

---

# Análise de Novas Abordagens para Mineração de Regras de Classificação Utilizando Algoritmos Genéticos

---

Alexandre Henrick da Silva Alves



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia  
2020



**Alexandre Henrick da Silva Alves**

**Análise de Novas Abordagens para Mineração  
de Regras de Classificação Utilizando  
Algoritmos Genéticos**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Laurence Rodrigues do Amaral

Uberlândia  
2020

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

A474 Alves, Alexandre Henrick da Silva, 1994-  
2020 Análise de Novas Abordagens para Mineração de Regras de  
Classificação Utilizando Algoritmos Genéticos [recurso eletrônico] /  
Alexandre Henrick da Silva Alves. - 2020.

Orientador: Laurence Rodrigues do Amaral.  
Dissertação (Mestrado) - Universidade Federal de Uberlândia,  
Pós-graduação em Ciência da Computação.  
Modo de acesso: Internet.  
Disponível em: <http://doi.org/10.14393/ufu.di.2020.260>  
Inclui bibliografia.

1. Computação. I. Amaral, Laurence Rodrigues do,1978-,  
(Orient.). II. Universidade Federal de Uberlândia. Pós-graduação  
em Ciência da Computação. III. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:  
Gizele Cristine Nunes do Couto - CRB6/2091  
Nelson Marcos Ferreira - CRB6/3074



## UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Ciência da Computação  
Av. João Naves de Ávila, nº 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
Telefone: (34) 3239-4470 - [www.ppgco.facom.ufu.br](http://www.ppgco.facom.ufu.br) - [cpgfacom@ufu.br](mailto:cpgfacom@ufu.br)



### ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação			
Defesa de:	Dissertação de Mestrado Acadêmico, 07/2020, PPGCO			
Data:	20 de fevereiro de 2020	Hora de início:	<b>09hr26min</b>	Hora de encerramento:
Matrícula do Discente:	11812CCP003			
Nome do Discente:	Alexandre Henrick da Silva Alves			
Título do Trabalho:	Análise de Novas Abordagens para Mineração de Regras de Classificação Utilizando Algoritmos Genéticos			
Área de concentração:	Ciência da Computação			
Linha de pesquisa:	Inteligência Artificial			
Projeto de Pesquisa de vinculação:	-			

Reuniu-se na sala 1B132, Bloco 1B, Campus Santa Mônica, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Murillo Guimarães Carneiro- FACOM/UFU, Carolina Ribeiro Xavier - DECOMP/UFSJ e Laurence Rodrigues do Amaral - FACOM/UFU, orientador do candidato.

Ressalta-se que a Prof.<sup>a</sup> Dr.<sup>a</sup> Carolina Ribeiro Xavier participou da defesa por meio de videoconferência desde a cidade de São de João Del Rei - MG. Os outros membros da banca e o aluno participaram in loco.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Laurence Rodrigues do Amaral, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

#### Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Carolina Ribeiro Xavier, Usuário Externo**, em 21/02/2020, às 10:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

Documento assinado eletronicamente por **Laurence Rodrigues do Amaral, Professor(a) do Magistério Superior**, em 21/02/2020, às 14:44, conforme horário oficial de Brasília, com

fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Murillo Guimarães Carneiro, Professor(a) do Magistério Superior**, em 21/02/2020, às 15:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1877290** e o código CRC **48801A8F**.

*Dedico este trabalho a todos os meus familiares  
e a minha namorada Lorrane.*



---

## Agradecimentos

Agradeço primeiramente a todos os meus familiares por todo apoio imprescindível. À minha namorada Lorrane, por estar sempre ao meu lado mesmo nos momentos mais difíceis. À meu orientador, professor Laurence, pelo exemplo de competência e profissionalismo. Aos professores do PPGCO-FACOM, pelo ensino de qualidade e o compromisso de formar profissionais de alto nível. Aos secretários Erisvaldo e Sônia, pelo excelente serviço prestado. Aos colegas, professores e técnicos do Laboratório de Bioinformática e Análises Moleculares (LBAM) campus Patos de Minas. Agradeço também a CAPES, pelo importante apoio financeiro. Todos vocês contribuiram grandiosamente para que este trabalho fosse concluído. Muito obrigado!



*“Só o que está morto não muda.”*  
*(Edson Marques)*



---

## Resumo

A tarefa de classificação está entre as mais utilizadas em Mineração de Dados e é amplamente pesquisada nos dias atuais. Vários trabalhos já foram desenvolvidos utilizando Algoritmos Genéticos para tarefa de classificação através da evolução de regras de classificação do tipo SE-ENTÃO e bons resultados já foram obtidos. Esses métodos geralmente utilizam a mesma estrutura cromossômica, números inteiros e reais, e essa estrutura pode impor algumas limitações na sua operação. Além da representação cromossômica, esses métodos também utilizam a mesma abordagem para escolha dos atributos para compor as regras. Nesse trabalho, foram propostos dois novos métodos, chamados de BIN-NLCEE e IG-CEE. O BIN-NLCEE utiliza uma nova estrutura cromossômica, através de valores binários. Já o método IG-CEE utiliza uma medida de avaliação de atributos, o *Information Gain*, para selecionar os atributos que podem compor as regras. Foram utilizados 4 *datasets* de domínio médico para a validação do BIN-NLCEE e 3 *datasets* sintéticos para a validação do IG-CEE. Ambos foram comparados com seus métodos de origem e outros 4 classificadores tradicionais (J48, IBK, *Naive Bayes* e SVM). Os resultados mostraram que os métodos propostos foram capazes de gerar melhores valores de *fitness* e melhores taxas de convergência.

**Palavras-chave:** Algoritmos Genéticos. Classificação de dados. Regras SE-ENTÃO. Cromossomo Binário. Ganho de Informação.



---

## Abstract

The classification task is among the most used in Data Mining and is widely researched nowadays. Several works have already been developed using Genetic Algorithms for classification tasks through the evolution of IF-THEN classification rules and good results have been obtained. These methods often use the same chromosome structure, integer, and real values, and this structure may impose some limitations on their operation. Also, these methods use the same approach for choosing the attributes that will compose the rules. In this work, two new methods were proposed, called BIN-NLCEE and IG-CEE. BIN-NLCEE uses a new chromosomal structure by binary values. The IG-CEE method uses an attribute evaluation measure, called Information Gain, to select the attributes that can compose the rules. Four medical domain datasets were used for BIN-NLCEE validation and 3 synthetic datasets for IG-CEE validation. Both were compared with their source methods and 4 other traditional classifiers (J48, IBK, Naive Bayes and SVM). The results showed that the proposed methods were able to generate better fitness values and better convergence rates.

**Keywords:** Genetic Algorithms. Data classification. IF-THEN Rules. Binary Chromosome. Information Gain.



---

# Lista de ilustrações

Figura 1 – Ciclo de execução de um AG . . . . .	30
Figura 2 – Representação binária . . . . .	31
Figura 3 – Representação por vetor de pesos . . . . .	31
Figura 4 – Seleção por roleta . . . . .	34
Figura 5 – <i>Crossover</i> 1 ponto . . . . .	35
Figura 6 – <i>Crossover</i> 2 pontos . . . . .	35
Figura 7 – Mutação binária . . . . .	36
Figura 8 – Mutação inteiros . . . . .	36
Figura 9 – Mapeamento de atributos para classe. Adaptado de (TAN, 2018) . . .	39
Figura 10 – Validação cruzada utilizando a estratégia <i>k-fold</i> . . . . .	40
Figura 11 – Exemplo de entropia dos conjuntos . . . . .	43
Figura 12 – Cromossomo CEE (AMARAL; HRUSCHKA, 2010) . . . . .	46
Figura 13 – Exemplo de cromossomo do CEE. Adaptado de (AMARAL; HRUSCHKA, 2011) . . . . .	46
Figura 14 – Cromossomo NLCEE (AMARAL; HRUSCHKA, 2011) . . . . .	48
Figura 15 – Exemplo de cromossomo do NLCEE. Adaptado de (AMARAL; HRUSCHKA, 2011) . . . . .	48
Figura 16 – Cromossomo do GASD. Adaptado de (AMARAL; JR, 2014) . . . .	50
Figura 17 – Diagrama dos trabalhos desenvolvidos . . . . .	54
Figura 18 – Conversão do gene binário . . . . .	56
Figura 19 – Ordenação dos atributos de acordo com o ganho de informação . . . .	57
Figura 20 – Gráficos de convergência dos métodos baseados em AG para o <i>breast-w dataset</i> . . . . .	65
Figura 21 – Gráficos de convergência dos métodos baseados em AG para o <i>diabetes dataset</i> . . . . .	65
Figura 22 – Gráficos de convergência dos métodos baseados em AG para o <i>parkinsons dataset</i> . . . . .	65



---

# **Lista de tabelas**

Tabela 1 – Amostra de registros do <i>Iris dataset</i> . . . . .	38
Tabela 2 – Resultados da classificação dos registros da amostra do <i>Iris dataset</i> . .	41
Tabela 3 – Ganho de Informação de cada atributo . . . . .	44
Tabela 4 – Resultados obtidos pelo CEE usando <i>10-fold</i> . Adaptado de (AMARAL; HRUSCHKA, 2010) . . . . .	47
Tabela 5 – Exemplo de regras geradas pelo CEE. Adaptado de (AMARAL; HRUSCHKA, 2010). . . . .	47
Tabela 6 – Resultados obtidos pelo NLCEE usando <i>10-fold</i> . Adaptado de (AMARAL; HRUSCHKA, 2011) . . . . .	49
Tabela 7 – Comparação dos resultados obtidos pelo HEE e outros classificadores. Adaptado de (Silva; de Souza Ribeiro; do Amaral, 2013). . . . .	50
Tabela 8 – Resultados do AG utilizando o operador <i>transgenic</i> . . . . .	51
Tabela 9 – Configurações de parâmetros utilizadas pelos AGs . . . . .	59
Tabela 10 – Distribuição de registros por classe do <i>diabetes dataset</i> . . . . .	60
Tabela 11 – Distribuição de registros por classe do <i>breast-w</i> . . . . .	60
Tabela 12 – Distribuição de registros por classe do <i>parkinson dataset</i> . . . . .	61
Tabela 13 – Distribuição de registros por classe do <i>mammography dataset</i> . . . . .	61
Tabela 14 – Distribuição de registros dos <i>datasets</i> $D_1$ , $D_2$ e $D_3$ . . . . .	62
Tabela 15 – Atributos relevantes para cada <i>dataset</i> . . . . .	62
Tabela 16 – Melhores configurações de parâmetro para cada método e para cada conjunto de dados. . . . .	62
Tabela 17 – Média de <i>fitness</i> das 100 execuções (com <i>seeds</i> distintas) para cada <i>dataset</i> e uma classe específica. . . . .	63
Tabela 18 – Média de <i>fitness</i> das 100 execuções (com <i>seeds</i> distintas) para cada <i>dataset</i> e uma classe específica. . . . .	63
Tabela 19 – Média de <i>fitness</i> das 100 execuções (com <i>seeds</i> distintas) para cada <i>dataset</i> usando classificadores tradicionais. . . . .	64
Tabela 21 – Melhores configurações para cada método e cada conjunto de dados. .	66

Tabela 20 – Conjunto de regras geradas pelo BIN-NLCEE . . . . .	67
Tabela 22 – Média de <i>fitness</i> das 100 execuções (com seeds distintas) para os <i>datasets</i> $D_1$ , $D_2$ e $D_3$ . . . . .	67
Tabela 23 – Média de <i>fitness</i> das 100 execuções (com sementes distintas) para os <i>datasets</i> $D_1$ , $D_2$ e $D_3$ usando os classificadores tradicionais. . . . .	68
Tabela 24 – Exemplo de regras obtidas pelos métodos IG-CEE e CEE para os <i>datasets</i> sintéticos. . . . .	70
Tabela 25 – Resultados das 100 execuções do BIN-NLCEE com o <i>breast-w dataset</i> nas configurações 1, 2 e 3. . . . .	83
Tabela 26 – Resultados das 100 execuções do BIN-NLCEE com o <i>breast-w dataset</i> nas configurações 4, 5 e 6. . . . .	86
Tabela 27 – Resultados das 100 execuções do BIN-NLCEE com o <i>diabetes dataset</i> nas configurações 1, 2 e 3. . . . .	89
Tabela 28 – Resultados das 100 execuções do BIN-NLCEE com o <i>diabetes dataset</i> nas configurações 3, 4 e 5. . . . .	92
Tabela 29 – Resultados das 100 execuções do BIN-NLCEE com o <i>parkinsons dataset</i> nas configurações 1, 2 e 3. . . . .	95
Tabela 30 – Resultados das 100 execuções do BIN-NLCEE com o <i>parkinsons dataset</i> nas configurações 4, 5 e 6. . . . .	98
Tabela 31 – Resultados das 100 execuções do BIN-NLCEE com o <i>mammography dataset</i> nas configurações 1, 2 e 3. . . . .	101
Tabela 32 – Resultados das 100 execuções do BIN-NLCEE com o <i>mammography dataset</i> nas configurações 4, 5 e 6. . . . .	104
Tabela 33 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D1 nas configurações 1 e 2. . . . .	107
Tabela 34 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D1 nas configurações 3 e 4. . . . .	110
Tabela 35 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D1 nas configurações 5 e 6. . . . .	113
Tabela 36 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D2 nas configurações 1 e 2. . . . .	116
Tabela 37 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D2 nas configurações 3 e 4. . . . .	119
Tabela 38 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D2 nas configurações 5 e 6. . . . .	122
Tabela 39 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D3 nas configurações 1 e 2. . . . .	125
Tabela 40 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D3 nas configurações 3 e 4. . . . .	128

Tabela 41 – Resultados das 100 execuções do IG-CEE com o <i>dataset</i> D3 nas configurações 5 e 6. . . . .	131
-------------------------------------------------------------------------------------------------------------	-----



---

## **Lista de siglas**

**AGs** Algoritmos Genéticos

**AM** Aprendizado de Máquina

**CEE** *Computational Evolutionary Environment*

**FF** Função *Fitness*

**GASD** *Genetic Algorithm for Synthetic Data*

**HEE** *Hybrid Evolutionary Environment*

**IA** Inteligência Artificial

**IG** *Information Gain*

**KDD** *Knowledge Discovery in Database*

**MD** Mineração de Dados

**MDRGA** *Multiple Disjunctions Rule Genetic Algorithm*

**NLCEE** *Nonlinear Computational Evolutionary Environment*

**NB** *Naive Bayes*

**GO** *Gene Ontology*

**Pmut** Probabilidade de mutação

**Pcross** Probabilidade de *crossover*

**Tp** Tamanho da população

**UCI** *University of California at Irvine*



---

# Sumário

1	<b>INTRODUÇÃO</b>	25
1.1	Motivação	26
1.2	Objetivos e Desafios da Pesquisa	27
1.3	Hipótese	27
1.4	Contribuições	28
1.5	Organização da Dissertação	28
2	<b>FUNDAMENTAÇÃO TEÓRICA</b>	29
2.1	Algoritmos Genéticos	29
2.1.1	Codificação dos indivíduos	31
2.1.2	Função <i>Fitness</i>	32
2.1.3	Parâmetros do AG	32
2.1.4	Operadores Genéticos	33
2.2	Mineração de Dados	37
2.2.1	Classificação	37
2.2.2	Seleção de atributos	42
3	<b>TRABALHOS CORRELATOS</b>	45
3.1	Introdução	45
3.2	CEE	45
3.3	NLCEE	48
3.4	HEE	48
3.5	<i>Transgenic</i>	49
4	<b>PROPOSTA</b>	53
4.1	Introdução	53
4.2	BIN-NLCEE	54
4.3	IG-CEE	56

<b>5</b>	<b>EXPERIMENTOS E ANÁLISE DOS RESULTADOS . . . . .</b>	<b>59</b>
5.1	Método para a Avaliação . . . . .	59
5.2	<b>Experimentos . . . . .</b>	<b>62</b>
5.2.1	Resultados BIN-NLCEE . . . . .	62
5.2.2	Resultados IG-CEE . . . . .	66
<b>5.3</b>	<b>Avaliação dos Resultados . . . . .</b>	<b>71</b>
5.3.1	Análise dos resultados do método BIN-NLCEE . . . . .	71
5.3.2	Análise dos resultados do método IG-CEE . . . . .	71
<b>6</b>	<b>CONCLUSÃO . . . . .</b>	<b>73</b>
6.1	Principais Contribuições . . . . .	74
6.2	Trabalhos Futuros . . . . .	75
6.3	Contribuições em Produção Bibliográfica . . . . .	75
	<b>REFERÊNCIAS . . . . .</b>	<b>77</b>
	<b>APÊNDICES</b>	<b>81</b>
	<b>APÊNDICE A – RESULTADOS DAS EXECUÇÕES . . . . .</b>	<b>83</b>
A.1	Resultados das execuções do método BIN-NLCEE . . . . .	83
A.2	Resultados das execuções do método IG-CEE . . . . .	107

# CAPÍTULO 1

---

## Introdução

Os Algoritmos Genéticos (AGs) são métodos computacionais de otimização e busca que se baseiam nos mecanismos da evolução por seleção natural (GOLDBERG, 1989). São bastante utilizados em problemas onde o espaço de busca é muito grande e frequentemente conseguem encontrar boas soluções. Seus operadores são inspirados nos fenômenos da genética e hereditariedade encontrados na natureza (HOLLAND, 1992), permitindo uma exploração do espaço de busca mais eficiente. Por serem bastante eficazes, vem sendo utilizados em diversas áreas para solucionar uma variedade de problemas.

Existe um grande desafio na atualidade que é o processamento de grandes massas de dados, que surgiu como consequência de avanços tecnológicos e da capacidade de extração de grandes volumes de informações. As técnicas tradicionais para tratamento e processamento de informações podem não ser eficientes quando aplicadas a uma grande quantidade de dados. Nesse sentido, a Mineração de Dados (MD) surge como uma alternativa, trazendo técnicas computacionais que permitem a extração eficiente de informações. As técnicas de MD conseguem descobrir novos padrões em dados e ainda podem fazer previsões de maneira automática (TAN, 2018).

Segundo (HAN; PEI; KAMBER, 2011), MD é o processo de descoberta de padrões escondidos em dados com o objetivo de obter algum tipo de conhecimento, processo conhecido também por *Knowledge Discovery in Database* (KDD). A MD possui diversas tarefas, sendo a classificação uma das mais importantes e utilizadas. Por meio dessa tarefa é possível ler um conjunto de dados e atribuir rótulos aos seus registros, através do mapeamento de suas características (TAN, 2018).

No processo para encontrar padrões em bases de dados, a MD utiliza algoritmos de Aprendizado de Máquina (AM), algoritmos esses que vem sendo constantemente utilizados em diferentes áreas devido a sua eficiência para trabalhar com grandes quantidades de dados. AM é uma sub-área de estudo da Inteligência Artificial (IA) que tem por objetivo construir algoritmos que resolvam problemas complexos utilizando uma abordagem de aprendizado por experiência. Segundo (MITCHELL, 1997), um algoritmo de AM é aquele que, através da experiência, consegue aperfeiçoar seu desempenho em uma determinada

tarefa.

Ao longo dos anos, diversos trabalhos vem sendo desenvolvidos utilizando AGs para tarefa de classificação, como pode ser visto em (FIDELIS; LOPES; FREITAS, 2000; CARVALHO; FREITAS, 2002; FREITAS, 2003; AMARAL et al., 2008; AMARAL; HRUSCHKA, 2010; AMARAL; HRUSCHKA, 2011; do Amaral; Junior, 2012; do Amaral; Hruschka, 2011; Silva; de Souza Ribeiro; do Amaral, 2013; FREITAS, 2013; AMARAL; JR, 2014; FREITAS, 2009; MATOS; AMARAL, 2018). Esses AGs são responsáveis por minrar regras de classificação do tipo SE-ENTÃO, com o objetivo de produzir boas taxas de classificação e regras de fácil interpretação. No entanto, em todos esses trabalhos utiliza-se números inteiros ou reais na representação cromossômica, o que pode comprometer a convergência do algoritmo e o funcionamento de alguns operadores. Nesse trabalho foi investigado os efeitos do uso da representação binária em AGs para mineração de regras de classificação. E ainda, com o intuito de aprimorar a capacidade do método de escolher os atributos mais relevantes para estarem presentes na regra, foi investigado uma nova estratégia baseada em uma medida de seleção de atributos chamada *Information Gain* (IG).

## 1.1 Motivação

Em (FIDELIS; LOPES; FREITAS, 2000) foi proposto uma estrutura cromossômica representada por números inteiros e em trabalhos posteriores essa representação foi mantida. No trabalho de (AMARAL; HRUSCHKA, 2010), que procede (FIDELIS; LOPES; FREITAS, 2000), foi proposto um AG chamado *Computational Evolutionary Environment* (CEE), para mineração de regras de classificação em um conjunto de dados de domínio biológico. No trabalho de (AMARAL; HRUSCHKA, 2011) foi proposto um AG que estende o método CEE, chamado *Nonlinear Computational Evolutionary Environment* (NLCEE), que produz regras de classificação para *datasets* (conjunto de dados) com distribuição não linear. (MATOS; AMARAL, 2018) estenderam o NLCEE, propondo o *Multiple Disjunctions Rule Genetic Algorithm* (MDRGA), também para classificação de dados não lineares. O MDRGA obteve boas taxas de classificação, mas ao custo de uma representação cromossômica extremamente complexa. Além de manter a representação por números inteiros ou reais em todos os métodos citados, a forma como um atributo é selecionado para compor a regra também se manteve. Portanto, levando em consideração esses problemas apontados, foi investigado quais seriam os impactos de modificações nos métodos NLCEE e CEE, visando gerar melhorias nos algoritmos e criando assim novas versões para os mesmos.

## 1.2 Objetivos e Desafios da Pesquisa

Esse trabalho tem como objetivo geral implementar novas versões para os métodos CEE e NLCEE, utilizando representação binária e um novo mecanismo para seleção de atributos codificado em seus genes. Os objetivos são motivados pelos fatores apontados na subseção 1.1.

Para isso, o trabalho tem como objetivos específicos:

- Desenvolver uma estratégia de representação binária para o método NLCEE;
- Desenvolver uma estratégia baseada em seleção de atributos para o método CEE;
- Executar e comparar os resultados das novas versões dos métodos com os trabalhos correlatos e outros classificadores;
- Comparar os resultados dos algoritmos em relação a taxa de classificação;
- Comparar a convergência dos algoritmos.

## 1.3 Hipótese

Com os problemas e as motivações apontados na subseção 1.1, algumas hipóteses podem ser levantadas, como, (a) a representação cromossômica utilizada nos métodos CEE e NLCEE pode diminuir o desempenho dos métodos em relação a convergência e capacidade de classificação; (b) a abordagem utilizada pelos métodos CEE e NLCEE para selecionar os atributos pode comprometer sua capacidade de gerar regras mais comprehensíveis para o especialista.

Portanto, as perguntas associadas as hipóteses são:

1. O uso da representação binária pode melhorar capacidade de classificação do algoritmo?
2. O uso de representação binária pode apresentar melhor convergência?
3. Um mecanismo de seleção de atributos codificado nos genes do algoritmo pode gerar melhores taxas de classificação?
4. Um mecanismo de seleção de atributos codificado nos genes do algoritmo pode gerar regras mais simples?
5. Um mecanismo de seleção de atributos codificado nos genes do algoritmo pode, efetivamente, reduzir o número de atributos presentes na regra?

## 1.4 Contribuições

No trabalho em questão foi investigado duas novas estratégias em AGs para mineração de regras de classificação. Com o desenvolvimento do mesmo, duas grandes contribuições podem ser mencionadas: (a) uma nova representação cromossômica para AGs usados para classificação de dados; (b) um novo mecanismo, baseado em IG, para selecionar os atributos que farão parte da regra. Portanto, as duas novas estratégias propostas neste trabalho contribuem para o melhoramento contínuo dos métodos abordados.

## 1.5 Organização da Dissertação

No Capítulo 2 são apresentados os principais conceitos teóricos utilizados para o desenvolvimento do trabalho. No Capítulo 3 os principais trabalhos correlatos são abordados. O Capítulo 4 apresenta com detalhes os dois métodos propostos. No Capítulo 5 são apresentadas as configurações de cada experimento e como os mesmos foram conduzidos, assim como os resultados e uma análise para cada método. Por fim, o Capítulo 6 conclui o trabalho apresentando as principais contribuições e os trabalhos futuros.

# CAPÍTULO 2

## Fundamentação Teórica

Nesse capítulo serão apresentados os conceitos utilizados para sustentar o desenvolvimento do trabalho. Nas Subseções 2.1 e 2.2 serão abordados os conceitos teóricos sobre AGs e mineração de dados, respectivamente.

### 2.1 Algoritmos Genéticos

A Computação Evolutiva (CE) é uma subárea de estudo da IA que busca na natureza, mais especificamente nos conceitos da evolução, inspiração para construir sistemas de otimização (MITCHELL, 1998). Programação Evolutiva, Estratégias Evolucionárias e Algoritmos Genéticos são alguns métodos pertencentes à CE. Esses métodos seguem as mesmas premissas da CE com pequenas diferenças em seu funcionamento. No entanto, o método mais utilizado na CE são os AGs, que tiveram seu marco inicial com o trabalho de John Holland (HOLLAND, 1992), onde buscou-se desenvolver sistemas adaptáveis e que, posteriormente, foram aprimorados por David E. Goldberg (1989). Os AGs buscam aplicar os princípios da seleção natural, ou seja, a sobrevivência do mais apto, para encontrar boas soluções em espaços de busca complexos. Atualmente são aplicados em diversos tipos de problemas, tais como:

- Otimização de campanhas publicitárias online (MIRALLES-PECHUÁN; PONCE; MARTÍNEZ-VILLASEÑOR, 2018);
- Seleção de atributos (DONG et al., 2018);
- Decisão de empréstimos bancários (METAWA; HASSAN; ELHOSENY, 2017);
- Evolução de redes neurais (SUCH et al., 2017).

No processo para encontrar boas soluções, um AG utiliza uma estrutura chamada cromossomo, também conhecida como indivíduo, que representam possíveis soluções no

espaço de busca do problema. Esses cromossomos são inicializados aleatoriamente formando uma população inicial e são submetidos a alguns operadores, conhecidos como operadores genéticos (GOLDBERG, 1989). Esses operadores realizam a troca de informações genéticas dos indivíduos (reprodução), a mutação e a seleção daqueles mais aptos.

Um AG passa por diversas gerações ou iterações, onde novas populações são geradas através dos operadores genéticos, sempre preservando aqueles indivíduos mais aptos. Para verificar a aptidão de um indivíduo, o AG utiliza uma função de avaliação que calcula o quanto bom é aquele indivíduo para resolver o problema que está sendo tratado. Dessa maneira os mais aptos possuem mais chances de reproduzir e perpetuar sua informação genética nas gerações seguintes. Outro ponto a ser levado em consideração é o critério de parada do algoritmo, que deve ser escolhido antes da execução do AG. Este critério pode ser, por exemplo, se o algoritmo encontrou a melhor solução possível ou, como na maioria das vezes, o número de gerações. A Figura 1 ilustra o ciclo de execução de um AG.

