções temporais na geração do conjunto e 500 fixas na predição.

Matriz de confusão				
Classe	Real			
Predita	1	2	3	4
1	627500	50500	0	0
2	627500	14380500	164000	248500
3	1000	9017500	5188000	2113500
4	0	245000	33500	71000
Acurácia global: 0,6185				

Tabela 5.18: Métricas da matriz de confusão do experimento 1 - Variável.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	627500	14380500	5188000	71000
Verdadeiro negativo	32140500	18387500	27580000	32697000
Falso positivo	50500	1040000	11132000	278500
Falso negativo	628500	9313000	197500	2362000
Taxa de verdadeiros positivos	0,4996	0,6069	0,9633	0,0292
Taxa de verdadeiros negativos	0,9984	0,9465	0,7124	0,9916
Valor preditivo positivo	0,9255	0,9326	0,3179	0,2031
Valor preditivo negativo	0,9808	0,6638	0,9929	0,9326
Acurácia	0,9797	0,7599	0,7431	0,9254
Taxa de falsos positivos	0,0016	0,0535	0,2876	0,0084
Taxa de falsos negativos	0,5004	0,3931	0,0367	0,9708
Taxa de falsas descobertas	0,0745	0,0674	0,6821	0,7969
Taxa de falsas omissões	0,0192	0,3362	0,0071	0,0674

Experimento 2 - Conjunto de referência gerado com quantidades fixas: A classificação obteve o resultado descrito na Tabela 5.19.

Tabela 5.19: Resultado da classificação do espaço raio 1,5 usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade fixa para gerar F .

Classes	Predições	Real	Percentual
1	2710	2512	107,88%
2	38219	47387	80,65%
3	22835	10771	212,00%
4	1772	4866	36,42%

As Tabelas 5.20 e 5.21, apresentam respectivamente a matriz de confusão e os valores obtidos para as taxas avaliadas.

Tabela 5.20: Matriz de confusão do experimento no espaço de raio 1,5. Conjunto de espectros de referência baseados nas regras do espaço elementar. Utilizando 500 evoluções temporais fixas na geração e predição.

Classe	Real			
Predita	1	2	3	4
1	1255000	100000	0	0
2	0	18339500	332500	437500
3	1000	4691000	4953000	1772500
4	0	563000	100000	223000
Acurácia global: 0,7559				

Tabela 5.21: Métricas da matriz de confusão do experimento 2 - Fixas.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	1255000	18339500	4953000	223000
Verdadeiro negativo	31513000	14428500	27815000	32545000
Falso positivo	100000	770000	6464500	663000
Falso negativo	1000	5354000	432500	2210000
Taxa de verdadeiros positivos	0,9992	0,7740	0,9197	0,0917
Taxa de verdadeiros negativos	0,9968	0,9493	0,8114	0,9800
Valor preditivo positivo	0,9262	0,9597	0,4338	0,2517
Valor preditivo negativo	1,0000	0,7294	0,9847	0,9364
Acurácia	0,9969	0,8425	0,8261	0,9194
Taxa de falsos positivos	0,0032	0,0507	0,1886	0,0200
Taxa de falsos negativos	0,0008	0,2260	0,0803	0,9083
Taxa de falsas descobertas	0,0738	0,0403	0,5662	0,7483
Taxa de falsas omissões	0,0000	0,2706	0,0153	0,0636

Experimento 2 - Conjunto de referência gerado com quantidades variáveis:

A classificação obteve o resultado descrito na Tabela 5.22.

Tabela 5.22: Resultado da classificação do espaço raio 1,5 usando análise de textura. Conjunto de dados baseados nas regras do espaço elementar e quantidade variável para gerar F .

Classes	Predições	Real	Percentual
1	2716	2512	108,12%
2	38214	47387	80,64%
3	22816	10771	211,83%
4	1790	4866	36,79%

As Tabelas 5.23 e 5.24, apresentam respectivamente a matriz de confusão e os valores obtidos para as taxas avaliadas.