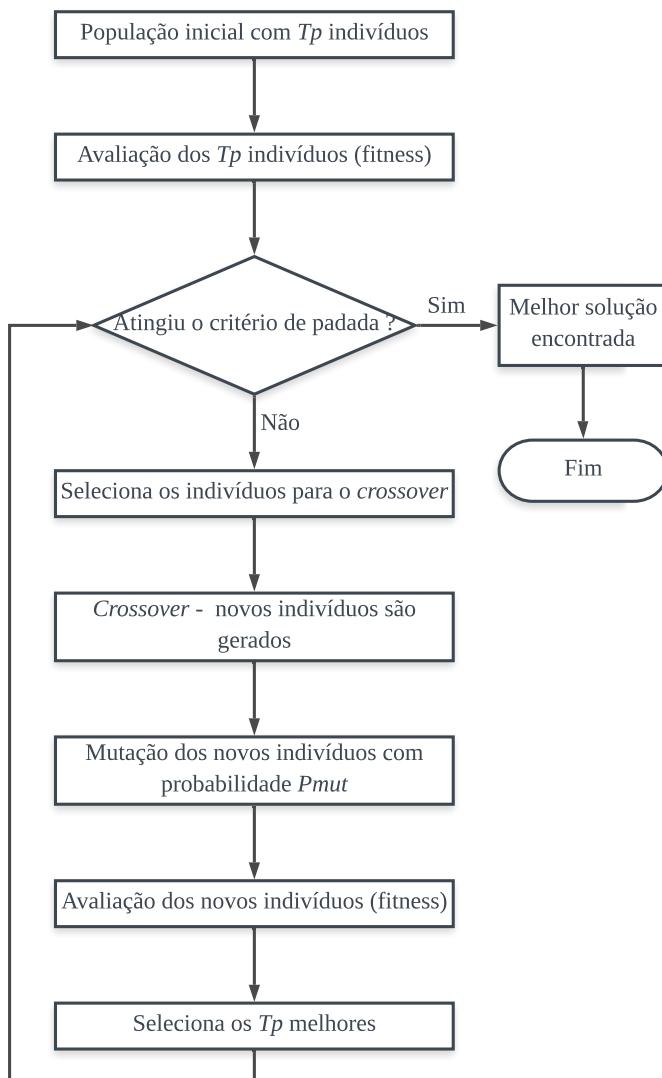


Figura 1 – Ciclo de execução de um AG

Os AGs são algoritmos não-determinísticos (AMARAL, 2007), que através da aleato-

riedade, conseguem explorar informações históricas para encontrar novos pontos de busca e assim aumentar seu desempenho (GOLDBERG, 1989). São vastamente utilizados em problemas onde o espaço de busca é muito complexo e conseguem encontrar boas soluções, principalmente por realizar a busca de forma global, minimizando as chances de cair em ótimos locais (SRINIVAS; DEB, 1994). Dentro de uma população, cada indivíduo representa um ponto no espaço de busca do problema (FILHO, 2001) e a abordagem utilizada por um AG busca evoluir paralelamente esses indivíduos através da aplicação dos operadores com o objetivo de convergir para a solução ótima (COLEY, 1999). Esses indivíduos devem ser codificados em estruturas que expressem uma solução para o problema. A Subseção 2.1.1 apresenta algumas formas de codificação dos indivíduos.

### 2.1.1 Codificação dos indivíduos

A forma como um cromossomo será representado compreende uma parte importante do AG. A estrutura do cromossomo deve estar alinhada às especificações do problema que está sendo tratado, caso contrário suas soluções não serão adequadas para o problema. Os AGs permitem uma grande flexibilidade na representação dos cromossomos e portanto é possível encontrar vários tipos, de acordo com o domínio da aplicação e suas necessidades, números inteiros, binários, reais e codificação de permutação (GEN; LIN, 2007).

Um cromossomo é composto por genes, que são unidades que carregam valores do domínio da aplicação. Na Figura 2, é apresentado um cromossomo que utiliza representação binária, uma abordagem bastante comum para codificar um indivíduo, por ser simples e trazer facilidades na aplicação dos operadores. Nesse tipo de codificação, cada gene pode possuir apenas um valor, sendo 0 ou 1.

0	1	1	1	0	1
---	---	---	---	---	---

Figura 2 – Representação binária

Esse tipo de representação é uma abstração das possíveis soluções para o problema e é bastante utilizada devido a sua simplicidade. No entanto, é possível ainda usar características mais explícitas para representar uma solução. Um exemplo seria um AG que tem como objetivo encontrar os valores de peso de uma rede neural articial que minimize o seu erro médio. Para isso, uma estratégia seria utilizar um vetor de pesos ( $W=[w_1, w_2, w_3, \dots, w_n]$ ) para codificar o indivíduo, como apresentado na Figura 3. Dessa forma, o AG irá evoluir para encontrar o vetor de pesos que minimize o erro da rede.

$W_1$	$W_2$	$W_3$	...	...	$W_n$
-------	-------	-------	-----	-----	-------

Figura 3 – Representação por vetor de pesos

### 2.1.2 Função *Fitness*

A Função *Fitness* (FF) é utilizada como uma maneira de avaliar cada indivíduo da população, avaliação essa que tem por objetivo dizer o quanto bom esse indivíduo é para resolver o problema que está sendo tratado (MITCHELL, 1998). A FF é utilizada durante a etapa de avaliação e atribui um valor relativo à cada cromossomo. Esse valor é o que define a aptidão de cada cromossomo, sendo esta utilizada posteriormente para definir se o indivíduo participará ou não das próximas gerações.

Um exemplo de FF é descrita a seguir para encontrar o mínimo de uma função. A função utilizada é a soma dos quadrados (MOLGA; SMUTNICKI, 2005), e é definida como:

$$f(x) = \sum_{i=1}^d ix_i^2, \quad (1)$$

onde  $d$  é a quantidade de dimensões e o mínimo da função é 0 para  $x^* = (0, \dots, 0)$ . Dessa forma o AG deve possuir uma FF que o faça convergir para  $f(x) = 0$ . A FF poderia ser, por exemplo:

$$FF = 100 - \sum_{i=1}^d ix_i^2 \quad (2)$$

Dessa maneira, quanto mais próximo de 0 for a soma dos quadrados maior será o valor da FF. Juntamente com a codificação dos indivíduos, a FF também é uma etapa importante no projeto do algoritmo e deve ser implementada com cuidado. Segundo (OLIVEIRA et al., 2004), a FF pode consumir até 95% do tempo de processamento de um AG.

### 2.1.3 Parâmetros do AG

Antes da execução de um AG, é necessário que sejam definidos alguns parâmetros, tais como:

- Tamanho da população (Tp);
- Probabilidade de mutação (Pmut);
- Probabilidade de *crossover* (Pcross);
- Número de gerações (iterações).

O Tp é o parâmetro que define o tamanho da população, ou seja, quantos indivíduos estarão presentes nela. Isso é, quanto mais indivíduos, mais pontos estarão presentes no espaço de busca do problema. A população em cada geração sempre manterá Tp indivíduos inseridos nela. Já o parâmetro Pmut define a probabilidade dos indivíduos sofrerem

mutação. A mutação geralmente ocorre em cada gene do cromossomo individualmente, portanto o  $P_{mut}$  pode ser tanto a probabilidade do cromossomo sofrer mutação quanto a probabilidade do gene sofrer mutação. A forma como um cromossomo será mutado pode variar bastante de acordo com o problema, mas detalhes serão apresentados na subseção 2.1.4.3. O parâmetro  $P_{cross}$  define a probabilidade de *crossover* dos cromossomos. E por fim, o número de gerações do AG também deve ser passado como parâmetro. A escolha dos parâmetros está diretamente relacionada com a eficiência do AG para encontrar boas soluções. Entende-se que é necessário testar diferentes configurações de parâmetros para obter a mais adequada, sendo essa etapa necessária para qualquer problema, já que uma configuração não irá obter os mesmos resultados para problemas diferentes.

## 2.1.4 Operadores Genéticos

Como apresentado anteriormente, os AGs possuem operadores que funcionam de forma análoga ao processo de seleção dos mais aptos encontrado na natureza. Esses operadores são:

- ❑ Seleção;
- ❑ Reprodução (*Crossover*);
- ❑ Mutação;
- ❑ Reinserção.

Existem diferentes versões para cada operador, que foram desenvolvidas ao longo dos anos. Algumas delas serão detalhadas nas subseções seguintes.

### 2.1.4.1 Seleção

O operador de seleção atua de forma a simular a principal característica da seleção natural: sobrevivência dos mais aptos. Como o próprio nome sugere, este operador seleciona aqueles indivíduos mais bem avaliados pela FF (mais aptos) para reproduzirem (*crossover*) e perpetuarem sua informação genética para as próximas gerações. De maneira geral, os operadores de seleção existentes escolhem com mais frequência aqueles indivíduos mais bem avaliados, mas sem descartar a possibilidade de selecionar os menos aptos, portanto garantindo a variedade genética nas populações (AMARAL, 2007).

Dentre os operadores de seleção mais utilizados está a seleção por roleta (*roulette wheel selection*), proposto por (HOLLAND, 1992). Neste método, cada indivíduo tem probabilidade de ser selecionado proporcionalmente ao seu *fitness*. Considerando que existem  $N$  indivíduos ( $1, 2, \dots, N$ ) e seus valores de *fitness* ( $f_i > 0$ ), a probabilidade do  $i$ ésimo indivíduo ser selecionado é dada por (LIPOWSKI; LIPOWSKA, 2012):

$$p_i = \frac{f_i}{\sum_{i=1}^N f_i} \quad (3)$$

Para ilustrar o funcionamento da roleta, a Figura 4 apresenta um exemplo onde uma população com 4 indivíduos é avaliada. O método simula uma roleta onde cada porção da mesma é proporcional ao *fitness* do indivíduo e, para realizar a seleção é necessário “rodar” a roleta virtualmente. O indivíduo 1 possui valor de *fitness* igual a 8 e portanto 6,78% de chances de ser selecionado. Para o indivíduo 2, que possui *fitness* igual a 35, suas chances são de 29,66%. O indivíduo 3 possui *fitness* igual a 15 e 12,71% de chances de ser selecionado. E por fim, o indivíduo 4 possui *fitness* 60 e 50,85% de chances de ser selecionado. É possível concluir então que o método da roleta prioriza a seleção dos mais aptos mas também dá espaço para aqueles com avaliação baixa.

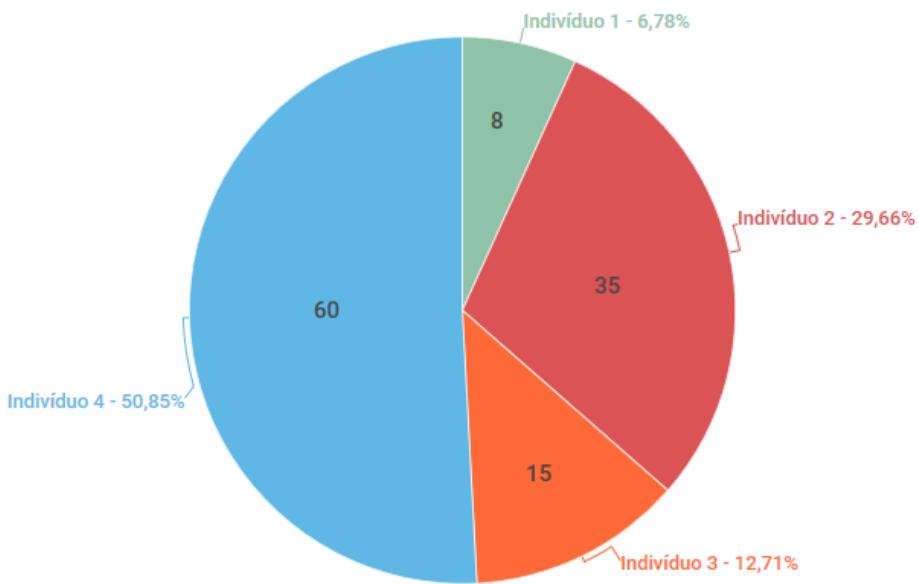


Figura 4 – Seleção por roleta

Outro método bastante utilizado é a seleção por torneio (*tournament*). É um método eficiente e de fácil implementação. Na seleção por torneio,  $t$  indivíduos da população são selecionados aleatoriamente ou através de uma roleta e, dentre eles, aquele com maior *fitness* é selecionado (SOKOLOV; WHITLEY, 2005). O parâmetro  $t$  deve ser definido antes da execução do AG e geralmente são utilizados valores entre 2 e 4. O processo é repetido até que se obtenha o número necessário de indivíduos para a reprodução.

#### 2.1.4.2 Crossover

Seguindo os princípios da seleção natural, os indivíduos escolhidos pelo operador de seleção, conhecidos como pais, irão reproduzir para gerar novos indivíduos filhos para a próxima geração. Em AGs, o operador que simula este fenômeno é o *crossover*. O

*crossover* simula a reprodução sexuada através da troca de informação genética entre indivíduos para explorar o espaço de busca (DEEP; THAKUR, 2007).

No trabalho de Holland (HOLLAND, 1992) é descrito o funcionamento do *crossover* de 1 ponto, sendo esta uma das maneiras mais utilizadas de cruzamento. Além do *crossover* de 1 ponto existem também diversas outras estratégias, como por exemplo o *crossover* de 2 pontos e o uniforme. Dentre todos os métodos de cruzamento, esses são os mais comuns e cada um pode operar de diferentes maneiras, dependendo da representação cromossômica utilizada. A Figura 5 mostra como é realizado o *crossover* de 1 ponto onde dois indivíduos sorteados (através do operador de seleção) para realizarem o cruzamento unem partes de seus genes para formarem dois novos indivíduos filhos, que farão parte de uma nova população. Para realizar esse processo, é sorteado um ponto de corte, onde os cromossomos são divididos em duas partes diferentes e essas são recombinadas formando assim dois novos filhos.

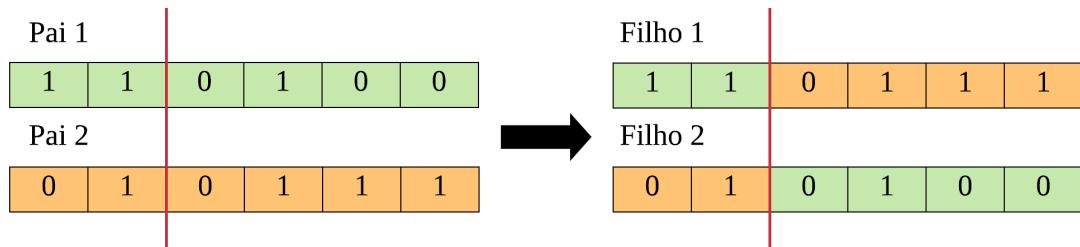


Figura 5 – *Crossover* 1 ponto

Já no *crossover* de 2 pontos, a única diferença está na quantidade de pontos de corte sorteados, neste caso são dois, como apresentado na Figura 6. Essa abordagem não se limita a apenas 2 pontos de corte, quando se tem 2 ou mais pontos de cortes, o método também é conhecido como *crossover* de múltiplos pontos.

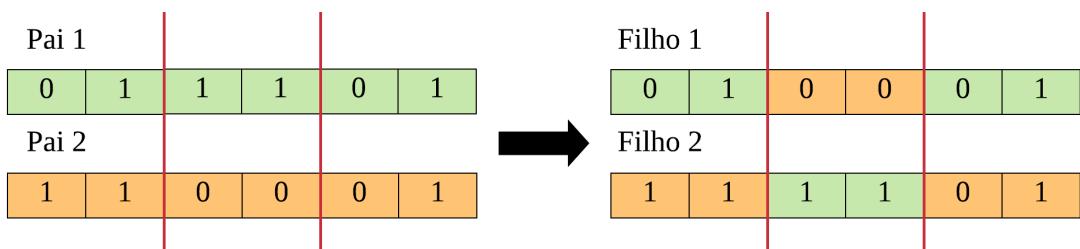


Figura 6 – *Crossover* 2 pontos

#### 2.1.4.3 Mutação

A seleção dos indivíduos prioriza aqueles com maior avaliação visando simular o processo de seleção natural. Isso permite que os candidatos mais aptos perpetuem seus genes nas gerações seguintes. Porém, esse comportamento elitista dos AGs pode gerar alguns problemas, como por exemplo a geração de indivíduos muito parecidos. Indivíduos muito

parecidos podem levar as soluções a ótimos locais, devido à falta de variação na população que não permite a exploração eficiente do espaço de busca.

Uma das formas de contornar esse problema é utilizar o operador de mutação. A mutação atua de maneira simples, alterando randomicamente o valor de um ou mais genes do cromossomo, como pode ser visto na Figura 7. Para se aplicar a mutação, é necessário definir uma taxa de mutação, ou probabilidade de mutação ( $P_{mut}$ ) como parâmetro do AG, esse parâmetro define as chances de um cromossomo ou gene sofrer mutação. Essa mudança garante a variedade genética da população e pode evitar a convergência para ótimos locais.

Diferentes tipos de mutação podem ser aplicados e estão diretamente relacionados ao tipo de codificação do cromossomo. Quando utilizada a codificação binária a mutação simplesmente troca o valor do *bit*, como é mostrado na Figura 7.

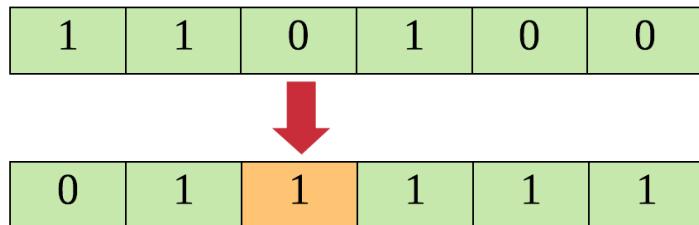


Figura 7 – Mutação binária

No caso do cromossomo estar representado por números inteiros, um exemplo de mutação seria incrementar ou decrementar um valor do gene, como é apresentado na Figura 8. O mesmo poderia ser feito se a representação fosse por números reais

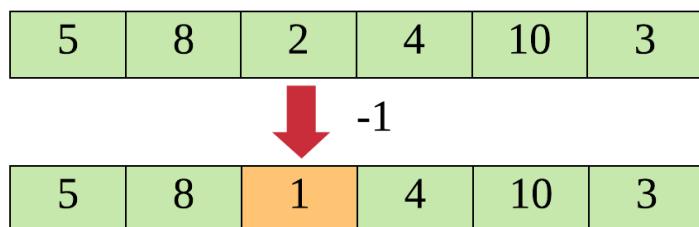


Figura 8 – Mutação inteiros

#### 2.1.4.4 Reinserção

O operador de reinserção é o responsável por escolher quais indivíduos serão inseridos na população da próxima geração. Esta etapa é feita após a seleção, cruzamento, mutação e avaliação da FF. Entre os operadores de reinserção mais conhecidos estão a reinserção pura, o método melhores pais e filhos e o método elitismo (AMARAL, 2007). Na reinserção pura todos os indivíduos da população atual são substituídos, isso é, apenas os filhos passarão para a próxima geração. No método melhores pais e filhos, todos os pais e

filhos da população corrente são colocados em uma população temporária, e aqueles  $T_p$  melhores são escolhidos para a próxima geração. E por fim, no método elitismo os  $n$  melhores indivíduos são escolhidos para fazerem parte da próxima geração. Isso garante que o melhor indivíduo da próxima geração será pelo menos igual ao da geração anterior. O parâmetro  $n$  deve ser definido antes da execução do AG.

## 2.2 Mineração de Dados

Atualmente a informação é obtida com muita velocidade e armazenada em grandes quantidades. Em todos os setores da indústria nota-se uma mudança de paradigma, voltado estritamente para os dados. Essa mudança agregou muito valor na forma como os dados são obtidos e tratados, tornando-os o recurso mais valioso. Não apenas na indústria é possível perceber a mudança mas também na academia, nas mais diversas áreas. Um exemplo disso pode ser encontrado nas áreas biológicas e médicas, onde são utilizadas grandes quantidades de informações para o desenvolvimento de várias soluções que impactam positivamente a área da saúde.

Essa necessidade de obter e processar grandes massas de dados trouxe um grande desafio para a área da Computação. Os métodos mais tradicionais para processamento de dados podem não ser tão eficientes quando o número de informações é muito grande. A computação então possui um papel importante no desenvolvimento de soluções que consigam processar grandes massas de dados de maneira mais rápida, eficiente e automática. A MD é uma área que nasceu dessa necessidade, trazendo diversas técnicas e métodos para o processamento de grandes conjuntos de dados. Segundo TAN (2018), a MD é composta por métodos capazes de encontrar padrões em dados de maneira eficiente e automática, colocando-a como principal alternativa para este desafio.

Um dos motivos da eficiência da MD para tratar grandes quantidades de dados se deve ao fato de se utilizar métodos de AM, métodos estes que conseguem aprender padrões em conjuntos de dados para a descoberta de conhecimento útil. Em AM, os algoritmos aprendem uma determinada tarefa através da experiência, considerando alguma medida de desempenho. Na subseção seguinte será apresentado uma das tarefas mais utilizadas em MD, a classificação.

### 2.2.1 Classificação

A tarefa de classificação é amplamente utilizada em MD e tem como objetivo atribuir algum tipo de rótulo para os registros. Para exemplificar, considere um conjunto de dados com registros referentes a pacientes com câncer de mama. Neste exemplo a função da tarefa de classificação poderia ser, por exemplo, classificar a severidade do câncer como benigno ou maligno, de acordo com as informações de cada paciente, onde benigno e maligno seriam os rótulos (ou classe).

Segundo TAN (2018), a tarefa de classificação é o processo de mapeamento de um conjunto de atributos  $X$  para uma classe  $y$ . De maneira geral, a classificação busca encontrar algum tipo de relação entre esse conjunto  $X$  e a classe  $y$ . Este mapeamento é feito pelo modelo criado pelo algoritmo, como é apresentado na Figura 9. Esse modelo depende do algoritmo que está sendo utilizado e alguns exemplos são:

- ❑ C4.5: Utiliza árvores de decisão como modelo (QUINLAN, 1993);
- ❑ *K-Nearest Neighbor* (KNN): Classifica um registro com a mesma classe que os  $k$  registros mais próximos. Utiliza uma medida de distância para encontrar os “vizinhos” mais próximos (AHA; KIBLER; ALBERT, 1991).
- ❑ *Naive Bayes* (NB): Classificador baseado no Teorema de Bayes. Considera que existe independência entre os atributos (HART, 1973).
- ❑ *Support Vector Machines* (SVM): Maximiza a margem do hiperplano que separa os elementos de cada classe (BOSER; GUYON; VAPNIK, 1992);

Um conjunto de dados é formado por vários registros, sendo esses utilizados para a construção do modelo de classificação. A Tabela 1 apresenta uma amostra do *Iris dataset*, que pode ser encontrado no repositório da *University of California at Irvine* (UCI) (DUA; GRAFF, 2017). O *dataset* em questão é referente a tipos de flores *Iris* e o mesmo possui 3 classes, 4 atributos e 150 registros. Os atributos são: *sepal length* (comprimento da sépala), *sepal width* (largura da sépala), *petal length* (comprimento da pétala) e *petal width* (largura da pétala). E as classes: *Iris Setosa*, *Iris Versicolor* e *Iris Virginica*. Na Tabela 1, cada linha representa um registro.

Tabela 1 – Amostra de registros do *Iris dataset*

Registro	<i>sepal length</i>	<i>sepal width</i>	<i>petal length</i>	<i>petal width</i>	Classe
1	5,1	3,5	1,4	0,2	Iris-setosa
2	5,0	3,3	1,4	0,2	Iris-setosa
3	7,0	3,2	4,7	1,4	Iris-versicolor
4	5,7	2,8	4,1	1,3	Iris-versicolor
5	6,3	3,3	6,0	2,5	Iris-virginica
6	5,9	3,0	5,1	1,8	Iris-virginica

A tarefa de classificação possui duas etapas: treino e teste. Durante a etapa de treino o algoritmo cria seu modelo de classificação através do conjunto de dados (entrada) que é fornecido. Após o treino e já com o modelo gerado, a etapa de teste é realizada, onde uma amostra do conjunto é fornecida para o modelo e sua capacidade de classificação é avaliada usando alguma medida específica.

Para ilustrar o processo de classificação será considerado um modelo que utiliza regras de classificação do tipo SE-ENTÃO. Neste tipo de modelo os valores dos atributos fazem

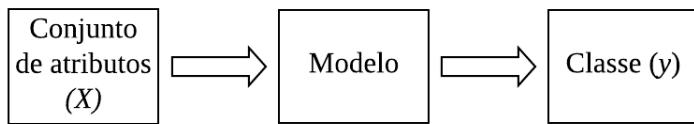


Figura 9 – Mapeamento de atributos para classe. Adaptado de (TAN, 2018)

parte do antecedente da regra (SE), enquanto a parte consequente (ENTÃO) possui a classe. As regras de classificação são geralmente avaliadas usando dois critérios, acurácia e interpretabilidade (FREITAS; LAVINGTON, 1997), ou seja, além de possuirem boas taxas de classificação as regras também devem ser de fácil interpretação. Usando a amostra do *dataset* citado anteriormente, o algoritmo poderia gerar a seguinte regra para uma classe.

$$\text{SE } \text{petal length} > 5 \text{ E } \text{petal width} > 2 \text{ ENTÃO } \text{Iris-virginica}$$

A regra pode ser vista como o modelo que foi gerado pelo classificador através do conjunto de amostras que foi fornecido como entrada e o mapeamento dos atributos (*petal length* e *petal width*) para a classe (*Iris-virginica*). Esse processo é realizado durante a etapa de treino e, posteriormente o desempenho do modelo será avaliado em uma outra amostra do conjunto de dados, etapa esta chamada de teste. Para validar e avaliar um modelo diversos métodos e medidas podem ser utilizadas. Algumas delas serão abordadas nas próximas subseções.

### 2.2.1.1 Validação Cruzada

Como supracitado, o algoritmo cria um modelo baseado nos dados que são conhecidos por ele, etapa essa conhecida como treino. Para que um modelo seja avaliado adequadamente, os dados que foram utilizados para treino não devem ser os mesmos utilizados para teste, pois espera-se que o modelo tenha um bom desempenho em dados ainda não vistos. Portanto, é necessário que sejam utilizados algumas técnicas de validação que sejam capazes de lidar com esse problema. A validação cruzada é bastante utilizada e possui dois métodos principais, o *k-fold* e o *leave-one-out*. No *k-fold*, o conjunto de dados é dividido em  $k$  partições com quantidades de registros proporcionais, onde  $k$  geralmente tem o valor de 5 ou 10, e o algoritmo é treinado em  $k-1$  partições e testado na partição restante (BENGIO; GRANDVALET, 2004). O processo se repete  $k$  vezes até que o teste tenha sido realizado em todas as partições e o resultado final é a média dos  $k$  valores de teste. A Figura 10 apresenta um exemplo usando o método *k-fold* com  $k = 5$ .

Para exemplificar o uso do *k-fold*, considerando que a medida de avaliação usada é a acurácia, após a execução das  $k$  partições de teste, se obtém a média dos resultados, de acordo com a Equação 4.

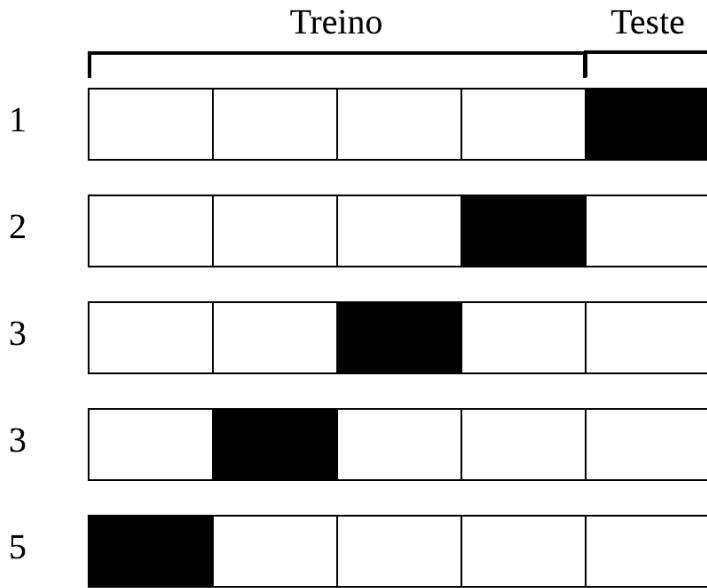


Figura 10 – Validação cruzada utilizando a estratégia  $k$ -fold

$$\frac{1}{k} \sum_{i=1}^k acc_i, \quad (4)$$

onde  $acc_i$  é a acurácia obtida no teste da partição  $i$ . Outro método de validação cruzada bastante conhecido é o *leave-one-out* (WONG, 2015). Diferentemente do  $k$ -fold, o número de partições no *leave-one-out* é igual ao número de registros. Portanto o modelo é testado em 1 registro e treinado nos  $N - 1$  restantes, onde  $N$  é o número de registros no conjunto de dados.

### 2.2.1.2 Medidas de avaliação

Para que o classificador consiga produzir uma melhoria contínua de sua capacidade de predição, seu desempenho deve ser avaliado para cada registro. Essa avaliação é feita de acordo com a quantidade de registros classificados corretamente e incorretamente. Para cada registro existem 4 possíveis resultados:

- ❑ Verdadeiro Positivo (VP): O registro é classificado como sendo da classe  $C$  e realmente pertence a classe  $C$ ;
- ❑ Verdadeiro Negativo (VN): O registro é classificado como não sendo da classe  $C$  e realmente não pertence a classe  $C$ ;
- ❑ Falso Positivo (FP): O registro é classificado como sendo da classe  $C$  mas não pertence a classe  $C$ ;
- ❑ Falso Negativo (FN): O registro é classificado como não sendo da classe  $C$  mas pertence a classe  $C$ .

Após obter os resultados de cada registro, é possível avaliar o desempenho do classificador utilizando diferentes medidas. Existem várias medidas na literatura, no entanto a escolha de qual medida utilizar deve ser feita de acordo com as necessidades do projeto. No entanto, algumas medidas podem ser descritas aqui por serem bastante utilizadas, como por exemplo a acurácia (acc), especificidade (sp) e sensibilidade (se). A acurácia, como mostrado na Equação 5, mostra qual a proporção de previsões corretas, isso é, a proporção de registros classificados corretamente levando em consideração todos os resultados.

$$acc = \frac{VP + VN}{VP + FN + FP + VN} \quad (5)$$

A sensibilidade, por sua vez, apresenta a proporção de verdadeiros positivos, isto é, com que frequência o algoritmo classifica registros como sendo da classe  $C$  e realmente pertencem a esta classe. A Equação 6 define o cálculo da sensibilidade.

$$se = \frac{VP}{VP + FN} \quad (6)$$

Já a especificidade, como pode ser visto na Equação 7, calcula a proporção de verdadeiros negativos, ou seja, a frequência com que o algoritmo classifica os registros como não pertencente a classe  $C$  e que realmente não pertencem a classe.

$$sp = \frac{VN}{VN + FP} \quad (7)$$

Para ilustrar o processo de validação, considere a Tabela 1 como uma amostra utilizada para teste e como modelo gerado a regra apresentada na subseção anterior. A regra foi gerada para classificar registros da classe *Iris-virginica* e a mesma será verificada para cada registro do conjunto de teste. A Tabela 2 apresenta os resultados para cada registro da amostra utilizando a regra supracitada. Na primeira coluna é mostrado a regra obtida, na segunda coluna o registro, referente a Tabela 1, a terceira coluna a classe do registro e na quarta coluna o resultado.

Tabela 2 – Resultados da classificação dos registros da amostra do *Iris dataset*

Regra	Registro	Classe	Resultado
SE <i>petal length</i> > 5 E <i>petal width</i> > 2 ENTÃO <i>Iris-virginica</i>	1	<i>Iris-setosa</i>	VN
	2	<i>Iris-setosa</i>	VN
	3	<i>Iris-versicolor</i>	VN
	4	<i>Iris-versicolor</i>	VN
	5	<i>Iris-virginica</i>	VP
	6	<i>Iris-virginica</i>	FP

Para verificar os resultados de uma regra, cada condição do seu antecedente é observado de acordo com o registro. É dito que uma regra satisfaz um registro quando todas as suas condições são verdadeiras. O primeiro registro não satisfaz a regra e a classe

em questão é *Iris-setosa*, portanto o resultado é verdadeiro negativo. O segundo registro também não satisfaz a regra e a classe em questão também é *Iris-setosa*, portanto o resultado também é verdadeiro negativo. O terceiro registro não satisfaz a regra e a classe em questão é *Iris-versicolor*, portanto o resultado é verdadeiro negativo. O quarto registro não satisfaz a regra e a classe em questão também é *Iris-versicolor*, portanto o resultado é verdadeiro negativo. O quinto registro satisfaz a regra e a classe em questão é *Iris-virginica*, portanto o resultado é verdadeiro positivo. E por fim, o sexto registro não satisfaz mas a classe em questão é *Iris-virginica*, portanto o resultado é falso positivo.

Com esses resultados já é possível calcular algumas medidas. A acurácia da regra, por exemplo, ficaria:

$$acc = \frac{1 + 4}{1 + 0 + 1 + 4} = 0.83 \quad (8)$$

Caso as medidas sensibilidade e especificidade também fossem necessárias, bastaria realizar o cálculo de ambas utilizando as Equações 6 e 7. O resultado seria:

$$se = \frac{1}{1 + 0} = 1 \quad (9)$$

$$sp = \frac{4}{4 + 1} = 0.8 \quad (10)$$

## 2.2.2 Seleção de atributos

Diversos fatores determinam a qualidade da performance de um classificador e um deles são os dados. Especialmente os conjuntos de dados biológicos podem conter poucos registros e muitos atributos (XING et al., 2001), esses conjuntos são conhecidos também como *datasets* com alta dimensionalidade. É intuitivo pensar que quanto maior a quantidade de atributos melhor será o modelo gerado, porém é muito comum que alguns conjuntos de dados contenham informações irrelevantes ou redundantes e isso pode comprometer o algoritmo na etapa de treino (HALL, 1999). Além disso, é necessário lembrar que quanto maior a quantidade de atributos, maior o número de dados para o algoritmo processar. Portanto, a qualidade dos dados pode determinar a eficiência do modelo gerado. A seleção de atributos busca, através do uso de algumas medidas de avaliação, selecionar aqueles que possuem mais chances de contribuir com o processo de classificação.

Existem 3 tipos de métodos para seleção de atributos: *filter*, *wrapper* e *embedded*. O método *filter* utiliza uma amostra dos dados e alguma medida para avaliar os atributos e o processo de classificação ocorre de maneira independente do classificador. Já no método *wrapper*, o próprio classificador é utilizado para avaliar quais são os atributos que aumentam a performance do mesmo (XING et al., 2001). Por fim, no método *embedded*, a seleção do atributo faz parte do algoritmo. Um exemplo deste modelo são as árvores de decisão, que selecionam os atributos durante o processo de construção da sua estrutura.

A entropia é uma grandeza que mede o grau de impureza ou desorganização de um conjunto (NOVAKOVIC, 2009), utilizada na teoria da informação e que foi adotada pelo AM. A entropia  $H$  de um conjunto  $Y$  pode ser definida de acordo com a Equação 11 (HALL, 1999):

$$H(Y) = - \sum_{i=1}^c p_i \log_2 p_i, \quad (11)$$

onde  $p_i$  é a proporção de registros pertencentes a classe  $i$  e  $c$  a quantidade de classes. Por exemplo, considere um conjunto  $Y$  com oito amostras de apenas um tipo (ou uma classe), sua entropia será:

$$H(Y) = -\frac{8}{8} \log_2 \left( \frac{8}{8} \right) = 0 \quad (12)$$

Como existem apenas amostras de uma classe, sua proporção é 8/8, seu valor de entropia é 0 (valor mínimo). Considerando outro caso, onde um conjunto possui oito amostras, sendo quatro de cada tipo (ou classe), sua entropia será:

$$H(Y) = -\frac{4}{8} \log_2 \left( \frac{4}{8} \right) - \frac{4}{8} \log_2 \left( \frac{4}{8} \right) = 1 \quad (13)$$

Visto que cada classe tem proporção igual a 4/8 sua entropia será 1 (valor máximo). As Figuras 11(a) e 11(b) ilustram esses dois casos. É possível, através da entropia, determinar a probabilidade de uma amostra de uma determinada classe ser selecionada. Considerando a Figura 11 (a), a probabilidade de uma determinada classe ser selecionada é de 50%. Já para a Figura 11 (b) poderá ser 0% ou 100%, dependendo da classe desejada.

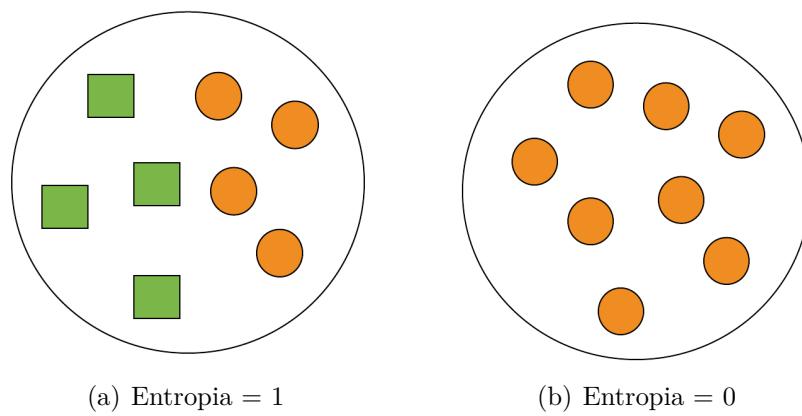


Figura 11 – Exemplo de entropia dos conjuntos

Portanto, através da entropia é possível medir o grau de impureza de um conjunto, e o mesmo procedimento pode ser feito para medir a entropia de um atributo. Conhecendo a entropia de um atributo, o processo de seleção pode ser conduzido com mais eficiência.

Uma das formas de avaliar um atributo é medir o quanto ele consegue discriminar os elementos, ou seja, o quanto um atributo consegue dividir o conjunto. Isso pode ser

feito através do cálculo do IG. Após o cálculo da entropia de todos os atributos de um conjunto de dados, o IG de um atributo pode ser obtido através da Equação 14:

$$IG = H(S) - H(X), \quad (14)$$

onde  $H(S)$  é a entropia de um determinado conjunto  $S$  e  $H(X)$  a entropia de um atributo  $X$ . Portanto, o IG mede a redução da entropia de um conjunto quando um determinado atributo é selecionado. Quanto maior a redução da entropia do conjunto, maior o ganho de informação do atributo. Além da seleção de atributos o IG também é utilizado para construção de árvores de decisão. Para exemplificar, considere 3 atributos,  $A$ ,  $B$  e  $C$  e um conjunto  $S$  com entropia igual a 0,974. A Tabela 3 apresenta os respectivos valores de entropia e IG para cada atributo através da aplicação da Equação 14:

Tabela 3 – Ganhos de Informação de cada atributo

Atributo	Entropia	IG
A	0,287	0,687
B	0,409	0,565
C	0,846	0,128

Como pode ser observado, o atributo  $A$  tem o maior valor de IG, já que o mesmo possui a menor entropia entre os atributos e consegue gerar maior redução na entropia do conjunto  $S$ .

# CAPÍTULO 3

## Trabalhos Correlatos

Nesse capítulo serão apresentados os trabalhos que foram utilizados como base teórica para o desenvolvimento da proposta dessa dissertação. A seção 3.1 apresenta uma breve introdução sobre AGs para tarefa de classificação. A Seção 3.2 detalha sobre o funcionamento do método CEE, a Seção 3.3 sobre o método NLCEE, a Seção 3.4 sobre o método HEE e a Seção 3.5 sobre o operador *Transgenic*.

### 3.1 Introdução

Nesse trabalho, propõe-se o uso de AGs para mineração de regras de classificação. Para este tipo de tarefa existem duas abordagens, conhecidas como abordagem *michigan* e *pitsburgh*. Na abordagem *michigan* a solução gerada é a própria população, enquanto que na abordagem *pitsburgh* cada indivíduo representa uma solução, ou uma regra (AMARAL, 2007). Em (FIDELIS; LOPES; FREITAS, 2000) foi proposto um AG utilizando a abordagem *pitsburgh* com o objetivo de obter regras com boas taxas de classificação e de fácil interpretação. Foram utilizados dois datasets, *dermatology* e *breast cancer*, disponíveis no repositório UCI. Nos anos seguintes, foram desenvolvidos trabalhos baseados na mesma abordagem que (FIDELIS; LOPES; FREITAS, 2000) e esses trabalhos serão detalhados nas próximas seções desse capítulo.

### 3.2 CEE

No trabalho de (AMARAL; HRUSCHKA, 2010) foi desenvolvido um AG para minerar regras de classificação de um *dataset* gerado através do consórcio *Gene Ontology* (GO) (ASHBURNER et al., 2000). O dataset em questão possui 34 atributos, 13.091 registros e 3 classes. O algoritmo foi chamado de CEE e trouxe uma representação cromossômica semelhante a de (FIDELIS; LOPES; FREITAS, 2000), como pode ser visto na Figura 12. Como se trata de uma abordagem *pitsburgh*, cada cromossomo representa uma regra e cada gene uma condição no antecedente da regra.

Gene <sub>1</sub>			...	...	Gene <sub>34</sub>		
W <sub>1</sub>	O <sub>1</sub>	V <sub>1</sub>	...	...	W <sub>34</sub>	O <sub>34</sub>	V <sub>34</sub>

Figura 12 – Cromossomo CEE (AMARAL; HRUSCHKA, 2010)

Neste cromossomo, cada gene representa um atributo, isso é, o número de genes é igual ao número de atributos no conjunto de dados. Como o *dataset* gerado possui 34 atributos, cada cromossomo do CEE possui 34 genes. Cada gene é subdividido em 3 segmentos, sendo estes o peso ( $W_i$ ), o operador ( $O_i$ ) e o valor ( $V_i$ ). O segmento  $W_i$  tem um papel importante nessa representação, ele é utilizado como um limiar que define se o atributo estará ou não presente na regra. No exemplo do CEE, esse segmento pode ter valores entre 0 e 10 e o valor escolhido como limiar foi 8, ou seja, o atributo estará presente na regra apenas se seu valor  $W_i$  for maior ou igual a 8. O segmento  $O_i$  é referente ao operador relacional utilizado, podendo ser  $<$  ou  $\geq$ . E por fim, o segmento  $V_i$  armazena um valor referente ao domínio do atributo. Como no trabalho de Fidelis (2000) foi utilizado *datasets* com atributos categóricos, o mesmo tinha a possibilidade de usar os operadores relacionais  $=$  e  $\neq$ , sendo essa a principal diferença entre a sua proposta e o CEE. Um exemplo de cromossomo e sua respectiva regra para uma classe do GO *dataset* pode ser visto na Figura 13. Nesse exemplo é apresentado um cromossomo com o gene referente ao atributo *intQtyAdjlsFather* e a classe *biological\_process*. Como o valor de peso ( $W_i$ ) é maior que o limiar (8), o atributo está presente na regra.

Gene <sub>1</sub>			...
9	$\geq$	128	...

→ SE (intQtyTermsAdjlsFather  $\geq$  128) E...  
ENTÃO *biological\_process*

Figura 13 – Exemplo de cromossomo do CEE. Adaptado de (AMARAL; HRUSCHKA, 2011)

Na etapa de mineração das regras de classificação o AG executa uma vez para cada classe. Ou seja, considerando que um *dataset* possui  $n$  classes ( $C_1, C_2, C_3, \dots, C_n$ ), o AG irá executar  $n$  vezes, sendo que em cada execução o AG pode classificar um registro como pertencente a classe  $C_i$  ou não pertencente a classe  $C_i$ . No exemplo do GO, que possui 3 classes, o AG irá executar 3 vezes buscando por regras que classifiquem os registros como sendo da classe em questão. Como método de validação foi utilizado *k-fold* com  $k = 10$  e a função *fitness* é a média das medidas de sensibilidade (*se*) e de especificidade (*sp*), como pode ser visto na Equação 15.

$$Fitness = (se + sp)/2 \quad (15)$$

Como pode ser observado, o AG irá executar para maximizar tanto *se* como *sp*. As duas medidas usadas para o cálculo do *fitness* foram apresentadas nas Equações 6 e 7.

A Tabela 4 apresenta os resultados de treino e teste do CEE utilizando *k-fold* com  $k = 10$  (*10-fold*) para cada uma das 3 classes. A primeira coluna indica qual composição do *10-fold*. Na primeira coluna, o valor 0 indica o *fold* de teste enquanto o valor 1 indica o *fold* de treino.

Tabela 4 – Resultados obtidos pelo CEE usando *10-fold*. Adaptado de (AMARAL; HRUSCHKA, 2010)

Composição	Biol.proc		Cell.comp.		Mol.Func	
	treino	teste	treino	teste	treino	teste
1111111110	0,692	0,701	0,649	0,643	0,767	0,764
1111111101	0,691	0,705	0,646	0,663	0,772	0,776
1111111011	0,694	0,680	0,657	0,647	0,756	0,755
1111110111	0,694	0,680	0,636	0,620	0,750	0,756
1111101111	0,693	0,696	0,628	0,628	0,763	0,768
1111011111	0,630	0,633	0,653	0,647	0,718	0,724
1101111111	0,694	0,679	0,652	0,640	0,776	0,757
1101111111	0,692	0,698	0,619	0,634	0,749	0,756
1011111111	0,690	0,718	0,651	0,671	0,762	0,775
0111111111	0,693	0,691	0,651	0,641	0,773	0,760
Média	0,686	0,688	0,644	0,643	0,758	0,759

Para exemplificar o processo de geração de regras do CEE e seus derivados, a Tabela 5 apresenta um exemplo de regra obtida para cada classe. A primeira coluna contém o nome de cada classe, a segunda a regra gerada, e a terceira e quarta o valor de *fitness* obtido pela regra nas etapas de treino e teste, respectivamente.

Tabela 5 – Exemplo de regras geradas pelo CEE. Adaptado de (AMARAL; HRUSCHKA, 2010).

Classe	Regra	Treino	Teste
<i>Biological Process</i>	IF (intQtyTermsAdjIsChildren < 1) AND (intNbrSpecies $\geq$ 337) THEN Biological Process	0,692	0,701
<i>Cellular Component</i>	IF (intQtyTermsAdjIsChildren $\geq$ 5) AND (intDistFatherMax < 9) AND (dblDistAVGFather < 2,1) AND (dblPercAsparticAcid < 0) THEN Cellular Component	0,649	0,643
<i>Molecular Function</i>	IF (intQtyTermsAdjIsChildren $\geq$ 2) AND (dblDistAVGChildren $\geq$ 0,5) AND (dblDistSTDChildren $\geq$ 0,1) AND (dblDistAVGFather $\geq$ 4,5) THEN Molecular Function	0,767	0,764

### 3.3 NLCEE

O CEE foi capaz de produzir boas taxas de classificação, se comparado aos classificadores clássicos, porém sua estrutura cromossômica não permite a representação de regras não lineares. Para contornar este problema foi proposto o NLCEE (AMARAL; HRUSCHKA, 2011). Para que fosse possível a representação de regras não-lineares cada gene do NLCEE passa a ter duas ocorrências, direita e esquerda, isso é, o  $i$ ésimo gene é subdividido em 6 segmentos, sendo, peso da esquerda ( $WL_i$ ), peso da direita ( $WR_i$ ), operador da esquerda ( $OL_i$ ), operador da direita ( $OR_i$ ), valor da esquerda ( $VL_i$ ) e valor da direita ( $VR_i$ ). A Figura 14 ilustra um exemplo de cromossomo do NLCEE.

$Gene_n$					
$WL_n$	$OL_n$	$VL_n$	$WR_n$	$OR_n$	$VR_n$

Figura 14 – Cromossomo NLCEE (AMARAL; HRUSCHKA, 2011)

Assim como no trabalho de (AMARAL; HRUSCHKA, 2010), também foi utilizado como método de validação  $k$ -fold com  $k = 10$  e a função *fitness* apresentada na Equação 15 assim como o GO *dataset*, citado na Seção 3.2. A Figura 15 apresenta um exemplo de regra gerada pelo NLCEE. É possível observar que, por permitir duas ocorrências de um mesmo gene, a regra utiliza o operador de disjunção OU, já que o atributo não pode ter valores maiores ou iguais a 128 e menores que 35. Este tipo de regras não eram possíveis de serem geradas no CEE.

Gene <sub>1</sub>							...
10	$\geq$	128	10	$<$	35	...	SE (intQtyTermsAdjlsFather $\geq$ 128 OU intQtyTermsAdjlsFather $<$ 35) E... ENTÃO biological_process

Figura 15 – Exemplo de cromossomo do NLCEE. Adaptado de (AMARAL; HRUSCHKA, 2011)

A Tabela 6 apresenta os resultados do NLCEE para cada uma das 3 classes usando 10-fold em cada uma das 10 composições.

### 3.4 HEE

Dando continuidade com os trabalhos utilizando AGs para classificação, (Silva; de Souza Ribeiro; do Amaral, 2013) utilizaram uma nova abordagem, integrando métodos de seleção de atributos, para classificação de um conjunto de dados chamado NCI60. O algoritmo foi nomeado *Hybrid Evolutionary Environment* (HEE), por utilizar um método híbrido. O *dataset* utilizado contém informações sobre tipos de câncer e de expressão gênica advindas de experimentos de *microarrays*. Trata-se de um conjunto de dados

Tabela 6 – Resultados obtidos pelo NLCEE usando *10-fold*. Adaptado de (AMARAL; HRUSCHKA, 2011)

Composição	Biol. proc.		Cell. comp.		Mol. func.	
	treino	teste	treino	teste	treino	teste
1111111110	0,693	0,702	0,619	0,623	0,762	0,764
1111111101	0,679	0,672	0,500	0,527	0,500	0,661
1111111011	0,682	0,673	0,624	0,500	0,500	0,543
1111110111	0,695	0,681	0,648	0,680	0,764	0,753
1111101111	0,678	0,685	0,593	0,588	0,500	0,505
1111011111	0,695	0,681	0,574	0,561	0,763	0,769
1101111111	0,652	0,628	0,506	0,507	0,779	0,766
1101111111	0,653	0,640	0,640	0,639	0,762	0,766
1011111111	0,675	0,712	0,655	0,634	0,772	0,791
0111111111	0,651	0,637	0,633	0,609	0,717	0,726
Média	0,676	0,681	0,600	0,587	0,682	0,705

com alta dimensionalidade, contendo 1.002 atributos, 61 registros e 9 classes. Alguns trabalhos usando abordagens híbridas, com AGs e métodos de seleção de atributos, já foram desenvolvidos anteriormente, como em (JIANG et al., 2008) e (YANG et al., 2010).

No HEE foi utilizado o software *Weka* (HALL et al., 2009) e alguns métodos de seleção de atributos, sendo esses o IG, *Chi-squared* e *Gain-ratio*. Os parâmetros do AG e o método de seleção de atributos são definidos a priori. Os atributos são ranqueados de acordo com sua avaliação pelo método de seleção e os  $n$  melhores são selecionados, onde  $n$  também é definido anteriormente. Portanto, o AG executa com os  $n$  melhores atributos escolhidos pelo usuário. A Tabela 7 mostra a acurácia obtida pelo HEE utilizando o *leave-one-out cross-validation* e o compara com outros 5 classificadores, tais como: PART (FRANK; WITTEN, 1998), J48 (QUINLAN, 1993), *Random Forest* (HO, 1995), *Naive Bayes* (HART, 1973) e IBK (AHA; KIBLER; ALBERT, 1991). Entre os métodos de seleção de atributos utilizados, o *Chi-squared* obteve os melhores resultados, sendos esses apresentados na Tabela 7. A primeira coluna apresenta as classes das instâncias do *dataset*. Entre as colunas 2 e 6 são mostrados os resultados obtidos por cada método e a média dos resultados. As colunas 7 e 8 mostram a média e o desvio padrão da acurácia para cada classe, respectivamente.

### 3.5 *Transgenic*

Em (AMARAL; JR, 2014) um AG foi implementado para executar em conjuntos de dados sintéticos e foi chamado de *Genetic Algorithm for Synthetic Data* (GASD). No GASD, similar ao CEE e NLCEE, cada gene representa um atributo, no entanto, a informação contida no seus genes é um valor booleano indicando se o atributo estará ou não presente na regra. Caso o valor seja *True* o atributo estará presente, e se o valor for

Tabela 7 – Comparação dos resultados obtidos pelo HEE e outros classificadores. Adaptado de (Silva; de Souza Ribeiro; do Amaral, 2013).

Classes	Métodos							
	PART	J48	RF	NB	IBK	HEE	Média	DP
$C_1$	0,55	0,59	0,51	0,59	0,63	0,73	0,600	0,069
$C_2$	0,74	0,64	0,71	0,50	0,62	0,89	0,683	0,120
$C_3$	0,83	0,85	0,78	0,79	0,98	0,98	0,868	0,082
$C_4$	0,79	0,77	0,78	0,74	0,82	0,72	0,777	0,038
$C_5$	0,70	0,67	0,82	0,81	0,94	0,73	0,778	0,090
$C_6$	0,45	0,49	0,64	0,49	0,54	0,60	0,535	0,067
$C_7$	0,48	0,38	0,64	0,50	0,37	0,73	0,477	0,123
$C_8$	0,35	0,45	0,50	0,75	0,56	0,61	0,537	0,126
$C_9$	0,49	0,62	0,62	0,62	0,72	0,78	0,642	0,091
Média	0,60	0,61	0,64	0,64	0,69	0,75		

*False*, não estará presente na regra. A Figura 16 apresenta um exemplo de cromossomo do GASD.

Gene <sub>1</sub>	Gene <sub>2</sub>	Gene <sub>3</sub>	Gene <sub>n</sub>
True	True	False	...