Tabela 5.23: Matriz de confusão do experimento no espaço de raio 1,5. Conjunto espectros de referência baseados nas regras do espaço elementar. Utilizando quantidades variáveis de evoluções temporais na geração e 500 para a predição.

Classe	Real			
Predita	1	2	3	4
1	1255000	103000	0	0
2	0	18335000	332500	439500
3	1000	4689500	4950000	1767500
4	0	566000	103000	226000
Acurácia global: 0.7558				

Tabela 5.24: Métricas da matriz de confusão do experimento 2 - Variáveis.

Resultados				
Classes	1	2	3	4
Verdadeiro positivo	1255000	18335000	4950000	226000
Verdadeiro negativo	31513000	14433000	27818000	32542000
Falso positivo	103000	772000	6458000	669000
Falso negativo	1000	5358500	435500	2207000
Taxa de verdadeiros positivos	0,9992	0,7738	0,9191	0,0929
Taxa de verdadeiros negativos	0,9967	0,9492	0,8116	0,9799
Valor preditivo positivo	0,9242	0,9596	0,4339	0,2525
Valor preditivo negativo	1,0000	0,7293	0,9846	0,9365
Acurácia	0,9968	0,8424	0,8262	0,9193
Taxa de falsos positivos	0,0033	0,0508	0,1884	0,0201
Taxa de falsos negativos	0,0008	0,2262	0,0809	0,9071
Taxa de falsas descobertas	0,0758	0,0404	0,5661	0,7475
Taxa de falsas omissões	0,0000	0,2707	0,0154	0,0635

Os dados apresentados pelos experimentos mostram que para o espaço de raio 1,5 utilizar espectros gerados a partir de todo o espaço elementar, gerando evoluções temporais utilizando condições iniciais aleatórias para cada regra, oferece a melhor configuração para se criar o conjunto de espectros utilizados como referência, seja com quantidades fixas ou variáveis. No entanto, comparar os resultados de classificar o espaço elementar e o de raio 1,5, força a conclusão de que escalonar o classificador de espaços menores para maiores não oferece resultados tão bons quanto aquele quando tanto os espectros de referência e os utilizados na predição são do mesmo espaço.

A característica local da extração de informações quando se utiliza análise de textura é uma limitação da técnica quando se utiliza um conjunto de espectros para classificar um espaço maior do que aquele usado para gerá-lo. Ou seja, o resultado menos expressivo da classificação do espaço de raio 1,5 se deve à característica local da análise de textura, pois no espaço maior há proporções de configurações de vizinhança distintas do elementar, tornando a classificação menos precisa. Ainda assim a análise de textura se mostrou útil para se obter uma primeira aproximação da classificação de um espaço desconhecido.

Ao se levar em conta os resultados de escalonar o classificador, a despeito das prováveis imprecisões ao se utilizar uma classificação visual como referência, o processo automático permitiu conhecer a classe correta superior a 70% das vezes. Um percentual interessante para espaços ainda maiores, mas que carece de se avaliar a redução ainda maior da acurácia nestes espaços, caso se limite o conjunto de espectros aos gerados a partir do espaço elementar.

Capítulo 6

Conclusão

Este trabalho abordou duas estratégias de aprendizado. A primeira delas foi baseada em redes neurais, onde o conhecimento está representado pelos pesos sinápticos da rede após o treinamento. Antes disso, apesar de os pesos estarem presentes na estrutura da rede, não representam algum conhecimento, sendo apenas valores aleatórios de inicialização. Como neste trabalho foi usada rede neural convolucional, o conhecimento está representado nos valores dos vários filtros utilizados pelas camadas convolucionais e ainda nos pesos sinápticos das camadas densas.

A segunda abordagem foi a construção de bases de conhecimento que associa informações extraídas das imagens às regras do espaço elementar, abordagem que utilizou a variação da entropia e espectros das configurações de vizinhança. Esta abordagem necessita ser complementada por um algoritmo de busca na base de conhecimento por elementos mais semelhantes possíveis àquele que se busca classificar. Uma vez que o formato utilizado para a base de conhecimento foram vetores associados às informações de classe das regras, foi escolhido o algoritmo k -NN, com $k = 1$, para a busca de elementos semelhantes ao que se propunha classificar.