Figura 16 – Cromossomo do GASD. Adaptado de (AMARAL; JR, 2014)

O ponto principal do trabalho foi a proposta de um novo operador, nomeado de *transgenic*. Este operador é inspirado nos organismos geneticamente modificados, onde as características mais relevantes são inseridas artificialmente em seu genoma, organismos esses chamados de transgênicos. O operador então seleciona as características mais importantes e as insere no cromossomo dos indivíduos com o objetivo de produzir uma convergência mais rápida para melhores resultados. Mais especificamente, se o algoritmo encontrar um valor  $V$  em um gene  $G$ , e esse valor produzir uma forte influência no *fitness* do indivíduo, o operador irá perpetuar esse valor  $V$  nos indivíduos das próximas gerações. De certa forma, é possível dizer que o operador *transgenic* atua como um método de seleção de atributos, já que avalia aqueles atributos mais relevantes, baseado em como os mesmos influenciam no *fitness* dos indivíduos, para mantê-los nas gerações seguintes.

Foram utilizados 3 *datasets* sintéticos gerados pelo ambiente *GeNIe*<sup>1</sup>, um software para criação de modelos de decisão e análises (SMILE, 1999). Os *dataset* gerados foram nomeados de  $D_1$ ,  $D_2$  e  $D_3$  e cada um possui 19 atributos, 3 classes e 12.000 instâncias. Para o  $D_1$  existe apenas um atributo que define cada classe, ou seja, apenas 3 atributos são relevantes. Para o  $D_2$ , apenas 2 atributos definem cada classe, ou seja, apenas 6

<sup>1</sup> <https://www.bayesfusion.com>

atributos são relevantes. E por fim, para o  $D_3$ , apenas 3 atributos definem cada classe, ou seja, apenas 9 atributos são relevantes. Portanto, o AG deve ser capaz de selecionar apenas aqueles atributos que sejam determinantes para o processo de classificação através do uso do operador *transgenic*. A Tabela 8 apresenta os resultados do AG com o operador *transgenic*. Para cada *dataset* é mostrado os valores médios de *fitness* em treino e teste e o desvio padrão obtidos em 34 execuções com sementes randômicas distintas. O *fitness* é calculado usando a Equação 15.

Tabela 8 – Resultados do AG utilizando o operador *transgenic*

		Classe 1		Classe 2		Classe 3	
Datasets		treino	teste	treino	teste	treino	teste
D1	média	0,720	0,721	0,718	0,718	0,729	0,730
	desvio padrão	0,001	0,015	0,001	0,023	0,001	0,011
D2	média	0,700	0,701	0,703	0,706	0,699	0,702
	desvio padrão	0,002	0,011	0,003	0,013	0,002	0,014
D3	média	0,627	0,630	0,626	0,628	0,617	0,619
	desvio padrão	0,001	0,011	0,002	0,010	0,003	0,013



# CAPÍTULO 4

## Proposta

Nesse capítulo será apresentado os AGs propostos, sendo a Seção 4.2 a versão do NLCEE com representação binária (BIN-NLCEE) e na Seção 4.3 o CEE com ganho de informação (IG-CEE).

### 4.1 Introdução

Como já foi apresentado nas seções anteriores, os AGs vêm sendo vastamente utilizados para tarefas de classificação, principalmente utilizando a abordagem *pitsburgh*. Portanto, entende-se que seja necessário aperfeiçoar constantemente essa abordagem buscando melhores resultados e até mesmo apontar possíveis fragilidades dos algoritmos. Visando isso, esse trabalho buscou analisar detalhes do funcionamento dos algoritmos CEE e NLCEE e investigar os efeitos de novas mudanças nos mesmos. Para isso, dois pontos principais foram considerados. Em primeiro lugar, em todos os trabalhos relacionados e citados no Capítulo 3, quando utilizados datasets com atributos contínuos, a representação cromossômica utiliza valores inteiros ou reais. Entende-se então que seja necessário investigar diferentes formas de representação cromossômica e comparar com aquelas já utilizadas. Com essa investigação é possível apontar os principais pontos positivos e negativos de cada uma, além de contribuir com uma discussão mais aprofundada acerca das possibilidades de representações cromossômicas.

O segundo ponto analisado foi o segmento peso ( $W$ ) dos AGs. Esse segmento, como já mencionado, possui um papel importante para a geração das regras, já que sua função é decidir se um atributo estará ou não presente na mesma. Nos trabalhos relacionados utiliza-se um valor inteiro como limiar, e se caso o valor seja maior que o limiar, o atributo estará presente na regra. Sendo assim, ele se torna mais um parâmetro que deve ser definido antes da execução do AG e para encontrar o melhor valor para o limiar é necessário testar cada um em diferentes execuções.

Levando em consideração esses dois problemas apontados, foram implementados duas novas versões. A primeira buscou utilizar representação binária nos cromossomos, com o

objetivo de alcançar melhores taxas de convergência. A segunda versão propõe o uso de uma medida de avaliação para seleção de atributos, o IG, para definir o valor do limiar do peso, fazendo com que cada gene tenha um valor diferente. Com o objetivo de ilustrar o desenvolvimento da abordagem utilizada nesse trabalho, a Figura 17 apresenta uma árvore cronológica dos trabalhos realizados até os dias atuais.

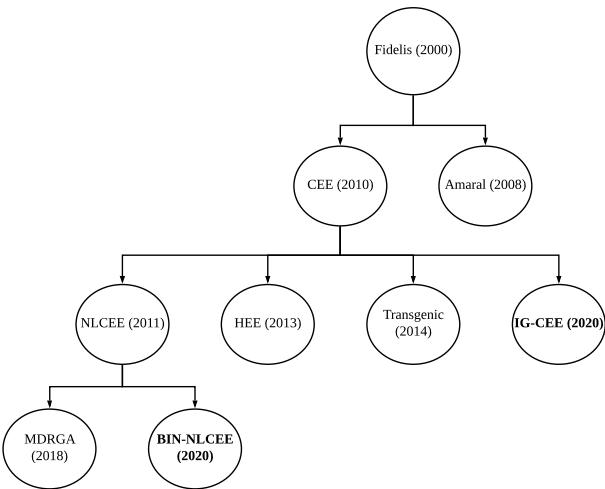


Figura 17 – Diagrama dos trabalhos desenvolvidos

Como pode ser visto, o marco inicial desta abordagem ocorre no trabalho de (FIDELIS; LOPES; FREITAS, 2000). Posteriormente, em Amaral (2008), um método baseado em Fidelis (2000) foi utilizado para classificação de um *dataset* com informações referentes a vários tipos de câncer. Em seguida foram propostos o CEE e suas extensões NLCEE, que modificou o cromossomo para gerar regras não lineares, o HEE que utilizou métodos de seleção de atributos e o *transgenic* com uma abordagem para selecionar as características (atributos) mais relevantes. Por fim o MDRGA (MATOS; AMARAL, 2018), que propôs uma nova estrutura cromossômica mais complexa capaz de gerar regras não lineares. Portanto, o trabalho em questão procede o MDRGA, mas apontando diferentes problemas e propondo dois novos métodos, dessa forma dando continuidade com o desenvolvimento dessas abordagens.

## 4.2 BIN-NLCEE

Como já mencionado, o NLCEE utiliza representação por números inteiros nos seus cromossomos. Na primeira versão implementada deste trabalho, buscou-se utilizar representação binária e verificar quais seriam seus efeitos nos valores de *fitness* do indivíduo. Portanto, a estrutura cromossômica foi alterada para suportar tal representação. O segmento operador (*O*), que indica qual operador relacional será utilizado, pode ter dois valores, 0 ou 1. Caso o valor seja 0, o operador será  $\geq$ , e caso o valor seja 1, o operador será  $<$ .

Para a binarização dos segmentos peso ( $W$ ) e valor ( $V$ ) foi necessário adotar duas estratégias diferentes, para cada um. Nos trabalhos anteriores, o segmento peso poderia ter valores entre 1 e 10 ( $W_i = 1\dots10$ ). Quando um número positivo inteiro é convertido para binário, algumas limitações podem ocorrer, por exemplo, para representar até o número 7, apenas 3 bits são necessários:

$$111 = 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 7$$

Já para os números 8, 9 e 10, são necessários 4 bits, como pode ser visto no exemplo abaixo:

$$\begin{aligned} 1000 &= 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 = 8 \\ 1001 &= 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 9 \\ 1010 &= 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 10 \end{aligned}$$

O problema encontra-se em uma possível mutação do valor com 4 bits, como por exemplo, se o segmento peso de um cromossomo possui como valor binário 1010, e sofrer uma mutação o transformando em 1110, o valor passa a ser 14, valor esse fora do intervalo previamente definido.

$$1110 = 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 14$$

Para contornar esse problema, foram utilizados valores no intervalo entre 0 e 1 para o segmento peso ( $W_i = 0\dots1$ ). Portanto, a transformação do valor binário para o decimal passa a ser feito pela equação:

$$d = \sum_{i=0}^n b_i 2^{-i}, \quad (16)$$

onde  $b_i$  é o valor do bit, podendo ser 0 ou 1 e  $n$  o número de bits. Dessa forma é possível definir um número de bits para representar o valor sem que a mutação transforme-o em um valor fora do intervalo. Por exemplo, considerando um valor com 6 bits, é possível obter:

$$111100 = 1 \times 2^{-5} + 1 \times 2^{-4} + 1 \times 2^{-3} + 1 \times 2^{-2} + 0 \times 2^{-1} + 0 \times 2^{-0} = 0.9375$$

Com essa estratégia, quanto maior o número de bits utilizado, maior a quantidade de números é possível representar, sem que o intervalo seja violado. Ou seja, a quantidade de números que passa a ser possível representar é igual a  $2^n$ , onde  $n$  é a quantidade de bits utilizados. Dessa maneira o segmento peso pode conter valores entre 0 a  $\approx 1$  ( $W_i = 0\dots \approx 1$ ), dependendo da quantidade de bits utilizados. Portanto, passa a ser necessário especificar o número de bits como parâmetro do método.

Para o segmento valor, a mesma estratégia foi utilizada, com o objetivo de evitar os mesmos problemas. No entanto, os valores dos atributos são diferentes de um *dataset*

para outro e possuem valores diferentes do intervalo [0...1]. Para trazer os valores para o intervalo desejado, foi aplicado a normalização *min-max*, definida pela Equação 17.

$$x' = \frac{x - \min(A)}{\max(A) - \min(A)}, \quad (17)$$

onde  $x$  é o valor original do atributo  $A$ . Com os valores normalizados, é possível aplicar a Equação 16 para transformar o binário em um valor válido. A Figura 18 apresenta um exemplo de conversão de um cromossomo binário.

<i>Gene<sub>n</sub></i>		
<i>WL<sub>n</sub></i>	<i>OL<sub>n</sub></i>	<i>VL<sub>n</sub></i>
111100	0	110111
		
0,9375	$\geq$	0,8594

Figura 18 – Conversão do gene binário

No BIN-NLCEE os indivíduos são inicializados com valores binários, e todas as operações são feitas com esses valores, exceto o cálculo do *fitness*. No cálculo do *fitness* os valores convertidos devem ser verificados, para determinar se uma regra satisfaz ou não um registro, e portanto os valores dos atributos também são transformados em uma escala [0...1] usando a Equação 17. Com essa abordagem, o operador de mutação atua de maneira simplificada, apenas trocando o valor do bit. A troca do bit pode levar a soluções mais próximas da solução ótima, por gerar valores completamente aleatórios. Portanto, além da simplicidade, a nova estratégia pode explorar o espaço de busca de maneira mais eficiente, se comparado aos métodos correlatos.

### 4.3 IG-CEE

Considerando o que foi apontado na Seção 4.1, entende-se que o segmento peso deve ser investigado com profundidade e novas abordagens devem ser propostas. Baseando-se no trabalho de (AMARAL; JR, 2014), apresentado na Seção 3.5, onde um novo operador para seleção das características mais relevantes foi desenvolvido, o trabalho em questão buscou investigar uma nova maneira de trabalhar o segmento peso para seleção dos melhores atributos para fazerem parte da regra.

Nos AGs CEE e NLCEE, um valor é escolhido antes da execução e ele será utilizado como limiar do segmento peso. Para que um atributo esteja presente na regra, ou seja, o gene seja ativado, o valor deve ser maior ou igual ao limiar. Sendo assim, o AG passa a ter mais um parâmetro e para encontrar o melhor limiar é necessário testar diferentes

valores. Para aprimorar a capacidade do AG de selecionar os atributos mais relevantes, o IG de cada um é calculado e utilizado para definir o valor do limiar. Os atributos são ordenados de acordo com seu valor de IG, de maneira que o atributo com maior valor ocupe a última posição. Considerando 4 atributos,  $A_1$ ,  $A_2$ ,  $A_3$  e  $A_4$ , e seus respectivos valores de IG, eles serão ordenados de acordo com a Figura 19.

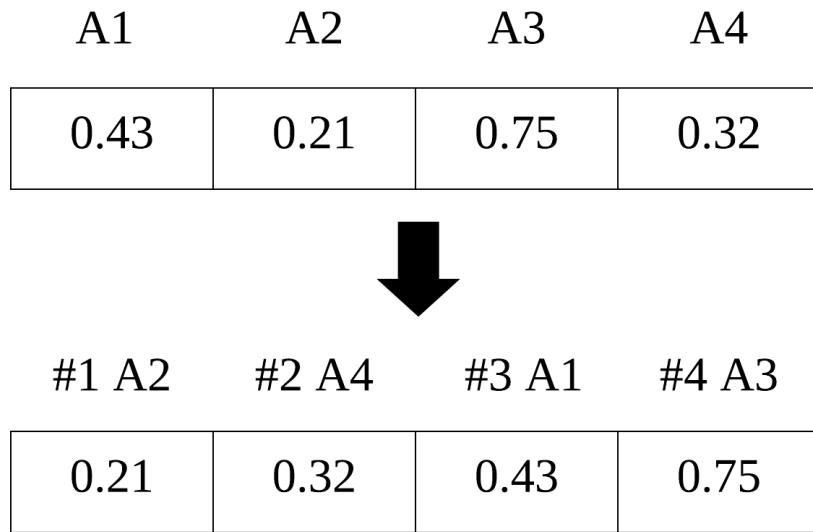


Figura 19 – Ordenação dos atributos de acordo com o ganho de informação

Com os atributos ordenados, calcula-se a probabilidade de cada um ser selecionado para estar presente na regra, usando a equação abaixo.

$$S = \frac{p}{n}, \quad (18)$$

onde  $S$  é a probabilidade do atributo ser ativo,  $p$  a sua posição e  $n$  a quantidade de atributos no *dataset*. O valor de  $S$  passa a ser o valor de limiar e, diferente das outras abordagens, o valor de peso agora precisa ser menor ou igual ao valor de  $S$  para que o atributo apareça na regra. Usando os atributos da Figura 19 como exemplo, a probabilidade de  $A_1$  aparecer na regra será de 75%,  $A_2$  de 25%,  $A_3$  de 100% e  $A_4$  50%. Como pode ser observado, o atributo com maior valor de IG sempre estará presente na regra ( $S = 1$  ou 100%), e quanto menor o IG de um atributo, menor suas chances de estar presente.



# CAPÍTULO 5

## Experimentos e Análise dos Resultados

O capítulo em questão irá discorrer sobre os experimentos realizados e suas configurações. Também serão apresentados os resultados obtidos pelos métodos propostos, os métodos relacionados e os classificadores tradicionais escolhidos. Por fim, uma análise dos resultados de cada método proposto é feita.

### 5.1 Método para a Avaliação

Nesse trabalho optou-se por utilizar como função *fitness* a média entre as medidas sensibilidade (se) e especificidade (sp), já apresentada na Equação 15. O motivo para isso foi o fato de todos os trabalhos relacionados, apresentados no Capítulo 3, utilizarem a mesma função, comprovando assim sua eficiência. Para validação dos resultados foi utilizado o método *k-fold* com  $k = 10$ . Cada *dataset* foi dividido em 10 partições proporcionalmente iguais e, a cada *fold* de treino o AG executa uma vez. Em seguida a melhor regra encontrada na etapa de treino será a regra testada. Por fim, a média de *fitness* das 10 etapas de teste é obtida.

Como o AG exige a definição de alguns parâmetros antes de sua execução, nesse trabalho foram montadas 6 configurações diferentes, como apresentado na Tabela 9.

Tabela 9 – Configurações de parâmetros utilizadas pelos AGs

Configuração	Pop	Gen	Mut	Tour
1	50	100	0,05	3
2	50	100	0,04	3
3	50	200	0,05	3
4	50	200	0,04	3
5	100	100	0,05	3
6	100	100	0,04	3

Além dos parâmetros apresentados acima, ainda existe a escolha dos operadores genéticos. Como método de seleção foi utilizado o *tournament* com *tour* = 3, além do

*Crossover* de dois pontos (com taxa de 100%) e método de reinserção melhores pais e filhos. No método BIN-NLCEE, foi utilizado 6 bits para representar os segmentos peso ( $W$ ) e valor ( $V$ ). E por fim, para validar e comparar os métodos, foi utilizado intervalo de confiança com  $\alpha = 0.05$ , ou seja, 95% de confiança. As duas propostas foram comparadas com os algoritmos CEE, NLCEE e outros 4 classificadores bastante conhecidos: J48, IBK, SVM e *Naive Bayes* (NB). Para executar esse classificadores foi utilizado a biblioteca de *Machine Learning* e *Data Mining* Weka

Foram escolhidos 4 *datasets* bem conhecidos com informações de domínio médico e que podem ser encontrados no repositório UCI <sup>1</sup>. Esses 4 *datasets* foram utilizados para validar os resultados do BIN-NLCEE. O primeiro a ser descrito é o *pima-indians-diabetes*, chamado aqui de *diabetes*. Trata-se de um conjunto de dados com informações sobre pacientes do sexo feminino de até 21 anos de idade e o seu objetivo é predizer se o paciente possui ou não diabetes. O conjunto de dados contém 768 registros, 9 atributos e 2 classes, sendo essas *teste\_negative* para aqueles com resultado negativo para diabetes e *teste\_positive* para aqueles com resultado positivo para diabetes. A Tabela 10 mostra a distribuição de registros por classe.

Tabela 10 – Distribuição de registros por classe do *diabetes dataset*.

Classe	# de registros	% de registros
<i>tested_negative</i>	500	65,11%
<i>tested_positive</i>	268	34,89%
Total	768	100%

O segundo *dataset* descrito aqui é o *breast cancer wisconsin*, chamado aqui de *breast-w*. O *dataset* foi criado pelos departamentos de Ciência da Computação e Ciências Clínicas da Universidade de Wisconsin e contém informações referentes a imagens de células de câncer de mama. Possui 560 registros, 32 atributos e duas classes, sendo estas *M (malignant)* e *B (benign)*. O objetivo então é predizer se a célula é maligna ou benigna. A Tabela 11 apresenta a distribuição de registros por classe do *breast-w dataset*.

Tabela 11 – Distribuição de registros por classe do *breast-w*

Classe	# de registros	% de registros
<i>malignant</i>	212	37,25%
<i>benign</i>	357	62,75%
Total	569	100%

O *parkinsons dataset* (LITTLE et al., 2007) contém informações referentes a medições de voz para 31 pessoas, sendo 23 com doença de Parkinson. Contém 195 registros, 22

<sup>1</sup> <https://archive.ics.uci.edu/ml/index.php>

atributos e duas classes, sendo estas *0* para *healthy* (saudável) e *1* para *PD* (doença de Parkinson). Portanto, o objetivo é classificar um indivíduo como saudável ou com doença de Parkinson. A Tabela 12 apresenta a distribuição de registros por classe do *parkinsons dataset*.

Tabela 12 – Distribuição de registros por classe do *parkinson dataset*.

Classe	# de registros	% de registros
<i>healthy</i>	48	24,61%
<i>PD</i>	147	75,38%
Total	195	100%

E por fim, o *mammography mass dataset*, chamado aqui apenas de *mammography*. Ele contém informações referentes a lesões de massa mamográfica e tem por objetivo dizer a severidade da lesão. Contém 961 registros, 6 atributos e 2 classes que indicam a severidade, chamadas *benign* e *malignant*. A Tabela 13 apresenta a distribuição de registros por classe do *mammography dataset*.

Tabela 13 – Distribuição de registros por classe do *mammography dataset*.

Classe	# de registros	% de registros
<i>benign</i>	516	53,69%
<i>malignant</i>	445	46,31%
Total	961	100%

Para validar o AG IG-CEE foram utilizados os 3 *datasets* mencionados na subseção 3.5. Como já foi apresentado, os 3 *datasets* foram gerados através do ambiente *GeNIe* e foram nomeados  $D_1$ ,  $D_2$  e  $D_3$ . Os mesmos foram criados para testar a capacidade do operador *transgenic* em selecionar as características mais relevantes, ou seja, os atributos mais importantes. Dessa maneira, entende-se que os mesmos possam ser utilizados para validação da proposta desse trabalho. Para o  $D_1$ , existem 3 atributos que definem cada classe, sendo eles o atributo *3* para a classe *0*, o atributo *14* para a classe *1* e o atributo *18* para a classe *2*. Para o  $D_2$ , 2 atributos definem cada classe, ou seja, são 6 atributos relevantes. Os atributos *7* e *11* para a classe *0*, os atributos *9* e *16* para a classe *1* e os atributos *2* e *17* para a classe *2*. Por fim, para o  $D_3$ , 3 atributos definem cada classe, ou seja, são 9 atributos relevantes. Os atributos *4*, *6* e *10* para classe *0*, os atributos *1*, *5* e *12* para a classe *1* e os atributos *8*, *13* e *15* para a classe *2*. A Tabela 14 mostra a distribuição dos registros por classe para os 3 *datasets*. A Tabela 15 apresenta os atributos relevantes para cada classe.

Tabela 14 – Distribuição de registros dos *datasets*  $D_1$ ,  $D_2$  e  $D_3$ .

Classes	$D_1$		$D_2$		$D_3$	
	#	%	#	%	#	%
0	4025	33,54	3942	32,85	3938	32,81
1	3859	32,151	3990	33,25	3944	32,86
2	4116	34,31	4068	33,9	4118	34,33
Total	12000	100	12000	100	12000	100

Tabela 15 – Atributos relevantes para cada *dataset*.

Classes	$D_1$		$D_2$		$D_3$		
	Atr.	Atr.	Atr.	Atr.	Atr.	Atr.	Atr.
0	3	7	11	4	6	10	
1	14	9	16	1	5	12	
2	18	2	17	8	13	15	

## 5.2 Experimentos

Nessa seção serão apresentados os resultados dos experimentos para as duas propostas aplicadas nos *datasets* apresentados anteriormente. A subseção 5.2.1 apresenta os resultados dos experimentos obtidos pelo BIN-NLCEE e a subseção 5.2.2 os resultados do IG-CEE.

### 5.2.1 Resultados BIN-NLCEE

A Tabela 16 apresenta as melhores configurações para cada método em cada conjunto de dados. A primeira coluna apresenta os métodos e as demais colunas os *datasets*. Todas as configurações analisadas podem ser visualizadas na Tabela 9.

Tabela 16 – Melhores configurações de parâmetro para cada método e para cada conjunto de dados.

	<i>breast-w</i>	<i>diabetes</i>	<i>parkinsons</i>	<i>mammography</i>
BIN-NLCEE	5	5	6	5
NLCEE	5	5	6	6
CEE	6	5	5	6

Para encontrar as melhores configurações de parâmetros para os AGs, cada método foi executado 100 vezes utilizando sementes randômicas distintas. O resultado de cada semente é a média das etapas de testes do *10-fold cross-validation*. As Tabelas 17 e 18 apresentam os resultados dos AGs para os *datasets* *breast-w*, *diabetes*, *parkinsons* e *mammography* e suas respectivas classes. Esses resultados foram obtidos utilizando as melhores configuração de parâmetros, como mostrado na Tabela 16. Como os métodos baseados em AG executam uma vez para cada classe, e cada *dataset* utilizado aqui possui duas classes, os resultados são apresentados em tabelas diferentes. Os resultados em negrito

são aqueles que obtiveram valores superiores de acordo com o intervalo de confiança. Na primeira coluna são mostrados os nomes dos *datasets* e respectivas classes, e as colunas restantes, os métodos analisados. Os resultados são apresentados no formato (*média de fitness*  $\pm$  *intervalo de confiança*).

Tabela 17 – Média de *fitness* das 100 execuções (com *seeds* distintas) para cada *dataset* e uma classe específica.

	BIN-NLCEE	NLCEE	CEE
<i>breast-w</i> ( <i>M</i> )	<b>97,89 <math>\pm</math> 0,35</b>	<b>97,18 <math>\pm</math> 0,37</b>	96,45 $\pm$ 0,42
<i>diabetes</i> ( <i>tested_negative</i> )	<b>83,34 <math>\pm</math> 0,65</b>	82,00 $\pm$ 0,65	80,92 $\pm$ 0,66
<i>parkinsons</i> ( <i>healthy</i> )	<b>99,27 <math>\pm</math> 0,33</b>	<b>98,53 <math>\pm</math> 0,42</b>	97,92 $\pm$ 0,51
<i>mammography</i> ( <i>benign</i> )	<b>85,86 <math>\pm</math> 0,65</b>	<b>85,64 <math>\pm</math> 0,64</b>	<b>85,23 <math>\pm</math> 0,63</b>

Tabela 18 – Média de *fitness* das 100 execuções (com *seeds* distintas) para cada *dataset* e uma classe específica.

	BIN-NLCEE	NLCEE	CEE
<i>breast-w</i> ( <i>B</i> )	<b>98,45 <math>\pm</math> 0,28</b>	<b>98,09 <math>\pm</math> 0,33</b>	97,46 $\pm$ 0,33
<i>diabetes</i> ( <i>tested_positive</i> )	<b>82,34 <math>\pm</math> 0,64</b>	80,93 $\pm$ 0,66	80,21 $\pm$ 0,72
<i>parkinsons</i> ( <i>PD</i> )	<b>97,34 <math>\pm</math> 0,55</b>	<b>95,78 <math>\pm</math> 1,12</b>	<b>96,13 <math>\pm</math> 0,72</b>
<i>mammography</i> ( <i>malignant</i> )	<b>84,23 <math>\pm</math> 0,59</b>	<b>84,28 <math>\pm</math> 0,64</b>	<b>83,78 <math>\pm</math> 0,62</b>

Analizando as Tabelas 17 e 18, é possível observar que para o *breast-w dataset*, o BIN-NLCEE obteve melhores resultados que o CEE, onde foi obtido média de 97,89 (variando de 97,54 a 98,24) para a classe *malignant* e 98,45 (variando de 98,17 a 98,73) para a classe *benign*. Entre o CEE e NLCEE foram obtidos médias de 96,45 (variando de 96,03 a 96,87) e 97,18 (variando de 96,81 a 9,55) para a classe *malignant* e 97,46 (variando de 9,13 a 97,79) e 98,09 (variando de 97,76 a 98,43) para a classe *benign*, respectivamente. Os métodos BIN-NLCEE e NLCEE obtiveram resultados estatisticamente iguais, considerando o intervalo de confiança.