Fato importante a destacar é que apesar dos resultados, ainda não se obteve uma forma genérica de classificação automática. Os classificadores desenvolvidos aqui possuem grande sensibilidade ao ser aplicada a espaços maiores que aquele usado no treinamento. A classificação visual feita para o espaço de raio 1,5 também contribuiu para a queda de acurácia neste espaço, pois ela insere erros de classificação devido ao processo ma-

nual. Além disso, todas as abordagens utilizadas usaram informações locais aos pixels das imagens binárias, representantes das evoluções temporais, todas tendo sensibilidade em relação às configurações de vizinhanças aprendidas. Isso foi feito, por exemplo, na classificação usando a variação da entropia, adaptando a proposta de Wuensche (1998) é avaliada no Capítulo 3. Tal abordagem não foi capaz de classificar corretamente o espaço elementar em sua totalidade, levando em conta o comportamento típico. O classificador foi capaz de classificar o espaço elementar com acurácia global de 90,71%. Os resultados das métricas avaliadas para o classificador mostrou fragilidades, principalmente para a classe 4. Ao aplicar o classificador ao espaço de raio 1,5 a acurácia global apresenta queda acentuada, muito provavelmente devido aos padrões visuais encontrados no espaço de raio 1,5 que não estão no elementar, espaço de referência para os classificadores.

O classificador baseado em rede neural convolucional, abordado no Capítulo 4, classificou o espaço elementar com 100% de precisão, ao levar em conta o comportamento típico das regras. E ao avaliar a qualidade do classificador individual de cada evolução temporal submetida à rede, alcançou 99,02% de acurácia global. Assim como a variação da entropia é sensível às características locais aos pixels na imagem binária, as redes neurais convolucionais também o são, pois extraem automaticamente tais características locais.

Neste caso, o uso do classificador no espaço de raio 1,5 apresentou queda significativa da acurácia global, que atingiu 54,68%; além disso, os resultados das métricas individuais mostram fragilidades, principalmente para a classe 4. Apesar de esta classe ter alcançado a maior acurácia individual de 87,32%, ela apresenta valores distantes do ideal para a taxa de verdadeiros positivos, o valor preditivo positivo e para as taxas falso negativo e falsa descoberta. Para as outras classes, situação semelhante se apresenta para outras métricas, o que justifica a queda significativa da acurácia global.

Em Silva et al. (2016) uma variação do algoritmo LBP foi utilizada para extrair características das texturas de evoluções temporais, usadas para criar vetores de característica que serviram de base para a utilização do algoritmo k -NN, obtendo-se 96,35% de acurácia. Também foi utilizada a transformada de Fourier para gerar espectros das evoluções temporais e deles também criar vetores de características, que forma uma base de comparação para o algoritmo k -NN, obtendo-se neste caso 99,42% de acurácia (SILVA et al., 2016).

Observe-se que a acurácia de 99,71% atingida pela análise de textura aqui apresentada foi a melhor entre os quatro experimentos feitos com este classificador, no contexto do espaço elementar. Acrescente que este patamar superior é equivalente ao obtido utilizando o algoritmo LBPV e a transformada de Fourier em (SILVA et al., 2016). No entanto, a melhora mais significativa do processo proposto aqui não diz respeito à acurácia, apesar de se ter obtido pequena melhora. O fato de o algoritmo k -NN ter sido utilizado em (SILVA et al., 2016) como decisor de qual histograma deverá ser associado à evolução temporal submetida ao classificador, implica em uma base de histogramas grande o suficiente para conter informações suficientes das evoluções temporais geradas por uma regra. Assumindo que deverão ser geradas 1000 evoluções temporais para cada regra, a base de comparação para o algoritmo k -NN será de 256.000 histogramas, o que contrasta muito com a proposta deste trabalho, em que apenas 256 espectros são necessários e suficientes, tornando a tarefa de prever a classe da regra geradora de uma evolução temporal, muito mais eficiente em termos computacionais. Ou seja, em situações em que se necessita prever a classe geradora de muitas evoluções temporais, o trabalho seria agilizado.