Para o *diabetes dataset*, o BIN-NLCEE obteve melhores resultados que o NLCEE e o CEE nas duas classes. Para a classe *tested\_negative* o BIN-NLCEE obteve média de 83,34 (variando de 82,69 a 83,99) e 82,34 (variando de 81,70 a 82,98) para a classe *tested\_positive*. Para o *parkinsons dataset*, os métodos BIN-NLCEE e o NLCEE obtiveram resultados estatisticamente iguais na classe *healthy*, considerando o intervalo de confiança,

onde o BIN-NLCEE obteve média de 99,27 (variando de 98,94 a 99,60), enquanto que o NLCEE obteve média de 98,53 (variando de 98,11 a 98,95). Para a classe *PD* todos os 3 métodos obtiveram resultados iguais, de acordo com o intervalo de confiança. BIN-NLCEE obteve média de 97,34 (variando de 96,79 a 97,89) enquanto que o CEE e o NLCEE obtiveram médias de 96,13 (variando de 95,41 a 96,85) e 95,78 (variando de 94,66 a 96,90), respectivamente.

Para o *mammography dataset* todos os três métodos obtiveram resultados iguais, de acordo com o intervalo de confiança. O BIN-NLCEE obteve média de 85,86 (variando de 85,21 a 86,51) para a classe *benign* e 84,23 (83,64 a 84,82) para a classe *malignant*. Já os métodos NLCEE e CEE obtiveram médias de 85,64 (85,00 a 86,28) e 85,23 (84,60 a 85,86) para a classe *benign* e 84,28 (83,64 a 84,92) e 83,78 (83,16 a 84,40) para a classe *malignant*, respectivamente.

A Tabela 19 apresenta os resultados obtidos por 4 classificadores tradicionais: J48, IBK, Naive Bayes e SVM. Para obter os resultados foi utilizado a ferramenta *Weka* e com o objetivo de fazer uma comparação justa, a mesma função *fitness*, apresentada na Equação 15, foi utilizada para avaliar cada método. Foram utilizados os parâmetros *default* desses classificadores na ferramenta. A primeira coluna contém os nomes dos conjuntos de dados e as 4 colunas restantes os métodos analisados. Os resultados são apresentados no formato (*média de fitness*  $\pm$  *intervalo de confiança*).

Tabela 19 – Média de *fitness* das 100 execuções (com *seeds* distintas) para cada *dataset* usando classificadores tradicionais.

	J48	IBK	NB	SVM
<i>breast-w</i>	92,94 $\pm$ 0,17	95,07 $\pm$ 0,08	92,07 $\pm$ 0,06	<b>97,02 <math>\pm</math> 0,06</b>
<i>diabetes</i>	70,66 $\pm$ 0,25	66,65 $\pm$ 0,16	<b>72,11 <math>\pm</math> 0,11</b>	71,38 $\pm$ 0,10
<i>parkinsons</i>	79,70 $\pm$ 0,61	<b>95,56 <math>\pm</math> 0,22</b>	76,80 $\pm$ 0,21	74,94 $\pm$ 0,21
<i>mammography</i>	<b>81,57 <math>\pm</math> 0,11</b>	74,68 $\pm$ 0,15	77,80 $\pm$ 0,10	79,65 $\pm$ 0,09

Para o *breast-w dataset*, o método SVM obteve os melhores resultados com média de 97,02 (variando de 96,96 a 97,08). Para o *diabetes dataset*, o método NB obteve os melhores resultados com média de 72,11 (variando de 72,00 a 72,22). Para o *parkinsons dataset*, o método IBK obteve o melhor resultado com média de 95,56 (variando de 95,34 a 95,78). Por fim, para o *mammography dataset*, o J48 obteve o melhor resultado com média de 81,57 (variando de 81,46 a 81,68).

Além das tabelas acima, foram gerados 8 gráficos mostrados nas Figuras 20, 21, 22 e 23 comparando a convergência dos 3 métodos, BIN-NLCEE, NLCEE e CEE. Foi selecionado uma semente aleatória (semente 15) para execução dos AGs e montagem dos gráficos.

Para o *breast-w dataset*, mostrado na Figura 20, o BIN-NLCEE obteve melhores taxas de convergência para as duas classes. Para o *diabetes dataset*, o BIN-NLCEE obteve melhores taxas de convergência para as duas classes e também melhores valores de *fitness* na população inicial. Para o *parkinsons dataset*, os três métodos tiveram resultados seme-

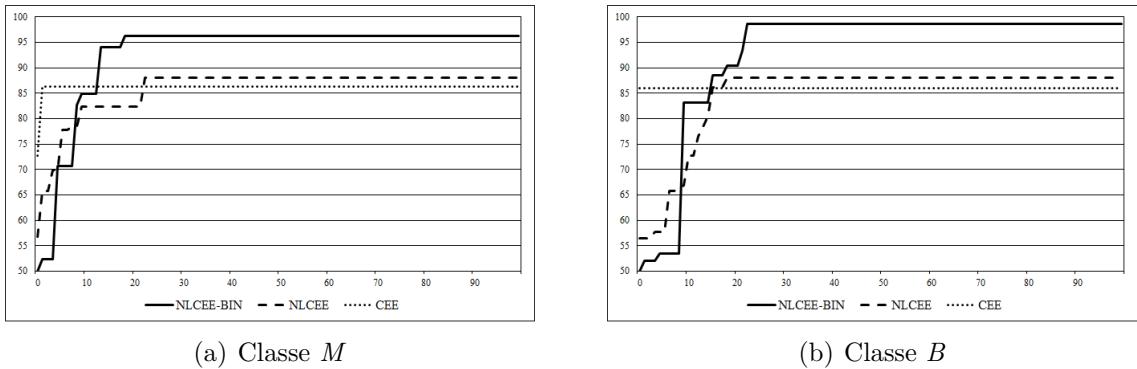


Figura 20 – Gráficos de convergência dos métodos baseados em AG para o *breast-w dataset*

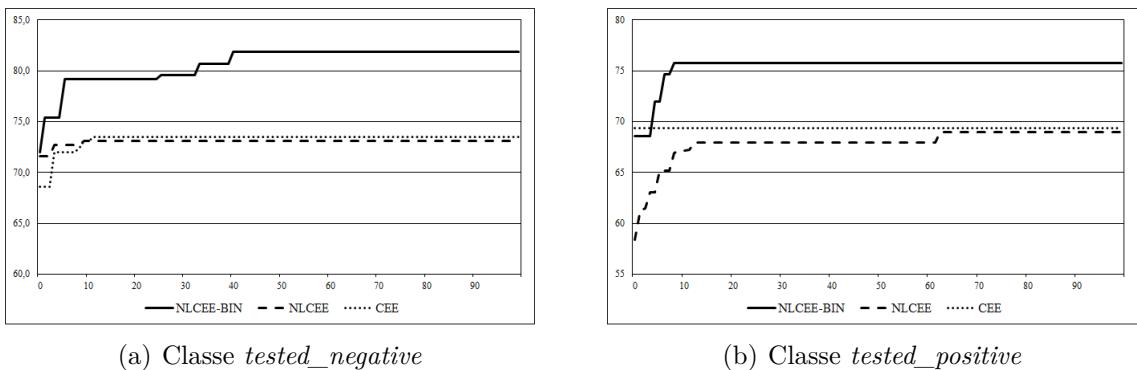


Figura 21 – Gráficos de convergência dos métodos baseados em AG para o *diabetes dataset*

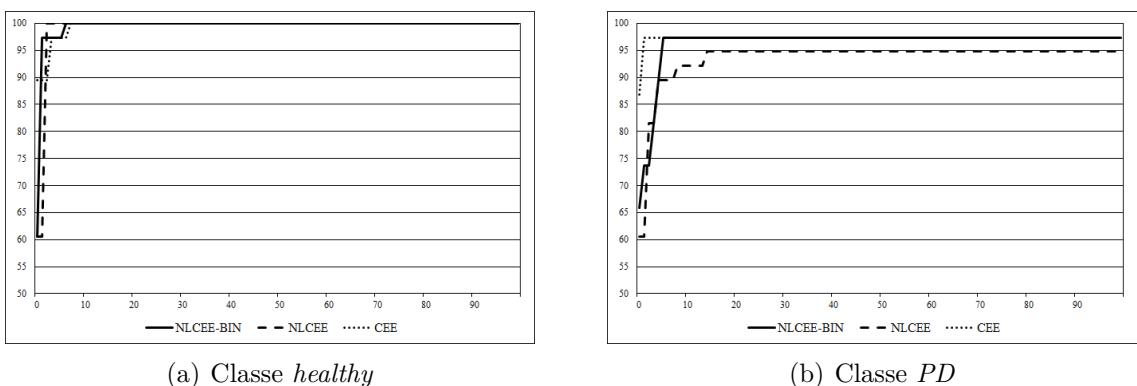


Figura 22 – Gráficos de convergência dos métodos baseados em AG para o *parkinsons dataset*

lhantes para a classe *healthy*, tanto nos valores de *fitness* quanto nas taxas de convergência. Para a classe *PD*, os métodos BIN-NLCEE e CEE conseguiram melhores resultados. No *mammography dataset*, para a classe *benign*, o BIN-NLCEE conseguiu melhor taxa de convergência e *fitness* igual ao NLCEE. E para a classe *malignant*, CEE e NLCEE obtiveram melhores valores de *fitness* e o CEE melhor taxa de convergência.

A Tabela 20 apresenta as regras geradas pelo BIN-NLCEE. Foi selecionado uma se-

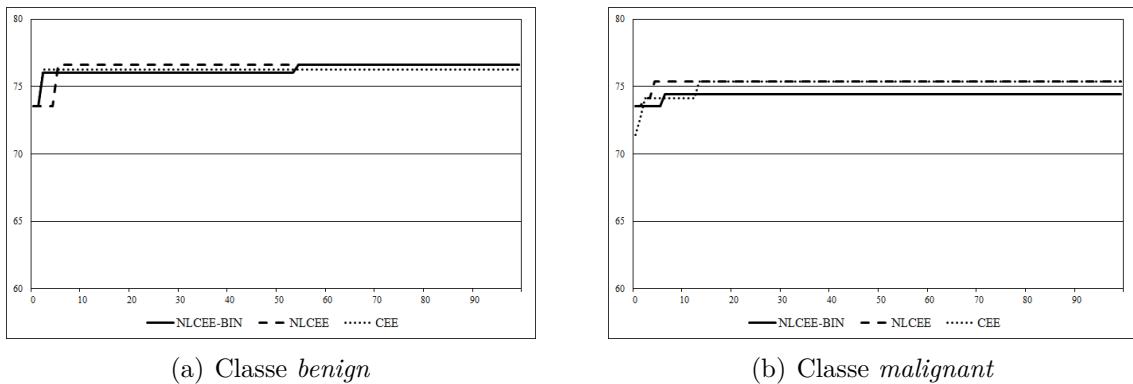


Figura 23 – Gráficos de convergência dos métodos baseados em AG para o *mammography dataset*

mente randômica (semente 15) e um *fold* aleatório para selecionar as regras. É possível observar que as regras geradas pelo BIN-NLCEE são simples, especialmente para o *mammography dataset*. Para a classe *benign* a regra obtida contém apenas 3 atributos (V2, V3 e V5) dos 6 presentes no *dataset* e para a classe *malignant*, a regra obtida contém 3 atributos (V3, V4 e V5).

### 5.2.2 Resultados IG-CEE

A Tabela 21 apresenta quais foram as melhores configurações para os métodos IG-CEE e CEE nos datasets  $D_1$ ,  $D_2$  e  $D_3$ . Optou-se por comparar os resultados apenas com o CEE, pelo fato dos *datasets* possuírem atributos nominais, podendo conter apenas dois valores (Ausente e Presente), e portanto não sendo capaz de gerar regras não lineares. Para calcular o ganho de informação de cada atributo, foi utilizado a biblioteca do *Weka*, que possui um método específico para tal.

Tabela 21 – Melhores configurações para cada método e cada conjunto de dados.

	$D_1$	$D_2$	$D_3$
IG-CEE	1	2	6
CEE	6	6	6

Para encontrar a melhor configuração, foram utilizadas as mesmas medidas e métodos citados na subseção 5.2.1. Já na Tabela 22 é apresentado os resultados obtidos pelos métodos IG-CEE e CEE. Na primeira coluna é apresentado o nome dos métodos e nas colunas restantes, a média de *fitness*, para cada *dataset*, das cem execuções utilizando sementes distintas. Foi utilizado o mesmo formato da Tabela 17 para apresentar os resultados.

Para o *dataset*  $D_1$ , o método IG-CEE obteve melhores resultados apenas para a classe 2, com média de 72,71 (variando de 72,46 a 72,96), enquanto o método CEE obteve

Tabela 20 – Conjunto de regras geradas pelo BIN-NLCEE

Dataset	Regra
<i>breast-w</i>	IF((A1<0.0625 OR A1≥0.4219)AND(A2≥0.2657) AND(A3≥0.1719)AND(A7≥0.0625) AND(A8≥0.1563)AND(A18<0.5938) AND(A20<0.6875)AND(A21<0.7188) AND(A30<0.7813)) THEN <i>M</i>
	IF((A15<0.8438)AND(A16<0.3125) AND(A17<0.9375)AND(A20≥0.1875 OR A20<0.1719) AND(A21<0.1875)AND(A23<0.5938) AND(A26<0.5157)AND(A29<0.4688) AND(A30<0.75)AND(A31<0.7188 OR A31≥0.8282)) THEN <i>B</i>
<i>diabetes</i>	IF((V1<0.25 OR V1≥0.5)AND(V2<0.8125) AND(V3≥0.0313)AND(V4<0.4219) AND(V8<0.25)) THEN <i>tested_negative</i>
	IF((V1≥0.0625)AND(V5<0.2969 OR V5≥0.625) AND(V6<0.625)AND(V8≥0.1094)) THEN <i>tested_positive</i>
<i>parkinsons</i>	IF((V1<0.5625)AND(V2≥0.375) AND(V3<0.375 OR V3≥0.875)AND(V16≥0.3125) AND(V20<0.4063)) THEN <i>healthy</i>
	IF((V1<0.6875 OR V1≥0.7969) AND(V18≥0.4375)) THEN <i>PD</i>
<i>mammography</i>	IF((V2<0.7032 OR V2≥0.8438)AND(V3<0.9063) AND(V5≥0.1875)) THEN <i>benign</i>
	IF((V3≥0.8282)AND(V4<0.0782 OR V4≥0.4844) AND(V5<0.9219)) THEN <i>malignant</i>

Tabela 22 – Média de *fitness* das 100 execuções (com seeds distintas) para os *datasets*  $D_1$ ,  $D_2$  e  $D_3$ .

	$D_1$			$D_2$			$D_3$		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
IG-CEE	<b>72,08</b> ±0,22	<b>72,71</b> ±0,25	<b>72,68</b> ±0,25	<b>70,23</b> ±0,25	<b>69,63</b> ±0,24	<b>70,86</b> ±0,28	<b>62,78</b> ±0,19	<b>62,82</b> ±0,20	<b>62,95</b> ±0,21
CEE	<b>71,63</b> ±0,26	72,02 ±0,29	<b>72,43</b> ±0,27	69,01 ±0,39	68,83 ±0,32	69,92 ±0,36	60,80 ±0,34	61,03 ±0,37	61,41 ±0,35

média de 72,02 (variando de 71,73 a 72,31). Para as classes 1 e 3, o método IG-CEE obteve médias de 72,08 (variando de 71,86 a 72,30) e 72,68 (variando de 72,43 a 72,93), respectivamente, e o método CEE obteve médias de 71,63 (variando de 71,37 a 71,89) e 72,43 (72,16 a 72,70), respectivamente. Para o *dataset*  $D_2$ , o método IG-CEE obteve melhores resultados em todas classes. Para a classe 1, o mesmo obteve média de 70,23 (variando de 69,98 a 70,48), para a classe 2 obteve 69,63 (variando de 69,39 a 69,87) e para a classe 3 obteve 70,86 (variando de 70,58 a 71,14). Já o método CEE obteve médias de 69,01 (variando de 68,62 a 69,40) para a classe 1, 68,83 (variando de 68,51 a 69,16) para a classe 2 e 69,92 (variando de 69,56 a 70,28) para a classe 3. Por fim, para o *dataset*  $D_3$ ,

o método IG-CEE obteve melhores resultados para as 3 classes. Para a classe 1, obteve média de 62,78 (variando de 62,61 a 62,95), para a classe 2 obteve 62,82 (variando de 62,62 a 63,02) e para a classe 3 obteve 62,95 (variando de 62,74 a 63,16).

A Tabela 23 apresenta os resultados obtidos pelos classificadores tradicionais nos *datasets*  $D_1$ ,  $D_2$  e  $D_3$ . As mesmas configurações da Tabela 19 foram utilizadas para obter os resultados. A primeira coluna apresenta o nome de cada classificador e as colunas restantes trazem os *datasets* analisados com os resultados obtidos para cada classe. Aqui, diferentemente da Tabela 19, os resultados são apresentados para cada classe. Isso porque os *datasets* são multiclasse, neste caso, 3 classes.

Tabela 23 – Média de *fitness* das 100 execuções (com sementes distintas) para os *datasets*  $D_1$ ,  $D_2$  e  $D_3$  usando os classificadores tradicionais.

	$D_1$			$D_2$			$D_3$		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
J48	67,78 ±0,09	68,30 ±0,09	69,01 ±0,07	68,76 ±0,09	68,47 ±0,10	<b>68,96</b> <b>±0,07</b>	62,31 ±0,09	<b>62,33</b> <b>±0,09</b>	<b>62,57</b> <b>±0,10</b>
IBK	64,02 ±0,03	64,42 ±0,10	64,74 ±0,09	64,57 ±0,10	62,98 ±0,03	62,95 ±0,04	58,93 ±0,05	58,34 ±0,09	58,57 ±0,10
NB	<b>68,99</b> <b>±0,03</b>	69,03 ±0,03	<b>69,42</b> <b>±0,10</b>	68,76 ±0,08	68,09 ±0,06	70,00 ±0,00	<b>62,98</b> <b>±0,08</b>	62,01 ±0,02	62,23 ±0,08
SVM	67,71 ±0,11	<b>70,41</b> <b>±0,10</b>	67,92 ±0,10	<b>68,98</b> <b>±0,08</b>	<b>68,88</b> <b>±0,08</b>	<b>68,98</b> <b>±0,07</b>	60,90 ±0,11	60,49 ±0,12	61,38 ±0,12

Para *dataset*  $D_1$ , o método NB obteve melhores resultados para a classe 1, com média de 68,99 (variando de 68,96 a 69,02), e 3, com média de 69,42 (variando de 69,32 a 69,52), e para classe 2 o método SVM obteve melhores resultados com média de 70,41 (variando de 70,31 a 70,51). Para o *dataset*  $D_2$ , o método SVM obteve melhores resultados para as classes 1 e 2 com médias de 68,98 (variando de 68,90 a 69,06) e 68,88 (variando de 68,80 a 68,96). Para a classe 3, os métodos SVM e J48 empataram, com médias de 68,98 (variando de 68,91 a 69,05) e 68,96 (variando de 68,89 a 69,03), respectivamente. Destaca-se a performance do NB por ter obtido baixa variação. Para o *dataset*  $D_3$ , o NB obteve melhores resultados para a classe 1, com média de 62,98 (variando de 62,90 a 63,06), e para as classes 2 e 3, o método J48 atingiu melhores resultados, com médias de 62,33 (variando de 62,24 a 62,42) e 62,57 (variando de 62,47 a 62,67).

Como já mencionado, o IG-CEE tem por objetivo selecionar os melhores atributos através do uso do IG para ativação dos seus genes. Sendo assim, a Tabela 24 apresenta as regras obtidas pelo método com o intuito de verificar o desempenho do mesmo. A análise é feita observando aqueles atributos já mencionados na Tabela 15. Como pode ser observado, o método IG-CEE de fato conseguiu selecionar os atributos mais relevantes para compor as regras. Em todos os *datasets* os atributos que definem as classes foram selecionados, exceto para o *dataset*  $D_2$  na classe 2 onde o atributo 9 não foi selecionado. No entanto, em alguns casos os atributos obtiveram valor “Ausente”. No *dataset*  $D_1$ , o atributo 14 está com valor “Presente” e os atributos 3 e 18 com valor “Ausente”. Para o

*dataset*  $D_2$ , os atributos 16 e 17 estão com valor “Ausente” e os atributos 2, 7, e 11 com valor “Presente”. E para o *dataset*  $D_3$ , os atributos 1, 6 e 15 estão com valor “Presente” e os atributos 4, 5, 8, 10 e 12 com valor “Ausente”. O método CEE, por outro lado, não conseguiu selecionar adequadamente todos os atributos relevantes. Para o *dataset*  $D_1$ , o atributo 14 está presente, o atributo 3 ausente e o restante não foi selecionado para compor a regra. Para o *dataset*  $D_2$ , os atributos 9 e 2 estão presentes nas regras e o restante não foi selecionado. E para o *dataset*  $D_3$ , os atributos 6 e 15 estão presentes nas regras e o restante dos atributos não foi selecionado.

Tabela 24 – Exemplo de regras obtidas pelos métodos IG-CEE e CEE para os datasets sintéticos.

	$D_1$	$D_2$	$D_3$
IG-CEE	<p>Classe 1</p> <p>IF((int01=Ausente) AND(int03=Ausente) AND(int05=Ausente) AND(int07=Ausente) AND(int08=Ausente) AND(int10=Presente) AND(int11=Presente) AND(int14=Presente) AND(int16=Presente) AND(int18=Presente) AND(int19=Presente)) THEN classe 1</p> <p>Classe 2</p> <p>IF((int01=Presente) AND(int03=Presente) AND(int05=Presente) AND(int07=Presente) AND(int08=Presente) AND(int10=Presente) AND(int11=Presente) AND(int14=Presente) AND(int16=Presente) AND(int18=Presente) AND(int19=Presente)) THEN classe 2</p> <p>Classe 3</p> <p>IF((int01=Presente) AND(int03=Presente) AND(int05=Presente) AND(int07=Presente) AND(int08=Presente) AND(int10=Presente) AND(int11=Presente) AND(int14=Presente) AND(int16=Presente) AND(int18=Presente) AND(int19=Presente)) THEN classe 3</p>	<p>Classe 1</p> <p>IF((int01=Presente) AND(int02=Presente) AND(int05=Presente) AND(int07=Presente) AND(int08=Presente) AND(int10=Presente) AND(int11=Presente) AND(int14=Presente) AND(int16=Presente) AND(int18=Presente) AND(int19=Presente)) THEN classe 1</p> <p>Classe 2</p> <p>IF((int01=Presente) AND(int02=Presente) AND(int05=Presente) AND(int07=Presente) AND(int08=Presente) AND(int10=Presente) AND(int11=Presente) AND(int14=Presente) AND(int16=Presente) AND(int18=Presente) AND(int19=Presente)) THEN classe 2</p> <p>Classe 3</p> <p>IF((int01=Presente) AND(int02=Presente) AND(int05=Presente) AND(int07=Presente) AND(int08=Presente) AND(int10=Presente) AND(int11=Presente) AND(int14=Presente) AND(int16=Presente) AND(int18=Presente) AND(int19=Presente)) THEN classe 3</p>	<p>Classe 1</p> <p>IF((int01=Ausente) AND(int04=Presente) AND(int05=Ausente) AND(int06=Presente) AND(int07=Ausente) AND(int08=Presente) AND(int09=Ausente) AND(int10=Ausente) AND(int11=Ausente) AND(int12=Ausente) AND(int13=Presente) AND(int15=Presente)) THEN classe 1</p> <p>Classe 2</p> <p>IF((int01=Ausente) AND(int04=Presente) AND(int05=Ausente) AND(int06=Presente) AND(int07=Ausente) AND(int08=Presente) AND(int09=Ausente) AND(int10=Ausente) AND(int11=Ausente) AND(int12=Presente) AND(int13=Presente) AND(int15=Presente)) THEN classe 2</p> <p>Classe 3</p> <p>IF((int01=Ausente) AND(int04=Presente) AND(int05=Ausente) AND(int06=Presente) AND(int07=Ausente) AND(int08=Presente) AND(int09=Ausente) AND(int10=Ausente) AND(int11=Ausente) AND(int12=Presente) AND(int13=Presente) AND(int15=Presente)) THEN classe 3</p>
CEE	<p>Classe 1</p> <p>IF((int02=Ausente) AND(int03=Ausente) AND(int05=Ausente) AND(int14=Presente) AND(int13=Presente) AND(int14=Presente) AND(int18=Presente) AND(int20=Presente)) THEN Class 1</p> <p>Classe 2</p> <p>IF((int03=Ausente) AND(int04=Ausente) AND(int15=Presente) THEN Class 2</p> <p>Classe 3</p> <p>IF((int03=Ausente) AND(int04=Ausente) AND(int15=Presente) THEN Class 3</p>	<p>Classe 1</p> <p>IF((int02=Presente) AND(int03=Presente) AND(int04=Presente) AND(int15=Presente)) THEN Class 1</p> <p>Classe 2</p> <p>IF((int02=Presente) AND(int03=Presente) AND(int04=Presente) AND(int15=Presente)) THEN Class 2</p> <p>Classe 3</p> <p>IF((int02=Presente) AND(int03=Presente) AND(int04=Presente) AND(int15=Presente)) THEN Class 3</p>	<p>Classe 1</p> <p>IF((int02=Presente) AND(int06=Presente) AND(int12=Presente) AND(int13=Presente) AND(int15=Presente)) THEN Class 1</p> <p>Classe 2</p> <p>IF((int02=Presente) AND(int06=Presente) AND(int12=Presente) AND(int13=Presente) AND(int15=Presente)) THEN Class 2</p> <p>Classe 3</p> <p>IF((int02=Presente) AND(int06=Presente) AND(int12=Presente) AND(int13=Presente) AND(int15=Presente)) THEN Class 3</p>

## 5.3 Avaliação dos Resultados

Nessa seção, os resultados obtidos pelos dois métodos propostos serão discutidos. A análise dos resultados será feita com o objetivo de determinar se as hipóteses previamente levantadas são válidas ou não, além de apontar as limitações das propostas. A subseção 5.3.1 apresenta a análise dos resultados do método BIN-NLCEE e a subseção 5.3.2 apresenta a análise dos resultados do método IG-CEE.

### 5.3.1 Análise dos resultados do método BIN-NLCEE

Foi possível observar, através dos resultados apresentados na seção anterior, que o método BIN-NLCEE atingiu uma performance igual ou melhor que o NLCEE e o CEE. Além disso, o método proposto conseguiu diminuir a variância, se comparado com os outros dois métodos (NLCEE e CEE). Nas 16 comparações feitas entre o método BIN-NLCEE e os AGs correlatos NLCEE e CEE, o método proposto obteve melhores resultados em 7 (43.75%) casos e iguais em 9 (56.25%). Se comparado com os classificadores tradicionais, o método proposto conseguiu resultados superiores em todos os casos. No entanto, é necessário destacar que os classificadores tradicionais conseguem obter variância menor se comparados aos AGs. Apesar dos bons resultados, outro ponto que deve ser mencionado é tempo de execução do algoritmo. Os AGs, por sua vez, podem consumir muito tempo de execução, o que faz com que os classificadores tradicionais sejam superiores neste quesito. Outra ponto negativo que deve ser mencionado, é a necessidade de especificar previamente a quantidade de bits que será utilizado para representar os genes, isso é, mais um parâmetro deve ser especificado antes da execução do AG.

Para compreender com mais detalhes os motivos da melhoria atingida pelo método proposto, entende-se que seja necessário analisar e comparar a convergência dos 3 métodos. Nas Tabelas 20, 21, 22 e 23 é possível observar que houve um ganho na taxa de convergência com o método BIN-NLCEE. Isso pode ser explicado pela forma como o operador de mutação atua nesse método. Para o método BIN-NLCEE a mutação ocorre através da troca do valor do bit, estratégia essa que pode levar o processo de busca para soluções melhores e evitar o ótimo local.

Por fim, as regras geradas pelo método mostraram-se bastante simples, principalmente para o *dataset mammography*, onde um número pequeno de atributos foi utilizado para compor as regras. Além disso, a Tabela 20 mostrou que as regras conseguiram bons valores de *fitness* e os valores de treino e teste são próximos.

### 5.3.2 Análise dos resultados do método IG-CEE

Analizando os resultados na Seção 5.2.2, é possível observar que o método IG-CEE conseguiu atingir melhores valores de *fitness* se comparado ao CEE. Observa-se também

uma menor variância do método proposto em relação ao CEE. Analisando os resultados entre os métodos IG-CEE e CEE globalmente, em 9 comparações, o método IG-CEE obteve melhores resultados em 7 (77.77%) e iguais em 2 (22.22%). Se comparados os resultados com os classificadores tradicionais, o método IG-CEE obteve melhores resultados em quase todos os *datasets* e classes, apenas o método NB obteve resultados iguais ao IG-CEE para a classe 1, de acordo com o intervalo de confiança. Esses resultados podem ser justificados pelo fato do método ter sido capaz de selecionar todos os atributos relevantes para compor as regras, ao contrário do CEE. No entanto, apenas a presença do atributo na regra não garante o sucesso na etapa de classificação. Além da presença do atributo na regra, seu valor deve ser “Presente”. Em alguns casos, como pode ser observado na Tabela 24, o IG-CEE foi capaz de selecionar o atributo mas com valor “Ausente”. Outro ponto que deve ser destacado é o fato de ambos os métodos selecionarem atributos que não são relevantes para a classificação, mesmo utilizando um método para seleção de atributos. O método proposto consegue, através do uso do GI, identificar os atributos mais relevantes, no entanto, não consegue descartar de maneira eficiente os não relevantes. Ainda na Tabela 24, é possível observar que ambos os métodos conseguem produzir regras comprehensíveis, no entanto, o CEE apresenta regras menores.

# CAPÍTULO 6

## Conclusão

Esse trabalho teve como objetivo geral estender e aprimorar duas versões de AGs para mineração de regras de classificação. Para isso foram implementados dois AGs, chamados CEE e NLCEE, e duas outras versões para os mesmos. Primeiro, buscou-se verificar quais seriam os efeitos de uma nova representação cromossômica para o método NLCEE. Optou-se por utilizar a representação binária, trabalhando na hipótese de que a mesmas poderiam influenciar nas taxas de classificação e convergência do algoritmo. Para validar a proposta, foram utilizados 4 *datasets* de domínio médico com distribuição não-linear. Foi utilizado o método de validação cruzada *10-fold* e 100 execuções com sementes randômicas distintas. Os resultados apontaram melhores valores de *fitness* e melhores taxas de convergência se comparadas com o método NLCEE e melhores taxas de classificação se comparado com os classificadores tradicionais utilizados (J48, IBK, NB e SVM).

A segunda proposta buscou investigar uma nova forma de definir o valor do segmento peso ( $W_i$ ). Para isso, foi implementado uma nova versão para o método CEE, onde seu valor de peso é guiado pelo IG do atributo, aumentando as chances dos atributos mais relevantes aparecerem nas regras. Foram utilizados 3 *datasets* sintéticos gerados pelo ambiente *GeNIE* e que também foram utilizados em (do Amaral; Hruschka, 2011). Assim como no primeiro método proposto, foi utilizado o *10-fold cross-validation* e 100 execuções com sementes randômicas distintas. Os resultados para o IG-CEE mostraram que o método é capaz de obter melhores valores de *fitness*, através da seleção dos melhores atributos, se comparado ao CEE e aos classificadores tradicionais utilizados no trabalho. No entanto, apenas a seleção do atributo não é o único fator determinante para o processo de classificação. Para que se obtenha sucesso na classificação, o atributo deve conter o valor “Presente”. Portanto, através das análises feitas no Capítulo anterior, as hipóteses levantadas inicialmente, na Seção 1.3, podem ser avaliadas:

- ❑ O uso da representação binária pode influenciar na capacidade de classificação do algoritmo? Sim, os resultados mostraram que o método proposto conseguiu elevar

seu valor de *fitness*.

- ❑ O uso de representação binária pode apresentar melhor conversão? Sim, os resultados apresentaram uma melhor taxa de convergência para o método proposto na maioria dos casos.
- ❑ Um mecanismo de seleção de atributos codificado nos genes do algoritmo pode gerar melhores taxas de classificação? Sim. Melhores taxas de classificação foram obtidas, devido a seleção dos atributos mais relevantes.
- ❑ Um mecanismo de seleção de atributos codificado nos genes do algoritmo pode gerar regras mais simples? Não. Os resultados mostraram que as regras geradas pelo método proposto são maiores que as regras geradas pelo CEE.
- ❑ Um mecanismo de seleção de atributos codificado nos genes do algoritmo pode, efetivamente, reduzir o número de atributos utilizados? Não. O método proposto deu espaço para mais atributos aparecerem nas regras, com o objetivo de selecionar os mais relevantes.

Como já mencionado no Capítulo 1, os AGs para mineração de regras de classificação vem sendo estudados e estão produzindo resultados interessantes, obtendo boas taxas de classificação e regras comprehensíveis. Portanto, entende-se que seja necessário buscar o melhoramento contínuo dessas abordagens, apontando suas principais dificuldades e propondo melhorias. Com o desenvolvimento desse trabalho, foi possível apontar algumas dificuldades dos métodos CEE e NLCEE e propor duas novas versões para os métodos, trazendo melhorias para os mesmos.

## 6.1 Principais Contribuições

A escolha da representação cromossômica deve ser feita observando as particularidades do problema a ser tratado. A maneira como um cromossomo será representado de fato influencia na performance do algoritmo e isso foi um ponto explorado nesse trabalho. Foi investigado o uso da representação binária em AGs para mineração de regras de classificação, representação essa que até então não foi utilizada para esse segmento. Foi possível constatar que a mesma pode trazer melhores valores de *fitness* e melhores taxas de convergência.

Seguindo na mesma linha de AGs para mineração de regras de classificação, foi proposta uma nova forma de ativação dos genes, buscando selecionar os atributos mais importantes para compor as regras. Para isso, foi utilizada uma medida de seleção de atributos, o Ganho de Informação (GI). Através dos resultados foi possível observar que todos os atributos relevantes foram selecionados e maiores valores de *fitness* foram obtidos.

## 6.2 Trabalhos Futuros

A partir da análise dos resultados e observando as limitações encontradas, algumas modificações podem ser propostas como trabalhos futuros:

- ❑ Explorar novas maneiras de representar o cromossomo binário;
- ❑ Testar diferentes configurações para o método BIN-NLCEE, visto que o mesmo possui um parâmetro extra (número de bits);
- ❑ Executar o IG-CEE proposto para *datasets* com atributos contínuos;
- ❑ Criar um mecanismo para o IG-CEE descartar os atributos irrelevantes, e não apenas selecionar os mais relevantes;
- ❑ Utilizar a mesma abordagem do IG-CEE, porém com uma medida de seleção de atributos diferente;
- ❑ Criar uma versão do IG-CEE para *datasets* com distribuição não-linear;
- ❑ Testar o método BIN-NLCEE em *datasets* multiclasse;
- ❑ Utilizar *datasets* reais de domínio médico e biológico.

## 6.3 Contribuições em Produção Bibliográfica

Foi submetido um artigo com os resultados obtidos pelo método BIN-NLCEE no *Congress on Evolutionary Computation (CEC) 2020 (Qualis A1)*, intitulado *Binary Nonlinear Computational Evolutionary Environment BIN-NLCEE: A Classification-based Genetic Algorithm for Nonlinear Datasets*.

Um segundo artigo, abordando o IG-CEE, está em fase de construção e será submetido ao congresso *ECML/PKDD 2020 (Qualis A1)*.



---

## Referências

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. **Machine learning**, Springer, v. 6, n. 1, p. 37–66, 1991. Disponível em: <<https://doi.org/10.1007/BF00153759>>.

AMARAL, L.; HRUSCHKA, E. Non-linear computational evolutionary environment (nlcee): Building high-level knowledge in complex biological databases. In: **ECML/PKDD-European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. No workshop: Data Mining in Functional Genomics and Proteomics: Current Trends and Future Directions**. [S.l.: s.n.], 2011.

AMARAL, L. R. do. Mestrado em Ciência da Computação, **Mineração de regras para classificação de oncogenes medidos por micro-array utilizando algoritmos genéticos**. [S.l.]: Universidade Federal de Uberlândia (UFU), 2007.

AMARAL, L. R. do; HRUSCHKA, E. R. Gene ontology classification: Building high-level knowledge using genetic algorithms. In: **IEEE. 2010 IEEE Congress on Evolutionary Computation (CEC)**. 2010. p. 1–7. Disponível em: <<https://doi.org/10.1109/CEC.2010.5586011>>.

AMARAL, L. R. do; JR, E. R. H. Transgenic: An evolutionary algorithm operator. **Neurocomputing**, Elsevier, v. 127, p. 104–113, 2014. Disponível em: <<https://doi.org/10.1016/j.neucom.2013.08.037>>.

AMARAL, L. R. do et al. Oncogenes classification measured by microarray using genetic algorithms. In: ACTA PRESS. **Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications**. [S.l.], 2008. p. 79–84.

ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. **Nature genetics**, Nature Publishing Group, v. 25, n. 1, p. 25–29, 2000. Disponível em: <<https://doi.org/10.1038/75556>>.

BENGIO, Y.; GRANDVALET, Y. No unbiased estimator of the variance of k-fold cross-validation. **Journal of machine learning research**, v. 5, n. Sep, p. 1089–1105, 2004.

BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: ACM. **Proceedings of the fifth annual workshop on**

**Computational learning theory.** 1992. p. 144–152. Disponível em: <<https://doi.org/10.1145/130385.130401>>.

CARVALHO, D. R.; FREITAS, A. A. A genetic-algorithm for discovering small-disjunct rules in data mining. **Applied Soft Computing**, Elsevier, v. 2, n. 2, p. 75–88, 2002. Disponível em: <[https://doi.org/10.1016/S1568-4946\(02\)00031-5](https://doi.org/10.1016/S1568-4946(02)00031-5)>.

COLEY, D. A. **An introduction to genetic algorithms for scientists and engineers**. World Scientific Publishing Company, 1999. Disponível em: <<https://doi.org/10.1142/3904>>.

DEEP, K.; THAKUR, M. A new crossover operator for real coded genetic algorithms. **Applied mathematics and computation**, Elsevier, v. 188, n. 1, p. 895–911, 2007. Disponível em: <<https://doi.org/10.1016/j.amc.2006.10.047>>.

do Amaral, L. R.; Hruschka, E. R. Transgenic, an operator for evolutionary algorithms. In: **2011 IEEE Congress of Evolutionary Computation (CEC)**. [s.n.], 2011. p. 1308–1314. Disponível em: <<https://doi.org/10.1109/CEC.2011.5949767>>.

do Amaral, L. R.; Junior, E. R. H. Never-ending learning principles in gene ontology classification using genetic algorithms. In: **2012 IEEE Congress on Evolutionary Computation**. [s.n.], 2012. p. 1–8. ISSN 1941-0026. Disponível em: <<https://doi.org/10.1109/CEC.2012.6256637>>.

DONG, H. et al. A novel hybrid genetic algorithm with granular information for feature selection and optimization. **Applied Soft Computing**, Elsevier, v. 65, p. 33–46, 2018. Disponível em: <<https://doi.org/10.1016/j.asoc.2017.12.048>>.

DUA, D.; GRAFF, C. **UCI Machine Learning Repository**. 2017. Disponível em: <<http://archive.ics.uci.edu/ml>>.

FIDELIS, M. V.; LOPES, H. S.; FREITAS, A. A. Discovering comprehensible classification rules with a genetic algorithm. In: **IEEE. Evolutionary Computation, 2000. Proceedings of the 2000 Congress on**. 2000. v. 1, p. 805–810. Disponível em: <<https://doi.org/10.1109/CEC.2000.870381>>.

FILHO, G. R. Melhoramentos no algoritmo genético construtivo e novas aplicações em problemas de agrupamento. **São José dos Campos, Instituto Nacional de Pesquisas Espaciais (INPE)**, 2001.

FRANK, E.; WITTEN, I. H. Generating accurate rule sets without global optimization. University of Waikato, Department of Computer Science, 1998.

FREITAS, A. A. A survey of evolutionary algorithms for data mining and knowledge discovery. In: **Advances in evolutionary computing**. Springer, 2003. p. 819–845. Disponível em: <[https://doi.org/10.1007/978-3-642-18965-4\\_33](https://doi.org/10.1007/978-3-642-18965-4_33)>.

\_\_\_\_\_. A review of evolutionary algorithms for data mining. In: **Data Mining and Knowledge Discovery Handbook**. Springer, 2009. p. 371–400. Disponível em: <[https://doi.org/10.1007/978-0-387-09823-4\\_19](https://doi.org/10.1007/978-0-387-09823-4_19)>.

\_\_\_\_\_. **Data mining and knowledge discovery with evolutionary algorithms**. [S.l.]: Springer Science & Business Media, 2013.

FREITAS, A. A.; LAVINGTON, S. H. **Mining very large databases with parallel processing**. [S.l.]: Springer Science & Business Media, 1997. v. 9.

GEN, M.; LIN, L. Genetic algorithms. **Wiley Encyclopedia of Computer Science and Engineering**, Wiley Online Library, p. 1–15, 2007. Disponível em: <<https://doi.org/10.1002/9780470050118.ecse169>>.

GOLDBERG, D. E. **Genetic algorithms in search optimization and machine learning**. [S.l.]: Addison-wesley Reading Menlo Park, 1989. v. 412.

HALL, M. et al. The WEKA data mining software: an update. **SIGKDD Explorations**, v. 11, n. 1, p. 10–18, 2009. Disponível em: <<https://doi.org/10.1145/1656274.1656278>>.

HALL, M. A. Correlation-based feature selection for machine learning. University of Waikato Hamilton, 1999.

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. [S.l.]: Elsevier, 2011.

HART, P. E. **Pattern classification and scene analysis**. [S.l.: s.n.], 1973.

HO, T. K. Random decision forests. In: **IEEE. Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on**. 1995. v. 1, p. 278–282. Disponível em: <<https://doi.org/10.1109/ICDAR.1995.598994>>.

HOLLAND, J. H. **Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence**. MIT Press, 1992. Disponível em: <<https://doi.org/10.7551/mitpress/1090.001.0001>>.

JIANG, B. et al. A hybrid feature selection algorithm: Combination of symmetrical uncertainty and genetic algorithms. In: . [S.l.: s.n.], 2008.

LIPOWSKI, A.; LIPOWSKA, D. Roulette-wheel selection via stochastic acceptance. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 391, n. 6, p. 2193–2196, 2012. Disponível em: <<https://doi.org/10.1016/j.physa.2011.12.004>>.

LITTLE, M. A. et al. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. **Biomedical engineering online**, BioMed Central, v. 6, n. 1, p. 23, 2007. Disponível em: <<https://doi.org/10.1186/1475-925X-6-23>>.

MATOS, M. D. S.; AMARAL, L. R. D. Multiple disjunctions rule genetic algorithm (mdrga): Inferring non-linear if-then rules in non-linear datasets. In: **IEEE. 2018 IEEE Congress on Evolutionary Computation (CEC)**. 2018. p. 1–6. Disponível em: <<https://doi.org/10.1109/CEC.2018.8477690>>.

METAWA, N.; HASSAN, M. K.; ELHOSENY, M. Genetic algorithm based model for optimizing bank lending decisions. **Expert Systems with Applications**, Elsevier, v. 80, p. 75–82, 2017. Disponível em: <<https://doi.org/10.1016/j.eswa.2017.03.021>>.

MIRALLES-PECHUÁN, L.; PONCE, H.; MARTÍNEZ-VILLASEÑOR, L. A novel methodology for optimizing display advertising campaigns using genetic algorithms. **Electronic Commerce Research and Applications**, Elsevier, v. 27, p. 39–51, 2018. Disponível em: <<https://doi.org/10.1016/j.elerap.2017.11.004>>.

MITCHELL, M. **An introduction to genetic algorithms**. [S.l.]: MIT press, 1998.

MITCHELL, T. M. **Machine learning**. [S.l.]: McGraw hill, 1997.

MOLGA, M.; SMUTNICKI, C. Test functions for optimization needs, 2005. **URL** <http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf>, 2005.

NOVAKOVIC, J. Using information gain attribute evaluation to classify sonar targets. In: **17th Telecommunications forum TELFOR**. [S.l.: s.n.], 2009. p. 1351–1354.

OLIVEIRA, S. A. d. et al. **Metaheuristicas aplicadas ao planejamento da expansão da transmissão de energia eletrica em ambientes de processamento distribuido**. Tese (Doutorado) — UNICAMP, 2004.

QUINLAN, J. C4. 5: Program for machine learning morgan kaufmann. **San Mateo, CA, USA**, 1993.

Silva, R. G. O.; de Souza Ribeiro, M. W.; do Amaral, L. R. Building high level knowledge from high dimensionality biological dataset (nci60) using genetic algorithms and feature selection strategies. In: **2013 IEEE Congress on Evolutionary Computation**. [s.n.], 2013. p. 578–583. Disponível em: <<https://doi.org/10.1109/CEC.2013.6557620>>.

SMILE, D. M. Structural modeling, inference, and learning engine and genie: A development environment for graphical decision-theoretic models. In: **Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)**, July. [S.l.: s.n.], 1999. p. 18–22.

SOKOLOV, A.; WHITLEY, D. Unbiased tournament selection. In: CITESEER. **GECCO**. 2005. v. 5, p. 1131–1138. Disponível em: <<https://doi.org/10.1145/1068009.1068198>>.

SRINIVAS, N.; DEB, K. Multiobjective optimization using nondominated sorting in genetic algorithms. **Evolutionary computation**, MIT Press, v. 2, n. 3, p. 221–248, 1994. Disponível em: <<https://doi.org/10.1162/evco.1994.2.3.221>>.

SUCH, F. P. et al. Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. **arXiv preprint arXiv:1712.06567**, 2017.

TAN, P.-N. **Introduction to data mining**. [S.l.]: Pearson Education India, 2018.

WONG, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. **Pattern Recognition**, Elsevier, v. 48, n. 9, p. 2839–2846, 2015. Disponível em: <<https://doi.org/10.1016/j.patcog.2015.03.009>>.

XING, E. P. et al. Feature selection for high-dimensional genomic microarray data. In: CITESEER. **ICML**. [S.l.], 2001. v. 1, p. 601–608.

YANG, C.-H. et al. Ig-ga: a hybrid filter/wrapper method for feature selection of microarray data. **Journal of Medical and Biological Engineering**, v. 30, n. 1, p. 23–28, 2010.

## Apêndices



# APÊNDICE A

## Resultados das execuções

As tabelas a seguir apresentam os valores absolutos dos resultados das execuções de cada método proposto para cada *dataset*. No apêndice A.1 é apresentado os resultados do método BIN-NLCEE para os *datasets* *breast-w*, *diabetes*, *parkinsons* e *mammography*, respectivamente. O apêndice 5.2.2 apresenta os resultados do método IG-CEE para os *datasets* D1, D2 e D3 respectivamente.

### A.1 Resultados das execuções do método BIN-NLCEE

Tabela 25 – Resultados das 100 execuções do BIN-NLCEE com o *breast-w dataset* nas configurações 1, 2 e 3.

Sementes	Conf. 1		Conf. 2		Conf. 3	
	M	B	M	B	M	B
1	0,9948	1	0,9852	0,9751	0,9852	0,975
2	0,9533	0,975	0,9648	0,97	0,9666	0,9755
3	0,9985	0,961	1	0,984	1	0,9784
4	0,9839	0,9744	0,9772	0,9852	0,9857	0,972
5	0,9657	1	0,9525	0,9933	0,967	0,9749
6	0,9775	1	0,9646	0,9744	0,9807	0,9851
7	0,9974	0,9615	0,9974	0,9772	1	0,943
8	0,9672	0,971	0,9533	0,9827	0,9598	0,9827
9	0,9971	0,9735	0,9774	0,9814	1	0,9605
10	0,9905	0,9819	0,9974	0,9974	1	0,9852
11	0,9416	0,9872	0,9431	1	0,9416	1
12	1	0,9836	1	0,9884	0,9964	1
13	0,9985	0,9545	1	0,949	1	0,9865
14	0,9882	0,9713	0,9808	0,9744	0,9955	1

15	0,9679	0,97	0,9635	0,9961	0,9767	0,9431
16	0,9773	1	0,9773	0,9846	0,9773	0,9971
17	0,9829	0,9787	0,9904	0,9633	0,9932	0,9748
18	0,9933	0,9826	0,9966	0,9442	0,9981	0,9714
19	1	0,9919	1	1	1	0,9891
20	0,9843	0,9665	0,9843	0,9638	0,9823	0,9848
21	0,9343	0,9686	0,9424	0,9718	0,9499	0,9802
22	0,9642	0,9952	0,9642	0,9674	0,9642	0,9342
23	0,9321	0,9919	0,9372	0,9845	0,9375	0,9788
24	0,9633	0,9869	0,9618	0,9848	0,9684	1
25	0,9471	0,9806	0,9608	0,99	0,9573	0,9599
26	0,9563	1	0,9549	0,9775	0,9563	0,9598
27	0,9704	1	0,9704	1	0,9648	0,989
28	0,9631	0,9875	0,9637	0,9846	0,9689	1
29	0,9389	0,9875	0,942	1	0,9461	0,9972
30	0,9839	0,9835	0,9834	0,9635	0,9808	0,9864
31	0,9832	1	0,9848	1	0,9812	0,9623
32	0,9988	0,9937	1	0,9577	1	0,9842
33	0,9683	0,9613	0,9736	0,9803	0,9803	0,9767
34	0,9861	1	0,9861	1	0,985	0,9953
35	1	0,9557	0,9986	0,9635	1	0,9737
36	0,971	0,9635	0,9684	0,9794	0,9737	0,9884
37	0,9815	0,969	0,982	0,9872	0,9827	0,996
38	0,9648	0,9944	0,9648	0,9634	0,9648	1
39	0,9323	0,9827	0,9315	1	0,9315	0,9875
40	0,9659	0,9988	0,9635	0,971	0,9629	1
41	0,9593	0,9946	0,9415	1	0,9503	0,9815
42	0,9642	1	0,9642	0,9852	0,9642	0,9813
43	0,9706	0,9706	0,9719	0,98	0,9645	0,9597
44	0,9735	0,9498	0,9827	0,9757	0,9749	0,9808
45	0,9627	0,9462	0,9804	0,9786	0,976	0,9853
46	0,9769	0,9665	0,9805	0,9537	0,9784	0,9642
47	0,9962	0,9861	0,9962	0,9878	0,9937	0,9625
48	0,9355	0,9743	0,9445	0,976	0,9342	0,9832
49	0,9759	0,974	0,9759	0,9958	0,9776	0,9697
50	1	0,962	0,9985	0,9852	0,9971	0,9861
51	0,9483	0,9473	0,9515	1	0,9608	0,9314

52	0,9792	0,9591	0,9772	1	0,9781	0,9831
53	0,9932	1	0,9986	0,9841	0,9905	0,9705
54	0,9632	0,975	0,9855	0,9861	0,9724	0,9419
55	0,9477	0,9939	0,9465	0,9749	0,9615	0,9752
56	0,9672	0,962	0,9729	0,9674	0,9757	0,9486
57	0,9696	0,9987	0,9623	1	0,977	0,9833
58	0,9159	0,9986	0,932	0,9739	0,9276	1
59	0,9614	1	0,9533	0,9814	0,9497	0,9696
60	0,9799	0,9973	0,9778	0,9246	0,9799	1
61	0,9865	0,9694	0,989	0,9868	0,9847	1
62	1	0,9737	1	0,9592	1	0,9784
63	0,9722	0,9848	0,9738	1	0,974	0,9732
64	0,9646	0,9707	0,9645	0,9851	0,9564	1
65	0,9696	0,9702	0,9675	0,9895	0,9772	0,9621
66	0,9589	0,9934	0,9463	0,9556	0,9529	0,9666
67	0,9417	0,9802	0,9344	0,9791	0,9363	1
68	0,975	0,9741	0,975	1	0,975	1
69	0,9849	1	0,9754	0,9837	0,9783	0,9561
70	0,9463	1	0,9318	0,9362	0,9461	0,9895
71	0,956	0,9675	0,9594	0,9986	0,964	0,9819
72	0,9985	0,9852	1	0,998	0,9971	0,9861
73	0,9849	0,9404	0,9857	0,944	0,9857	0,9865
74	1	0,9684	1	0,9825	1	0,9686
75	0,9624	0,9961	0,9568	0,9706	0,954	0,9939
76	0,9194	1	0,9166	0,9961	0,925	0,9749
77	0,9589	0,9799	0,9582	0,9743	0,9696	0,9892
78	0,9612	0,9964	0,9608	0,9027	0,9681	0,9977
79	0,963	1	0,9722	0,9886	0,9658	0,9985
80	0,9978	0,9693	0,9889	0,9853	0,9922	0,9825
81	0,9795	0,97	0,9647	0,9314	0,9597	0,9815
82	0,9808	0,973	0,9678	0,9868	0,9841	0,9853
83	0,9243	1	0,92	0,9531	0,9243	0,953
84	0,985	1	0,9783	0,9985	0,9744	1
85	0,9565	0,9857	0,9565	0,9833	0,9605	0,9512
86	0,9757	0,9592	0,9762	0,9864	0,9831	0,9828
87	0,9695	0,9472	0,9708	0,9831	0,9614	0,9736
88	0,9844	0,9781	1	1	0,9844	0,9839

89	0,9757	1	0,9654	0,9825	0,9857	1
90	0,9714	1	0,9771	1	0,9843	0,9689
91	0,9889	1	0,9852	0,9948	0,9894	0,9608
92	0,966	0,9582	0,953	1	0,9515	0,9963
93	1	1	1	1	1	0,9944
94	0,9722	1	0,9733	0,9847	0,9724	1
95	0,9762	0,9833	0,9623	1	0,9585	0,9551
96	0,9365	0,9839	0,9363	0,9987	0,9365	0,987
97	0,9861	0,9791	0,9721	0,9818	0,9861	0,9767
98	0,9817	0,9722	0,9878	1	0,9835	0,9985
99	0,969	0,9546	0,9542	0,9986	0,9557	0,9614
100	1	0,9772	0,9974	0,9545	0,9987	0,9752
MÉDIA	0,9719	0,9808	0,9709	0,9806	0,9728	0,9796
DESVIO PADRÃO	0,0201	0,0161	0,0203	0,0186	0,0195	0,0166
VARIÂNCIA	0,0004	0,0003	0,0004	0,0003	0,0004	0,0003
INT. CONFIANÇA	0,0039	0,0032	0,0040	0,0037	0,0038	0,0032

Tabela 26 – Resultados das 100 execuções do BIN-NLCEE com o *breast-w dataset* nas configurações 4, 5 e 6.