Dada a possibilidade de se aplicar não apenas a mesma técnica, mas também os mesmos 4 conjuntos de espectros gerados para o espaço elementar, em espaços maiores, foi definido o de raio 1,5 como alvo para avaliar quão escalável seria o processo. No entanto, o espaço 1,5 não possuía classificação prévia da dinâmica de suas regras, impedindo uma avaliação dos resultados. Por isso foi realizada a classificação visual desse espaço, para servir de referência e se repetir os mesmos experimentos feitos no espaço elementar.

Apesar da boa performance no espaço elementar, observou-se uma limitação importante ao se aplicar o classificador ao espaço de raio 1,5. As imprecisões observadas na predição se devem às diferenças nas frequências das configurações de vizinhança existentes entre o espaço elementar e o de raio 1,5. Tais diferenças provocaram falsos positivos, como a classificação de regras da classe 2 em classe 1, pois havia poucos elementos que se repetiam na evolução temporal. Apesar da condição de localidade da análise de textura se mostrar a fonte de tais imprecisões na classificação, a técnica se mostrou útil em obter uma primeira aproximação da classificação de um espaço desconhecido.

Um ponto importante na estratégia de classificação utilizada é o fato de ela não possuir limitações de aplicabilidade quanto ao espaço de regras dos autômatos celulares binários e

unidimensionais, a despeito da acurácia diminuir consideravelmente ao escalar para espaço maiores. Entretanto, e naturalmente, possui limitação quanto a outros ACs não binários ou de dimensões maiores, pois o esforço computacional aumentaria a ponto de inviabilizar o processo. Por exemplo, um AC com 3 estados significaria um espectro como $3^9 = 19.683$ componentes, e o aumento para mais estados apenas torna a computação impraticável.

Uma abordagem possível do problema, para superar a limitação computacional referente ao número de estados, pode ser aplicar a mesma estratégia de classificação, mas mapeando a configuração da vizinhança para um número fixo de componentes no espectro, fixando o custo computacional. Haveria evidente perda de informação quanto menor fosse a proporção entre o número de componentes do vetor e o total de conversões possíveis para o número de estados. Como 19.683 configurações de vizinhança diferentes, um AC com 3 estados perderia informações ao mapear estas configurações para 512 componentes, pois haveria mais de uma configuração mapeada para uma mesma componente. Porém, se esta perda não for suficiente para impedir uma distinção entre os espectros, o método ainda poderia ser aplicado.

Limitando-se aos AC binários e unidimensionais, a limitação de escalonar o classificador pode ser solucionado com a realimentação do processo. Ao identificar a não aderência da classificação retornada para uma evolução temporal, mas sabendo-se a regra que a gerou, é possível gerar um espectro associado a ela e incluí-lo no conjunto de espectros de referência, denominado aqui como F . Um ponto a ser explorado é avaliar como realimentar F para elevar a acurácia da classificação do espaço 1,5.

É importante destacar a similaridade entre a análise de espectro e redes neurais convolucionais, pois para se classificar uma imagem usando redes neurais é necessária a extração de características e a análise de textura pode ser uma das técnicas de extração. No caso das redes neurais convolucionais, o trabalho de extração de características é feito automaticamente pela própria rede de forma local, através dos filtros, similar ao que foi feito utilizando análise de textura.

Uma avaliação de como a acurácia varia conforme se escala o classificador para espaços maiores, bem como a concepção de um meio ótimo de realimentar o processo e melhorar a acurácia, são pontos importantes para se abordar no futuro. Outro ponto importante é a busca de meios para se definir a arquitetura da rede neural, pois atualmente não se

conta com algum algoritmo ou meios analíticos para sua definição. Ainda assim o trabalho desenvolvido oferece meios de se extrair informações de espaços desconhecidos utilizando técnicas com custo computacional que permite sua execução em tempo factível.

Referências Bibliográficas

ACHARYA, T.; RAY, A. K. Image Processing - Principles and Applications. New York, NY, USA: Wiley-Interscience, 2005. ISBN 0471719986.

AGGARWAL, C. C. Data Classification: Algorithms and Applications 1st. Chapman & Hall/CRC 2014. ISBN: 1466586745, 9781466586741

ALBREGTSEN, F. Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices. 5 Image Processing Laboratory. Department of Informatics, University of Oslo, 2008.

BANERJEE, I.; BHADURY, T. Self-medication practice among undergraduate medical students in a tertiary care medical college, West Bengal. Industrial Psychiatry Journal, v. 18, n. 2, p. 127–131, 2009. Disponível em: <http://www.industrialpsychiatry.org/article.asp?issn=0972-6748;year=2009;volume=18;issue=2;spage=127;epage=131;aulast=Banerjee;t=6>.

BATISTA, G.; PRATI, R.; MONARD, M. C. Balancing Strategies and Class Overlapping. Advances in Intelligent Data Analysis VI, 6th International Symposium on Intelligent Data Analysis, Madrid, Spain, September 8-10, pp.24-35. 2005.

COVER, T.; HART, P. Nearest neighbor pattern classification. IEEE Trans. Inf. Theor., IEEE Press, Piscataway, NJ, USA, v. 13, n. 1, p. 21–27, set. 2006. ISSN 0018-9448. Disponível em: <https://doi.org/10.1109/TIT.1967.1053964>.

CULIK II, K.; YU, S. Undecidability of CA classification schemes. Complex Syst., Complex Systems Publications, Inc., Champaign, IL, USA, v. 2, n. 2, p. 177–190, abr. 1988. ISSN 0891-2513.

- FAWCETT, T. An introduction to ROC analysis. *Pattern Recogn. Lett.*, Elsevier ScienceInc., New York, NY, USA, v. 27, n. 8, p. 861–874, jun. 2006. ISSN 0167-8655. Disponível em: <http://dx.doi.org/10.1016/j.patrec.2005.10.010>.
- GUO, Z.; ZHANG, L.; ZHANG, D. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 43, n. 3, p. 706–719, mar. 2010. ISSN 0031-3203. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2009.08.017>.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: The MIT Press, 2016. ISBN 0262035618, 9780262035613.
- HAYKIN, S.; VEEN, B. V. *Signals and Systems*. 1st. ed. John Wiley & Sons, Inc., New York, NY, USA, 1998. ISBN 0471138207.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2nd. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. ISBN 0132733501.
- HE, D. chen; WANG, L. Texture unit, texture spectrum, and texture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, v. 28, n. 4, p. 509–512, Jul 1990. ISSN0196-2892.
- HEBB, D. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, 1949. (A Wiley book in clinical psychology). ISBN 9780471367277. Disponível em: <https://books.google.com.br/books?id=dZ0eDiLTwuEC>.
- KARI, J. Theory of cellular automata: A survey. *Theor. Comput. Sci.*, Elsevier Science Publishers Ltd., Essex, UK, v. 334, n. 1-3, p. 3–33, abr. 2005. ISSN 0304-3975. Disponível em: <http://dx.doi.org/10.1016/j.tcs.2004.11.021>.
- LAURIKKALA, J. Improving identification of difficult small classes by balancing class distribution. In: QUAGLINI, S.; BARAHONA, P.; ANDREASSEN, S. (Ed.). *Artificial Intelligence in Medicine*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 63–66. ISBN 978-3-540-48229-1
- LI, W. and PACKARD, N. The Structure of the Elementary Cellular Automata Rule Space. *Complex Systems*, v. 4, p. 281–297. 1990.