Sementes	Conf. 4		Conf. 5		Conf. 6	
	M	B	M	B	M	B
1	1	0,9412	0,9911	1	1	0,9656
2	0,9645	0,9973	0,9664	0,9984	0,9657	0,9811
3	1	0,9452	1	1	0,9985	0,9844
4	0,9772	0,9852	0,9772	0,9755	0,9831	0,9839
5	0,9658	0,9506	0,9697	0,9988	0,9737	0,9635
6	0,9742	0,9853	0,9791	0,9969	0,9791	0,9692
7	1	0,9852	0,9987	0,9799	1	1
8	0,9598	0,9676	0,9853	1	0,983	0,9862
9	0,99	1	1	0,9712	0,9957	1
10	0,9946	1	1	1	0,9986	0,9633
11	0,9568	0,9986	0,9553	0,9822	0,9443	1
12	1	0,9828	0,9981	0,9974	0,9948	0,95
13	1	0,9643	1	1	1	1
14	0,997	0,9655	0,9955	1	0,9941	1
15	0,9724	0,9861	0,9753	0,969	0,9783	0,968

16	0,9773	0,9926	0,9773	1	0,9773	0,9986
17	1	0,9714	1	0,9988	1	0,99
18	1	1	0,9966	0,9972	1	0,972
19	1	0,9853	1	0,9783	1	0,9906
20	0,9843	0,9714	0,9843	0,9611	0,9843	0,9506
21	0,9497	0,9975	0,9523	0,9852	0,9408	0,9792
22	0,9642	1	0,9751	0,9865	0,9802	0,9608
23	0,9356	1	0,948	0,9304	0,9443	0,9875
24	0,9699	0,9863	0,9685	0,9824	0,9726	0,9827
25	0,961	0,9974	0,9629	1	0,9581	0,9974
26	0,9534	1	0,9612	0,9973	0,9563	1
27	0,9681	0,981	0,9617	0,9588	0,9704	0,9808
28	0,9722	1	0,9672	0,9706	0,9671	1
29	0,9568	0,9608	0,948	0,975	0,9646	0,9826
30	0,9848	0,9606	0,9839	0,965	0,9831	0,9973
31	0,9848	0,9826	0,9848	0,973	0,9832	0,9693
32	0,9988	0,9984	1	0,9889	1	0,9505
33	0,971	0,9622	0,9723	0,9831	0,967	0,9839
34	0,9861	1	0,9861	0,9832	0,9861	0,9859
35	1	0,9729	1	0,9831	1	0,975
36	0,9723	1	0,9855	1	0,9737	0,9944
37	0,9857	0,9821	0,9842	1	0,9813	0,9526
38	0,9717	1	0,9666	0,9722	0,9648	0,971
39	0,932	0,9725	0,9468	0,9829	0,9325	0,9976
40	0,9614	0,9857	0,9738	0,9857	0,962	0,9917
41	0,9629	0,9948	0,9729	0,9848	0,9647	1
42	0,9642	0,9756	0,9642	0,9846	0,9642	0,9782
43	0,9853	0,9557	0,9853	1	0,9768	0,9623
44	0,9749	0,9614	0,9865	0,9602	0,9749	1
45	0,9737	0,9623	1	0,9865	0,9791	0,9613
46	0,9784	1	0,9821	0,971	0,9754	0,9478
47	0,9975	0,972	0,9975	0,9903	0,9988	0,9864
48	0,9368	0,994	0,9419	0,997	0,9471	0,9622
49	0,981	0,9838	0,9828	0,9737	0,9828	0,997
50	1	0,9829	1	1	1	1
51	0,9623	0,9544	0,9623	1	0,9702	0,9799
52	0,9797	0,9643	0,9781	0,9837	0,9814	0,9957

53	1	0,9737	1	0,9875	1	0,9512
54	0,9829	0,9462	0,9961	0,9925	0,9947	0,9828
55	0,9483	0,9653	0,9632	0,9986	0,9552	0,962
56	0,9772	0,9851	0,9848	1	1	1
57	0,9789	0,9795	0,9813	0,9842	0,9718	0,973
58	0,94	0,986	0,9217	0,9737	0,9304	0,9789
59	0,9578	1	0,9614	0,9839	0,9614	0,9732
60	0,9799	0,9743	0,9799	0,981	0,9799	1
61	0,9927	0,9633	1	0,95	0,9988	0,9696
62	1	0,9709	1	0,9799	1	0,996
63	0,9775	0,9971	0,9827	1	0,9827	0,997
64	0,9643	0,9802	0,9724	0,9584	0,966	0,9969
65	0,9848	0,9828	0,9832	1	0,9848	1
66	0,9525	0,9844	0,9587	0,9793	0,9595	1
67	0,9521	0,9844	0,9526	0,9749	0,9401	0,9716
68	0,975	0,9762	0,975	0,9848	0,9773	0,9682
69	0,9752	0,9773	0,9883	1	0,9834	1
70	0,9597	0,9953	0,9582	0,9675	0,9646	0,9986
71	0,9649	0,9723	0,9762	0,9639	0,9737	0,9837
72	1	1	1	1	1	0,9538
73	0,9942	0,9655	1	0,975	1	1
74	1	0,9966	1	1	1	0,9926
75	0,9706	0,9565	0,964	0,9853	0,9607	0,9964
76	0,9239	1	0,9343	0,9861	0,9224	0,9737
77	0,9673	0,9868	0,9726	0,9703	0,9761	1
78	0,9629	0,9696	0,9756	0,9828	0,9796	0,9749
79	0,9722	1	0,9722	0,9861	0,9641	0,9571
80	0,9978	0,9744	1	0,9703	1	0,9881
81	0,9766	0,9721	0,9814	0,9734	0,9814	0,9824
82	0,9861	0,9833	0,9851	1	0,9841	0,9984
83	0,9213	0,9972	0,9259	0,9865	0,92	1
84	0,9875	1	0,9937	0,9774	0,9875	0,9962
85	0,9565	0,975	0,9591	0,9743	0,9566	0,9659
86	0,9802	0,9697	0,9846	0,9723	0,9787	0,9791
87	0,9722	0,9883	0,9791	1	0,9681	0,9855
88	1	0,9746	1	0,9843	0,9984	0,9855
89	0,9743	0,975	0,9857	0,9649	0,9744	0,9803

90	0,9562	0,9855	0,9857	0,9868	0,9755	1
91	0,9928	1	1	0,9875	0,9983	0,9773
92	0,9644	0,9868	0,9644	1	0,9627	1
93	1	0,9861	1	0,9868	1	0,9843
94	0,9739	0,9988	0,9743	0,9484	0,9871	0,9877
95	0,9637	0,9689	0,9762	1	0,9762	0,9906
96	0,9367	0,9819	0,965	0,9605	0,9607	0,9683
97	0,9791	0,9977	0,9846	0,9932	0,9861	0,9838
98	0,9854	0,9476	0,9878	1	0,9878	1
99	0,9676	0,9723	0,969	1	0,9822	0,9826
100	1	1	0,9987	0,9958	1	0,9783
MÉDIA	0,9752	0,9812	0,9789	0,9845	0,9775	0,9829
DESVIO PADRÃO	0,0190	0,0154	0,0179	0,0144	0,0189	0,0152
VARIÂNCIA	0,0004	0,0002	0,0003	0,0002	0,0004	0,0002
INT. CONFIANÇA	0,0037	0,0030	0,0035	0,0028	0,0037	0,0030

Tabela 27 – Resultados das 100 execuções do BIN-NLCEE com o *diabetes dataset* nas configurações 1, 2 e 3.

Sementes	Conf. 1		Conf. 2		Conf. 3	
	<i>tn</i>	<i>tp</i>	<i>tp</i>	<i>tp</i>	<i>tn</i>	<i>tp</i>
1	0,7828	0,7993	0,7797	0,8547	0,7903	0,8643
2	0,8204	0,8126	0,8193	0,8528	0,8208	0,7735
3	0,8361	0,8473	0,8488	0,7971	0,8534	0,7615
4	0,8144	0,7725	0,8045	0,7989	0,8104	0,7933
5	0,8332	0,8688	0,8332	0,8496	0,8332	0,8288
6	0,8234	0,8217	0,8124	0,8518	0,8212	0,8337
7	0,8541	0,8186	0,8546	0,8312	0,8566	0,7457
8	0,8403	0,8576	0,8386	0,7809	0,8367	0,7887
9	0,8035	0,8271	0,7797	0,7928	0,826	0,8434
10	0,8541	0,7685	0,8686	0,7595	0,8652	0,8064
11	0,8085	0,796	0,8038	0,7873	0,7987	0,7977
12	0,8404	0,8073	0,8404	0,7913	0,8384	0,7952
13	0,7898	0,8398	0,7817	0,7923	0,7888	0,7965
14	0,8604	0,8051	0,8619	0,7922	0,8654	0,8555
15	0,8005	0,8155	0,7809	0,8324	0,7904	0,8259

16	0,791	0,7737	0,7952	0,8415	0,8083	0,8315
17	0,8129	0,9015	0,795	0,8014	0,8291	0,8182
18	0,834	0,7868	0,8336	0,8	0,823	0,8233
19	0,747	0,8451	0,7431	0,7865	0,7629	0,8004
20	0,862	0,7802	0,8633	0,7878	0,8587	0,8072
21	0,8474	0,815	0,8372	0,7934	0,8416	0,852
22	0,8374	0,8098	0,8448	0,8454	0,8398	0,7822
23	0,8772	0,8307	0,8731	0,8368	0,8738	0,791
24	0,8539	0,8166	0,8585	0,7955	0,8756	0,7584
25	0,8419	0,871	0,8414	0,7712	0,844	0,8137
26	0,8327	0,881	0,8385	0,8313	0,8472	0,8404
27	0,8185	0,8623	0,8197	0,8236	0,8182	0,8204
28	0,8165	0,8931	0,8262	0,7606	0,8203	0,8081
29	0,8355	0,8306	0,8279	0,8305	0,8265	0,7823
30	0,9095	0,8437	0,907	0,8462	0,9132	0,7899
31	0,8156	0,8259	0,8122	0,7656	0,8233	0,8273
32	0,7987	0,8247	0,7998	0,803	0,8068	0,8206
33	0,738	0,8443	0,7446	0,8752	0,7585	0,8002
34	0,8171	0,7998	0,8009	0,8246	0,8232	0,9046
35	0,7651	0,7996	0,7469	0,8121	0,7669	0,8447
36	0,7991	0,8211	0,8218	0,7927	0,7997	0,7483
37	0,8332	0,7559	0,8395	0,8435	0,8448	0,8086
38	0,7485	0,8451	0,7482	0,7577	0,754	0,7883
39	0,8747	0,8819	0,8889	0,8176	0,8819	0,8053
40	0,7948	0,844	0,7872	0,822	0,803	0,7691
41	0,8251	0,8419	0,8176	0,8448	0,8209	0,7808
42	0,8691	0,7987	0,8472	0,8519	0,8542	0,8529
43	0,7983	0,786	0,7839	0,8132	0,8069	0,7668
44	0,7932	0,8154	0,7999	0,9187	0,819	0,8437
45	0,8151	0,8197	0,8131	0,7838	0,8461	0,8502
46	0,8252	0,8206	0,8221	0,7893	0,8298	0,7515
47	0,8454	0,7515	0,8668	0,8466	0,8577	0,842
48	0,8171	0,7702	0,8224	0,8405	0,8216	0,857
49	0,845	0,728	0,8439	0,7866	0,8501	0,8077
50	0,7997	0,8392	0,7971	0,8037	0,8026	0,8212
51	0,852	0,7939	0,8709	0,7416	0,8489	0,7776
52	0,8426	0,8148	0,8235	0,8104	0,8347	0,8157

53	0,8093	0,8121	0,8024	0,7897	0,807	0,8223
54	0,7847	0,7974	0,7872	0,8061	0,7927	0,8069
55	0,7773	0,81	0,7762	0,8429	0,7744	0,8558
56	0,7664	0,7773	0,7664	0,7572	0,7844	0,8132
57	0,8614	0,8072	0,8477	0,8026	0,8443	0,8313
58	0,7908	0,7857	0,7832	0,8214	0,7918	0,7598
59	0,8649	0,8193	0,8631	0,782	0,8626	0,8011
60	0,8543	0,8444	0,8601	0,8689	0,8577	0,744
61	0,7832	0,8487	0,775	0,8401	0,7792	0,7971
62	0,8144	0,8162	0,8125	0,7604	0,8125	0,8356
63	0,8146	0,8598	0,8092	0,802	0,8208	0,8298
64	0,8005	0,7751	0,7885	0,8508	0,7993	0,7594
65	0,7908	0,7899	0,7942	0,8161	0,8007	0,7859
66	0,7828	0,7699	0,7933	0,7898	0,7944	0,8077
67	0,7974	0,8494	0,8065	0,8282	0,7936	0,8072
68	0,7881	0,8404	0,7788	0,831	0,8036	0,8487
69	0,854	0,8077	0,8503	0,7755	0,8625	0,8521
70	0,7932	0,831	0,7912	0,7867	0,7966	0,8644
71	0,8315	0,791	0,832	0,8542	0,8362	0,8393
72	0,8545	0,8064	0,8525	0,788	0,8539	0,882
73	0,7894	0,8338	0,7907	0,7578	0,7884	0,8434
74	0,7948	0,7984	0,7789	0,8153	0,7922	0,7908
75	0,8844	0,7841	0,8762	0,8916	0,885	0,7611
76	0,8491	0,8308	0,8375	0,8269	0,8544	0,8255
77	0,7601	0,868	0,7594	0,8153	0,7787	0,7968
78	0,8435	0,8127	0,8375	0,8428	0,8408	0,7974
79	0,897	0,7868	0,8871	0,806	0,8957	0,7851
80	0,9032	0,8289	0,8877	0,8181	0,8998	0,8001
81	0,7895	0,7963	0,7932	0,7665	0,8205	0,8451
82	0,8379	0,8126	0,8419	0,8641	0,8568	0,8437
83	0,8588	0,8445	0,8421	0,8054	0,8526	0,7951
84	0,8262	0,803	0,8055	0,8292	0,8158	0,8551
85	0,7982	0,7891	0,8103	0,8643	0,8184	0,875
86	0,8486	0,8099	0,845	0,8953	0,845	0,8082
87	0,8225	0,8051	0,8234	0,7983	0,8143	0,8149
88	0,841	0,803	0,8534	0,7974	0,8454	0,7671
89	0,8266	0,8054	0,8295	0,8028	0,8371	0,8319

90	0,7838	0,7783	0,789	0,7815	0,7988	0,806
91	0,8803	0,803	0,8685	0,7697	0,8745	0,8604
92	0,843	0,8224	0,8242	0,797	0,8266	0,7735
93	0,7616	0,8121	0,7419	0,7523	0,757	0,8123
94	0,7894	0,839	0,7787	0,7869	0,7849	0,82
95	0,8072	0,8245	0,8085	0,8165	0,8108	0,8299
96	0,8679	0,8297	0,8693	0,8174	0,8693	0,7945
97	0,8329	0,8128	0,8137	0,831	0,8414	0,8236
98	0,8481	0,764	0,8787	0,761	0,8521	0,8032
99	0,8184	0,7719	0,8236	0,7969	0,8296	0,7781
100	0,8478	0,816	0,8544	0,8549	0,8469	0,7772
MÉDIA	0,8226	0,8155	0,8203	0,8122	0,8263	0,8117
DESVIO PADRÃO	0,0347	0,0313	0,0364	0,0342	0,0328	0,0327
VARIÂNCIA	0,0012	0,0010	0,0013	0,0012	0,0011	0,0011
INT. CONFIANÇA	0,0068	0,0061	0,0071	0,0067	0,0064	0,0064

Tabela 28 – Resultados das 100 execuções do BIN-NLCEE com o *diabetes dataset* nas configurações 3, 4 e 5.

Sementes	Conf. 4		Conf. 5		Conf. 6	
	<i>tn</i>	<i>tp</i>	<i>tp</i>	<i>tp</i>	<i>tn</i>	<i>tp</i>
1	0,7831	0,813	0,7992	0,8521	0,7907	0,8725
2	0,8168	0,7954	0,8279	0,7716	0,8284	0,8951
3	0,8516	0,8397	0,8577	0,8125	0,8615	0,8239
4	0,8095	0,7875	0,8217	0,8705	0,8154	0,8263
5	0,8332	0,8138	0,8332	0,8052	0,8332	0,8598
6	0,8195	0,7528	0,8381	0,7778	0,8248	0,7954
7	0,8559	0,7832	0,8562	0,8194	0,8565	0,9054
8	0,8351	0,7952	0,8483	0,8459	0,8446	0,7962
9	0,8161	0,7994	0,8132	0,8349	0,7918	0,8162
10	0,8637	0,8062	0,8691	0,7835	0,8639	0,8488
11	0,7976	0,8221	0,7917	0,8511	0,7992	0,8316
12	0,8364	0,8194	0,8404	0,8158	0,8551	0,8473
13	0,8036	0,7983	0,7949	0,8432	0,7873	0,8115
14	0,8673	0,7416	0,8671	0,7493	0,8661	0,7976
15	0,7873	0,7892	0,7985	0,8045	0,7931	0,7394

16	0,8027	0,8242	0,8179	0,7878	0,7966	0,7876
17	0,8179	0,7671	0,8165	0,8388	0,8192	0,7923
18	0,8242	0,7713	0,8415	0,844	0,8415	0,84
19	0,7438	0,8138	0,7586	0,8437	0,764	0,8146
20	0,8567	0,7933	0,8633	0,7537	0,8633	0,8612
21	0,8454	0,813	0,8652	0,8108	0,8539	0,8189
22	0,8404	0,8284	0,8426	0,9389	0,844	0,8995
23	0,8844	0,8841	0,8831	0,7936	0,8755	0,8125
24	0,8636	0,8048	0,8749	0,8472	0,866	0,7666
25	0,8485	0,8105	0,85	0,8172	0,8501	0,7735
26	0,8407	0,805	0,8478	0,8411	0,8456	0,82
27	0,8289	0,8276	0,8288	0,8539	0,8112	0,797
28	0,8351	0,7486	0,8207	0,7962	0,8349	0,8256
29	0,8228	0,8295	0,8275	0,7827	0,8258	0,8136
30	0,9132	0,8183	0,9207	0,865	0,9225	0,8279
31	0,8264	0,7664	0,8421	0,8012	0,8315	0,81
32	0,7938	0,824	0,8125	0,8762	0,8003	0,8086
33	0,7615	0,8346	0,7658	0,8348	0,7671	0,8135
34	0,8154	0,8156	0,8301	0,8163	0,8292	0,8683
35	0,7649	0,7493	0,775	0,8264	0,7759	0,7954
36	0,808	0,8276	0,8331	0,8745	0,8307	0,8402
37	0,8463	0,8259	0,8512	0,8327	0,8512	0,7961
38	0,7501	0,8138	0,7694	0,7551	0,7544	0,8417
39	0,8889	0,8432	0,8813	0,8392	0,8911	0,8304
40	0,7892	0,7787	0,8034	0,7883	0,8085	0,901
41	0,8209	0,7879	0,8306	0,8654	0,8197	0,8334
42	0,8523	0,7851	0,8809	0,81	0,8781	0,8116
43	0,8265	0,8452	0,8357	0,7885	0,8162	0,8452
44	0,8049	0,8317	0,7901	0,8052	0,8242	0,87
45	0,8246	0,8131	0,8212	0,8765	0,8371	0,7853
46	0,8266	0,805	0,8301	0,7755	0,8288	0,8286
47	0,8638	0,7916	0,8638	0,7783	0,8547	0,841
48	0,8206	0,8361	0,8284	0,8374	0,831	0,8153
49	0,8405	0,8286	0,8508	0,806	0,8544	0,8237
50	0,7988	0,7812	0,8048	0,8404	0,7997	0,7968
51	0,8589	0,8011	0,8544	0,8547	0,8709	0,8718
52	0,8357	0,7434	0,8439	0,8129	0,8337	0,8188

53	0,8033	0,8554	0,8228	0,779	0,8341	0,8452
54	0,7991	0,7541	0,7956	0,7913	0,788	0,8033
55	0,765	0,7869	0,7744	0,7919	0,7803	0,8128
56	0,7654	0,7937	0,7683	0,8015	0,768	0,78
57	0,8506	0,8363	0,8583	0,791	0,8632	0,8149
58	0,7963	0,7861	0,7963	0,837	0,7881	0,8106
59	0,8599	0,7425	0,8727	0,8322	0,8641	0,8683
60	0,8707	0,8096	0,8525	0,7808	0,869	0,8237
61	0,7835	0,8368	0,8141	0,8317	0,7995	0,806
62	0,8152	0,785	0,8125	0,8291	0,8119	0,8743
63	0,8125	0,85	0,8228	0,7849	0,8165	0,8354
64	0,7956	0,8072	0,7965	0,8277	0,8166	0,7952
65	0,7911	0,8179	0,8034	0,8232	0,8019	0,7755
66	0,7885	0,8059	0,8062	0,8384	0,8075	0,8067
67	0,8075	0,8232	0,8061	0,8285	0,7974	0,7928
68	0,7906	0,79	0,7945	0,8275	0,783	0,8496
69	0,86	0,7963	0,8697	0,7985	0,8656	0,7847
70	0,7939	0,8118	0,7949	0,8228	0,8121	0,8255
71	0,8343	0,8516	0,8409	0,8499	0,8362	0,7876
72	0,8515	0,861	0,8518	0,8314	0,8754	0,8325
73	0,7893	0,8588	0,791	0,7861	0,791	0,7611
74	0,8017	0,8377	0,7971	0,8444	0,8085	0,8227
75	0,8525	0,7817	0,8924	0,8486	0,8775	0,7937
76	0,8515	0,7892	0,8743	0,8201	0,8593	0,7979
77	0,774	0,8731	0,7802	0,8536	0,7859	0,7958
78	0,8435	0,7935	0,8475	0,8869	0,8488	0,8077
79	0,8893	0,8468	0,9	0,8376	0,8977	0,8739
80	0,8859	0,8629	0,8978	0,7926	0,9055	0,8482
81	0,7928	0,8552	0,8136	0,839	0,8023	0,8387
82	0,833	0,7898	0,8493	0,8783	0,8613	0,7738
83	0,8503	0,7798	0,857	0,8678	0,8598	0,7715
84	0,8194	0,8081	0,8319	0,7828	0,8126	0,9146
85	0,818	0,7845	0,8584	0,8481	0,8099	0,7756
86	0,8414	0,8232	0,8726	0,8135	0,8533	0,8417
87	0,8136	0,833	0,8348	0,8495	0,8268	0,7583
88	0,8563	0,841	0,8598	0,8557	0,8591	0,7847
89	0,8187	0,7875	0,8251	0,8413	0,8205	0,8407

90	0,7729	0,8163	0,7906	0,8038	0,7875	0,8328
91	0,8822	0,8529	0,8853	0,801	0,8978	0,9342
92	0,8336	0,828	0,8542	0,8052	0,8416	0,8575
93	0,7644	0,7783	0,8635	0,8517	0,7692	0,863
94	0,781	0,8368	0,8114	0,7976	0,795	0,7775
95	0,8143	0,7471	0,8147	0,791	0,834	0,7597
96	0,8625	0,7824	0,8731	0,8217	0,8633	0,858
97	0,8155	0,7787	0,8344	0,8328	0,8217	0,8229
98	0,8715	0,779	0,8693	0,8522	0,8893	0,7714
99	0,8254	0,8503	0,8411	0,8466	0,8275	0,8033
100	0,8423	0,8185	0,8538	0,8446	0,8583	0,8047
MÉDIA	0,8239	0,8084	0,8334	0,8234	0,8306	0,8217
DESVIO PADRÃO	0,0339	0,0310	0,0334	0,0325	0,0349	0,0369
VARIÂNCIA	0,0011	0,0010	0,0011	0,0011	0,0012	0,0014
INT. CONFIANÇA	0,0066	0,0061	0,0065	0,0064	0,0068	0,0072

Tabela 29 – Resultados das 100 execuções do BIN-NLCEE com o *parkinsons dataset* nas configurações 1, 2 e 3.