- MARTINEZ, G. J. A note on elementary cellular automata classification. *J. Cellular Automata*, p. 233–259, 2013.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. In: [S.l.]: *Bulletin of Mathematical Biophysics*, 1943. v. 5, p. 115–133.
- METZ, C. E. CE: Basic principles of ROC analysis. In: *Seminars in Nuclear Medicine*. [S.l.: s.n.], 1978. p. 8–283.
- MINSKY, M.; PAPER, S. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press, 1969.
- MUCHERINO, A.; PAPAJOGEORGIOU, P. J.; PARDALOS, P. M. *Data Mining in Agriculture*. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2009. ISBN 0387886141,9780387886145.
- NEUMANN, J. V. *Theory of Self-Reproducing Automata*. Champaign, IL, USA: University of Illinois Press, 1966.
- NOGUEIRA, M. A. Classificação automática baseada em rede neural convolucional da dinâmica de autômatos celulares binários unidimensionais. *Caderno de Trabalhos Completos e Resumos: VI Encontro Regional de Matemática Aplicada e Computacional (realizado na UNESP, Bauru, SP, de 17 a 19 de junho de 2019)*, pp. 91-97, 2019.
- NOGUEIRA, M.A. and DE OLIVEIRA, P.P.B. “Automatic classification of the dynamical behavior of one-dimensional binary cellular automata by means of the spectrum of neighborhood configurations”, *International Journal of Natural Computing Research*, 8(4):41-61, 2019. DOI: <https://doi.org/10.4018/IJNCR.2019100104>.
- PATIL, S.; NEMADE, V.; SONI, P. K. Predictive modelling for credit card fraud detection using data analytics. *Procedia Computer Science*, v. 132, p. 385–395, 2018. ISSN 1877-0509. *International Conference on Computational Intelligence and Data Science*. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1877050918309347>.
- PANIGRAHI, S.; NANDA, A.; SWARNKAR, T. Deep learning approach for image classification. In: *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*. [S.l.: s.n.], 2018. p. 511–516.

MACHICAO, J.; RIBAS, L. C.; SCABINI, L. F. S.; BRUNO, O. M. Cellular automata rule characterization and classification using texture descriptors. *Physica A: Statistical Mechanics and its Applications*, v. 497, n. C, p. 109–117, 2018. Disponível em: <https://ideas.repec.org/a/eee/phsmap-/v497y2018icp109-117.html>.

ROCHESTER, N.; HOLLAND, J. H.; HAIBT, L. H.; DUDA W. L. Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, v. 2, n. 3, p. 80–93, September 1956. ISSN 0096-1000.

RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. Third. Upper Saddle River, NJ: Prentice Hall, 2010. (Series in Artificial Intelligence). Disponível em: <http://aima.cs.berkeley.edu/>.

SARKAR, P. A brief history of cellular automata. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 32, n. 1, p. 80–107, mar. 2000. ISSN 0360-0300. Disponível em: <http://doi.acm.org/10.1145/349194.349202>.

SILVA, N. R. da et al. Classification of cellular automata through texture analysis. *Information Sciences*, v. 370-371, p. 33–49, 2016. ISSN 0020-0255. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0020025516304881>.

STOREY, J. D. A direct approach to false discovery rates. 2002

SRIVASTAVA, Nitish; HINTON, Geoffrey; KRIZHEVSKY, Alex; SUTSKEVER, Ilya; SALAKHUTDINOV, Ruslan. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, JMLR.org, v. 15, n. 1, p. 1929–1958, jan. 2014. ISSN 1532-4435. Disponível em: <http://dl.acm.org/citation.cfm?id=2627435.2670313>.

WOLFRAM, S. Universality and complexity in cellular automata. *Physica D: Nonlinear Phenomena*, v. 10, n. 1-2, p. 1–37, 1984.

WOLFRAM, S. *A New Kind of Science*. Champaign, Illinois, US, United States: Wolfram Media Inc., 2002. ISBN 1-57955-008-8.

WUENSCHÉ, A. *Classifying Cellular Automata Automatically*. [S.l.], 1998. Disponível em: <https://EconPapers.repec.org/RePEc:wop:safiwop:98-02-018>.