Sementes	Conf. 1		Conf. 2		Conf. 3	
	<i>healthy</i>	<i>PD</i>	<i>PD</i>	<i>healthy</i>	<i>healthy</i>	<i>healthy</i>
1	0,9665	0,9471	0,9846	0,98	0,9962	1
2	1	0,9773	1	0,9619	0,9964	0,9941
3	0,9683	0,9833	0,9683	0,9031	0,9766	0,9655
4	0,9375	0,9697	0,9375	1	0,9375	0,9258
5	1	0,9697	1	0,9923	1	1
6	1	1	1	0,8964	1	0,9777
7	1	0,99	1	0,9423	1	0,9577
8	1	1	0,9964	0,9964	1	0,9558
9	1	1	1	1	1	1
10	1	0,911	1	1	1	0,8875
11	1	0,9695	1	1	1	0,9655
12	1	0,9607	1	0,9818	1	0,5
13	1	0,9393	0,9906	0,9488	1	1
14	1	0,884	0,9964	0,9893	0,9655	0,9629
15	0,99	0,9031	1	0,9615	1	0,9937

16	1	1	1	0,9697	1	0,9853
17	1	0,9769	1	0,9629	1	0,9962
18	0,9375	0,9792	1	0,975	1	0,9564
19	1	0,9629	0,9964	0,9688	1	0,99
20	1	0,9971	1	0,9643	1	1
21	1	0,8894	1	0,87	1	0,5
22	0,9583	1	1	0,8887	0,9583	1
23	1	0,9644	1	0,9467	1	1
24	1	0,9393	1	0,9265	1	1
25	1	1	1	0,9449	1	0,9295
26	0,9545	1	0,9545	0,9354	1	1
27	1	0,9473	1	1	1	0,9745
28	0,965	0,9655	0,9889	1	0,9622	0,9665
29	1	1	1	0,9769	1	0,8887
30	0,9853	0,909	0,9728	0,9643	0,9853	1
31	0,9821	0,9962	1	0,9321	1	0,931
32	1	0,9476	1	1	1	0,9688
33	1	0,9629	1	0,9246	1	0,9226
34	1	0,9969	0,9643	1	1	0,9964
35	1	0,9792	1	0,9577	1	0,9941
36	1	0,9199	0,9844	0,9354	1	0,9933
37	1	1	0,99	1	0,97	0,9667
38	1	1	1	1	1	0,951
39	1	0,9523	0,9933	0,9677	1	0,9634
40	0,9804	1	0,9637	0,9677	0,9494	1
41	1	0,8633	1	0,895	1	1
42	1	1	1	1	1	0,9861
43	1	1	1	0,9629	1	0,9944
44	0,9923	1	1	0,9677	0,9923	0,9615
45	1	0,9861	1	0,9853	0,9467	0,9065
46	1	0,9697	1	0,9258	1	0,9937
47	1	0,9575	0,9777	1	1	1
48	1	0,9246	1	1	1	0,9258
49	0,9591	1	0,9591	0,9031	1	1
50	0,955	0,941	0,95	0,9677	0,95	0,9853
51	1	0,9923	1	0,9697	1	0,9542
52	1	0,9923	1	0,9544	1	0,9655

53	1	0,9167	1	0,9417	1	0,9571
54	0,9141	0,9755	0,9141	0,9599	0,9186	1
55	0,9507	1	0,9607	0,9655	0,9643	0,9591
56	1	0,9714	1	0,9697	0,9167	0,9583
57	1	0,931	1	0,9875	1	0,95
58	0,9659	0,9599	0,9944	0,892	0,9523	0,919
59	1	1	1	0,9971	1	0,9962
60	1	0,9577	1	0,9808	1	0,9167
61	0,95	0,9655	0,95	0,9688	0,95	0,9962
62	1	0,9756	1	0,9274	1	0,9312
63	1	0,9656	1	0,9967	1	0,9655
64	1	0,9697	1	0,9647	1	0,9552
65	1	0,9311	0,9853	0,9643	1	1
66	0,9783	0,8929	0,9783	0,9516	0,9584	0,9608
67	0,9643	0,9354	0,9679	1	0,9393	1
68	0,9383	1	0,9542	0,9544	0,9701	0,9286
69	1	0,9412	1	0,9875	1	0,9629
70	1	1	1	0,9271	1	0,9179
71	0,9375	0,9697	0,9375	0,9955	0,9375	1
72	1	0,9511	1	0,9621	1	1
73	1	0,9677	1	0,941	1	1
74	0,9467	0,931	0,9433	0,9591	0,9933	1
75	1	0,8594	1	0,9677	1	0,9104
76	0,9615	0,9464	0,9962	1	0,9808	0,9502
77	1	0,9752	1	0,9354	1	0,9615
78	0,9955	0,9362	0,9955	0,9583	0,9955	0,9697
79	1	0,9677	1	0,9563	1	0,9929
80	1	1	1	1	0,9933	0,9967
81	0,9844	0,8969	0,9844	0,9906	1	0,9619
82	1	0,9129	1	0,9354	1	0,9655
83	0,9701	0,9719	0,9808	0,9929	0,9885	0,9417
84	1	1	1	1	1	1
85	0,9564	0,9411	0,9547	0,9655	0,9547	1
86	1	0,9655	0,9962	0,9655	0,9923	0,9532
87	1	1	1	1	1	1
88	0,9833	1	0,9802	1	0,9933	0,9688
89	1	0,9258	0,95	0,9596	1	0,9958

90	0,9615	0,9348	0,9571	0,9286	0,9846	0,9354
91	1	1	1	0,95	1	0,9697
92	1	0,9523	1	0,9563	1	0,9543
93	0,9821	0,9906	0,951	1	0,9821	0,957
94	1	1	1	0,9644	1	1
95	1	0,9773	1	1	1	0,9824
96	1	1	1	0,875	1	0,8964
97	1	0,9792	1	0,961	1	0,9549
98	0,9572	0,931	0,9348	0,9697	0,9847	0,9923
99	0,9375	0,9697	1	1	1	0,9677
100	1	0,9719	1	0,8893	1	0,8798
MÉDIA	0,9877	0,9643	0,9884	0,9638	0,9894	0,9597
DESVIO PADRÃO	0,0208	0,0342	0,0198	0,0325	0,0201	0,0728
VARIÂNCIA	0,0004	0,0012	0,0004	0,0011	0,0004	0,0053
INT. CONFIANÇA	0,0041	0,0067	0,0039	0,0064	0,0039	0,0143

Tabela 30 – Resultados das 100 execuções do BIN-NLCEE com o *parkinsons dataset* nas configurações 4, 5 e 6.

Sementes	Conf. 4		Conf. 5		Conf. 6	
	<i>healthy</i>	<i>PD</i>	<i>healthy</i>	<i>PD</i>	<i>healthy</i>	<i>PD</i>
1	0,9962	0,9167	1	0,9667	1	0,9629
2	1	0,9314	1	0,9885	1	0,9821
3	0,9808	1	0,9808	0,9655	0,9683	0,9964
4	0,9375	0,9488	0,9375	1	1	1
5	1	1	1	0,9969	1	1
6	1	1	1	1	1	0,9619
7	1	0,9929	1	1	1	0,9655
8	1	1	1	0,9937	1	1
9	1	1	1	0,9844	1	0,9665
10	1	0,9393	0,9964	1	1	1
11	1	0,9667	1	0,9514	1	0,9643
12	1	0,91	1	1	1	1
13	0,9969	0,9629	1	0,9867	1	1
14	1	1	1	1	0,9964	1
15	0,9821	0,9137	0,9929	0,9786	1	0,9655

16	1	0,9677	1	0,9473	1	1
17	1	1	1	0,9967	1	0,931
18	0,9375	0,9714	0,9375	1	1	0,9677
19	1	0,9552	1	1	1	1
20	1	0,9714	1	0,9875	1	0,9885
21	1	0,9967	1	1	1	1
22	1	0,9415	1	0,9677	0,9583	1
23	1	0,9375	1	1	1	0,9619
24	1	0,9467	1	0,95	1	0,9688
25	1	1	1	1	0,9955	0,9448
26	0,9545	0,9814	0,9682	0,9577	1	0,9909
27	1	0,9792	1	0,9814	1	0,9545
28	0,95	1	0,96	0,9962	0,98	0,9964
29	1	0,9794	1	1	1	0,9629
30	1	1	1	1	0,9822	0,951
31	1	0,98	1	0,9677	1	0,9644
32	1	0,9629	1	0,9552	1	0,9359
33	1	0,9962	1	0,931	1	1
34	1	1	1	1	1	0,9777
35	1	0,8839	1	0,951	0,9583	0,9629
36	1	0,9769	1	0,9792	1	0,9577
37	0,9733	0,9599	1	0,9655	0,9967	0,9405
38	0,9344	0,9765	1	0,9564	1	1
39	1	1	1	1	1	1
40	0,9533	0,9964	0,9616	0,9231	0,9875	0,9962
41	0,9583	0,9199	1	0,9967	1	1
42	1	0,8581	1	0,9629	1	0,9969
43	1	0,903	1	0,9655	1	1
44	0,9801	1	0,9962	0,9354	0,9962	0,9742
45	1	0,9964	1	0,9143	1	0,9667
46	1	0,9941	1	1	1	0,9498
47	0,9967	0,9604	1	0,9906	1	0,9288
48	1	1	1	0,9619	1	0,9577
49	0,9629	1	0,9701	1	1	0,9555
50	0,95	0,9514	0,95	0,9833	0,95	0,9964
51	1	1	1	1	1	0,9479
52	1	0,9677	1	0,9929	1	0,9333

53	1	0,9697	1	0,9327	1	0,95
54	0,9252	1	0,9538	0,9643	0,9236	0,9297
55	0,9893	1	0,9964	0,9533	0,9476	1
56	1	0,9615	0,9167	0,9058	1	1
57	1	0,9599	1	1	1	0,9104
58	0,98	0,9655	0,99	0,9736	1	0,9756
59	1	1	1	0,9599	1	0,9393
60	1	1	1	0,5	1	1
61	0,95	1	1	0,9875	1	0,9885
62	1	0,9629	1	0,9577	1	1
63	1	0,9354	1	1	1	1
64	1	0,9548	1	1	1	0,9655
65	1	0,9095	1	0,951	1	0,9655
66	0,9821	0,9677	0,9821	0,951	1	0,9354
67	1	1	1	0,9625	0,9643	0,9199
68	0,9792	0,9797	0,9564	0,9451	0,9564	0,9514
69	1	0,9117	1	0,9344	1	0,9864
70	1	0,9808	0,9375	1	0,9344	0,9655
71	1	0,9409	0,9375	0,9599	0,9375	0,9125
72	1	0,9933	1	0,9655	1	1
73	1	0,9697	1	1	1	1
74	0,9867	0,9286	1	1	0,9933	0,9479
75	1	0,8474	1	1	1	1
76	1	0,9129	0,9846	0,5	0,9846	0,9469
77	1	0,9354	1	0,9867	1	0,9964
78	1	0,9857	1	0,9677	0,9955	0,931
79	1	0,9566	1	0,9967	1	0,9482
80	1	0,9599	1	1	1	1
81	1	0,9634	1	1	1	0,9599
82	1	0,9393	1	0,9644	1	1
83	0,9923	0,8433	0,9885	0,9375	0,9962	0,9655
84	1	0,9451	1	1	1	0,9893
85	0,9564	0,9697	0,9564	0,9655	0,9547	0,9798
86	0,9821	1	0,9772	0,99	0,9962	1
87	1	1	1	0,9655	1	0,9312
88	0,961	0,9861	0,99	1	0,9933	1
89	1	1	1	1	1	0,9788

90	1	0,9115	0,9615	0,9615	0,9923	0,8594
91	1	0,9937	1	0,9548	1	0,9844
92	1	0,9437	1	1	1	1
93	0,9783	0,8804	0,9783	0,9923	0,9745	0,9887
94	1	1	1	0,9199	1	0,9812
95	1	0,9937	1	0,9591	1	1
96	1	1	1	1	1	1
97	1	1	1	0,9655	1	0,9964
98	0,9576	0,9677	0,9909	0,9629	0,9564	1
99	1	0,96	1	0,9542	1	0,9345
100	1	1	1	0,9933	1	1
MÉDIA	0,9906	0,9658	0,9915	0,9667	0,9927	0,9734
DESVIO PADRÃO	0,0183	0,0369	0,0182	0,0712	0,0167	0,0281
VARIÂNCIA	0,0003	0,0014	0,0003	0,0051	0,0003	0,0008
INT. CONFIANÇA	0,0036	0,0072	0,0036	0,0140	0,0033	0,0055

Tabela 31 – Resultados das 100 execuções do BIN-NLCEE com o *mammography dataset* nas configurações 1, 2 e 3.

Sementes	Conf. 1		Conf. 2		Conf. 3	
	<i>benign</i>	<i>malignant</i>	<i>benign</i>	<i>malignant</i>	<i>benign</i>	<i>malignant</i>
1	0,8849	0,8642	0,8906	0,8655	0,8934	0,8105
2	0,8744	0,8173	0,8757	0,8159	0,8757	0,8966
3	0,8498	0,7992	0,8498	0,8334	0,8438	0,8104
4	0,8646	0,8306	0,8714	0,7959	0,8633	0,8532
5	0,8492	0,822	0,8427	0,8582	0,8311	0,8745
6	0,8339	0,8289	0,8334	0,8568	0,8312	0,859
7	0,8552	0,8584	0,8559	0,8432	0,8557	0,8984
8	0,8962	0,782	0,9013	0,8279	0,8962	0,8233
9	0,8631	0,8733	0,8625	0,7765	0,8344	0,8543
10	0,8636	0,8174	0,8606	0,8331	0,8636	0,9002
11	0,8668	0,8735	0,8752	0,8279	0,872	0,8207
12	0,79	0,8472	0,8192	0,8338	0,808	0,8823
13	0,8141	0,8097	0,8138	0,8035	0,814	0,8565
14	0,8571	0,7959	0,8571	0,823	0,8571	0,8122
15	0,7993	0,8303	0,7931	0,8188	0,7939	0,821

16	0,7694	0,8407	0,7688	0,8255	0,7694	0,8037
17	0,9107	0,8371	0,8996	0,8361	0,9298	0,8329
18	0,8459	0,8115	0,8465	0,8251	0,846	0,8281
19	0,861	0,8154	0,8601	0,8629	0,8704	0,8541
20	0,8771	0,8555	0,8793	0,8467	0,8795	0,8619
21	0,8609	0,7866	0,8644	0,9171	0,8614	0,8575
22	0,8218	0,8252	0,8197	0,8648	0,8155	0,8305
23	0,8563	0,764	0,8563	0,8721	0,8551	0,8054
24	0,8654	0,8729	0,8654	0,8566	0,8627	0,877
25	0,848	0,8429	0,8426	0,8288	0,8448	0,8376
26	0,9003	0,9234	0,8973	0,8487	0,8953	0,8235
27	0,8546	0,8524	0,8516	0,8315	0,8487	0,8114
28	0,8555	0,8242	0,8547	0,8354	0,8576	0,7796
29	0,8017	0,8181	0,8037	0,8316	0,8017	0,8012
30	0,821	0,8333	0,821	0,8441	0,8015	0,8209
31	0,895	0,8283	0,895	0,8436	0,895	0,8442
32	0,8589	0,8385	0,8594	0,8511	0,8745	0,8523
33	0,8802	0,7881	0,8854	0,84	0,8854	0,8277
34	0,8451	0,8271	0,8477	0,7863	0,8426	0,9042
35	0,8256	0,8656	0,8259	0,8768	0,8259	0,8631
36	0,8192	0,8615	0,8237	0,839	0,8124	0,8362
37	0,887	0,8543	0,8839	0,8778	0,8795	0,8589
38	0,8795	0,8847	0,8795	0,7935	0,8032	0,8337
39	0,8047	0,8818	0,8045	0,8452	0,8774	0,8324
40	0,8758	0,8421	0,8768	0,8742	0,9057	0,8112
41	0,9085	0,8322	0,9004	0,8335	0,8939	0,7985
42	0,8902	0,8153	0,8952	0,8784	0,8453	0,8436
43	0,8453	0,8759	0,8453	0,834	0,8203	0,8535
44	0,8343	0,8561	0,8276	0,8858	0,8189	0,9003
45	0,8194	0,8334	0,819	0,8481	0,7688	0,8856
46	0,7688	0,8107	0,7644	0,8208	0,8654	0,8441
47	0,8593	0,805	0,8678	0,8492	0,8575	0,8131
48	0,8589	0,8444	0,8604	0,8223	0,8169	0,8893
49	0,816	0,7898	0,8196	0,8336	0,9188	0,8191
50	0,9188	0,8126	0,9188	0,7518	0,8322	0,8474
51	0,8158	0,8256	0,8167	0,8129	0,8541	0,7907
52	0,8529	0,8127	0,8517	0,8543	0,8396	0,8213

53	0,8395	0,7969	0,8403	0,8362	0,8397	0,8609
54	0,8426	0,7983	0,8348	0,839	0,8185	0,8253
55	0,8167	0,8514	0,8189	0,8257	0,8803	0,8508
56	0,8765	0,8116	0,8891	0,8592	0,8845	0,8203
57	0,8845	0,8767	0,8752	0,8419	0,8751	0,7974
58	0,873	0,8088	0,8793	0,7912	0,8628	0,8317
59	0,8692	0,8483	0,8628	0,8974	0,8177	0,8483
60	0,8226	0,8236	0,8177	0,8484	0,8824	0,8457
61	0,8824	0,8651	0,8793	0,8483	0,8473	0,8384
62	0,8534	0,8631	0,8546	0,8142	0,8676	0,7806
63	0,8747	0,8362	0,8683	0,8119	0,8283	0,8882
64	0,8283	0,8247	0,8283	0,7836	0,8872	0,8308
65	0,8895	0,886	0,9139	0,7945	0,8353	0,8281
66	0,8296	0,8008	0,8405	0,8065	0,8848	0,8168
67	0,8816	0,8152	0,878	0,8752	0,8525	0,8388
68	0,8568	0,8654	0,8512	0,8155	0,8122	0,8598
69	0,8122	0,8999	0,8271	0,8092	0,8413	0,864
70	0,8423	0,8048	0,8398	0,8363	0,8793	0,8117
71	0,8788	0,8227	0,8649	0,8181	0,878	0,8132
72	0,8902	0,9173	0,8778	0,8218	0,8664	0,8676
73	0,848	0,8645	0,8445	0,8746	0,8746	0,8742
74	0,8772	0,8666	0,887	0,8858	0,8782	0,8512
75	0,8907	0,8865	0,8862	0,874	0,8578	0,8627
76	0,8578	0,8108	0,8578	0,811	0,8178	0,8439
77	0,8294	0,9063	0,8124	0,8723	0,9286	0,8508
78	0,9225	0,836	0,8961	0,8377	0,8394	0,8664
79	0,8288	0,8546	0,8244	0,8283	0,8482	0,8466
80	0,8532	0,8418	0,8506	0,833	0,8502	0,8728
81	0,8764	0,8312	0,8557	0,8767	0,8459	0,8131
82	0,8458	0,8712	0,8458	0,8657	0,8772	0,8935
83	0,8772	0,8348	0,8772	0,8314	0,7962	0,8677
84	0,7951	0,8884	0,8029	0,8243	0,8882	0,8423
85	0,8893	0,8548	0,8888	0,811	0,8583	0,7924
86	0,8546	0,8274	0,8578	0,8817	0,8561	0,7888
87	0,8561	0,8467	0,8529	0,8621	0,7844	0,8017
88	0,7761	0,8716	0,7844	0,7808	0,8733	0,8873
89	0,8456	0,8291	0,8429	0,8368	0,8466	0,839

90	0,8466	0,7808	0,8466	0,8615	0,8877	0,8077
91	0,8877	0,8569	0,8816	0,8347	0,8657	0,8183
92	0,8804	0,814	0,8649	0,82	0,8299	0,8044
93	0,8249	0,8434	0,8249	0,8624	0,8702	0,8275
94	0,8807	0,8662	0,8705	0,7965	0,7928	0,8837
95	0,7874	0,8099	0,7871	0,8165	0,8496	0,8154
96	0,8496	0,8556	0,8454	0,7953	0,8952	0,8845
97	0,8994	0,8279	0,9001	0,8537	0,8341	0,8067
98	0,8341	0,8116	0,8341	0,8566	0,8823	0,8442
99	0,8769	0,7948	0,8742	0,8426	0,867	0,8761
100	0,8716	0,841	0,8656	0,8523	0,853	0,8409
MÉDIA	0,8538	0,8378	0,8531	0,8375	0,0328	0,0302
DESVIO PADRÃO	0,0325	0,0312	0,0314	0,0286	0,0011	0,0009
VARIÂNCIA	0,0011	0,0010	0,0010	0,0008	0,0064	0,0059
INT. CONFIANÇA	0,0064	0,0061	0,0062	0,0056	0,0033	0,0055

Tabela 32 – Resultados das 100 execuções do BIN-NLCEE com o *mammography dataset* nas configurações 4, 5 e 6.

Sementes	Conf. 4		Conf. 5		Conf. 6	
	<i>benign</i>	<i>malignant</i>	<i>benign</i>	<i>malignant</i>	<i>benign</i>	<i>malignant</i>
1	0,9062	0,8483	0,9025	0,8597	0,9061	0,8517
2	0,8744	0,8177	0,8763	0,8483	0,8757	0,8243
3	0,8498	0,864	0,8465	0,8064	0,8479	0,8905
4	0,8616	0,8445	0,8654	0,7911	0,8655	0,8686
5	0,8382	0,8626	0,8631	0,827	0,8488	0,8114
6	0,8281	0,839	0,8342	0,8277	0,8292	0,8445
7	0,8557	0,8875	0,8594	0,8124	0,8589	0,8506
8	0,8996	0,8401	0,9013	0,8687	0,8996	0,9154
9	0,8593	0,8429	0,8625	0,8586	0,8629	0,8677
10	0,8636	0,8406	0,8677	0,8417	0,8718	0,8512
11	0,8752	0,8006	0,8752	0,8937	0,8752	0,8104
12	0,826	0,7884	0,82	0,9075	0,826	0,8826
13	0,8142	0,7859	0,8352	0,8325	0,8144	0,8468
14	0,8571	0,806	0,8571	0,8337	0,8571	0,8545
15	0,7929	0,7838	0,7948	0,8171	0,7949	0,893

16	0,7679	0,8235	0,7688	0,8825	0,7691	0,8127
17	0,917	0,8302	0,9259	0,8021	0,9245	0,8312
18	0,8465	0,8127	0,847	0,8242	0,8477	0,8514
19	0,8648	0,864	0,8737	0,8312	0,8648	0,8479
20	0,8785	0,8277	0,8798	0,8745	0,8798	0,7858
21	0,8634	0,8316	0,8645	0,7842	0,8651	0,8415
22	0,8197	0,9029	0,8218	0,8342	0,8218	0,827
23	0,8563	0,857	0,8563	0,8363	0,8563	0,8212
24	0,8654	0,8654	0,8654	0,8232	0,8654	0,8543
25	0,848	0,835	0,8417	0,8523	0,848	0,8022
26	0,9067	0,8492	0,9079	0,8787	0,9043	0,8001
27	0,8546	0,7993	0,8546	0,7813	0,8546	0,8277
28	0,8589	0,8797	0,8534	0,8615	0,8576	0,8162
29	0,8017	0,8091	0,8017	0,8021	0,8017	0,7902
30	0,8192	0,8665	0,823	0,8513	0,8175	0,8128
31	0,895	0,8323	0,895	0,7814	0,895	0,8546
32	0,8552	0,8842	0,8618	0,91	0,8705	0,8215
33	0,8761	0,8152	0,8823	0,8291	0,8855	0,818
34	0,8426	0,8709	0,8426	0,8698	0,8426	0,8349
35	0,8259	0,8298	0,8391	0,8286	0,8256	0,8416
36	0,8237	0,8614	0,8237	0,8549	0,8237	0,8144
37	0,9266	0,8291	0,9259	0,7837	0,9237	0,8722
38	0,8795	0,8221	0,8795	0,8853	0,8795	0,8331
39	0,8082	0,8282	0,8046	0,8493	0,8044	0,8808
40	0,8758	0,8629	0,8758	0,8323	0,8758	0,8506
41	0,8968	0,876	0,9063	0,8329	0,9065	0,8077
42	0,8911	0,8485	0,8924	0,8752	0,8902	0,8533
43	0,8453	0,8551	0,8521	0,795	0,8453	0,8865
44	0,8227	0,8693	0,8322	0,8725	0,8361	0,8647
45	0,8197	0,888	0,8197	0,9115	0,8197	0,8117
46	0,7688	0,837	0,7777	0,8341	0,7692	0,8466
47	0,8715	0,8351	0,8703	0,8378	0,8691	0,8028
48	0,8575	0,8577	0,8615	0,8362	0,8615	0,8757
49	0,8068	0,8324	0,8287	0,8038	0,8233	0,8651
50	0,9175	0,8035	0,9188	0,83	0,9188	0,8693
51	0,818	0,8609	0,8216	0,8323	0,8293	0,7924
52	0,8529	0,8558	0,8541	0,7804	0,8541	0,8662

53	0,8395	0,8126	0,8404	0,8274	0,8432	0,8424
54	0,8426	0,8155	0,8329	0,8734	0,8426	0,8347
55	0,8189	0,8637	0,8189	0,8418	0,8189	0,8423
56	0,8891	0,855	0,8891	0,8475	0,8849	0,8576
57	0,8845	0,8656	0,8845	0,8259	0,8845	0,8507
58	0,8751	0,8724	0,8814	0,8866	0,8793	0,7914
59	0,866	0,8086	0,866	0,8639	0,866	0,8322
60	0,8316	0,8584	0,8275	0,8261	0,8275	0,8402
61	0,8793	0,8541	0,8855	0,8302	0,8855	0,8198
62	0,8534	0,849	0,8568	0,8151	0,8571	0,8881
63	0,8676	0,8596	0,8703	0,8906	0,8707	0,864
64	0,8283	0,8421	0,8283	0,8466	0,8283	0,7825
65	0,9157	0,8341	0,9146	0,8803	0,9126	0,8318
66	0,8253	0,8408	0,8275	0,8763	0,8484	0,87
67	0,8901	0,8537	0,8944	0,8453	0,8934	0,8508
68	0,8568	0,8076	0,8585	0,8648	0,8568	0,833
69	0,8241	0,7862	0,8182	0,8335	0,8241	0,8065
70	0,8356	0,8327	0,8453	0,8419	0,8453	0,8855
71	0,8691	0,8551	0,8825	0,8807	0,8821	0,7947
72	0,8841	0,824	0,888	0,8456	0,889	0,8782
73	0,8565	0,8699	0,8553	0,8511	0,8495	0,8362
74	0,883	0,819	0,8849	0,8534	0,8849	0,7985
75	0,8824	0,81	0,8866	0,8657	0,8902	0,8434
76	0,8578	0,8485	0,8578	0,8464	0,8578	0,8147
77	0,8245	0,8625	0,833	0,8242	0,8257	0,8091
78	0,9132	0,7984	0,9275	0,8215	0,9272	0,8232
79	0,8227	0,8533	0,8458	0,8409	0,8305	0,8082
80	0,8367	0,8874	0,8503	0,8112	0,8513	0,7896
81	0,8801	0,8202	0,8837	0,8164	0,8758	0,8444
82	0,8474	0,8491	0,8492	0,8463	0,8489	0,8431
83	0,8772	0,8409	0,8772	0,8241	0,8772	0,8389
84	0,7988	0,8111	0,8077	0,8303	0,7981	0,8771
85	0,8882	0,8866	0,89	0,8711	0,89	0,8201
86	0,8555	0,8268	0,8661	0,8441	0,8587	0,8787
87	0,8464	0,8141	0,8561	0,8347	0,8561	0,8347
88	0,7844	0,8644	0,7916	0,9041	0,7844	0,8358
89	0,8668	0,8112	0,8768	0,8369	0,8599	0,8204

90	0,819	0,7907	0,8327	0,7817	0,8426	0,8878
91	0,8877	0,8232	0,8834	0,8118	0,8834	0,8056
92	0,8755	0,7991	0,8755	0,8466	0,8706	0,7783
93	0,8249	0,822	0,8516	0,8446	0,8247	0,8267
94	0,8751	0,8069	0,8852	0,8857	0,874	0,8658
95	0,7958	0,8763	0,7939	0,8735	0,8168	0,8002
96	0,8496	0,7666	0,8636	0,8453	0,8538	0,8334
97	0,9023	0,8863	0,9059	0,8007	0,9042	0,8263
98	0,8341	0,8259	0,8341	0,861	0,8341	0,8141
99	0,8883	0,7928	0,8969	0,8669	0,8744	0,7713
100	0,8716	0,8287	0,8706	0,832	0,8706	0,8776
MÉDIA	0,8547	0,8384	0,8586	0,8423	0,8572	0,8375
DESVIO PADRÃO	0,0334	0,0284	0,0329	0,0303	0,0328	0,0301
VARIÂNCIA	0,0011	0,0008	0,0011	0,0009	0,0011	0,0009
INT. CONFIANÇA	0,0065	0,0056	0,0065	0,0059	0,0064	0,0059

## A.2 Resultados das execuções do método IG-CEE

Tabela 33 – Resultados das 100 execuções do IG-CEE com o *dataset* D1 nas configurações 1 e 2.

Sementes	Conf.1			Conf. 2		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,7367	0,7043	0,7568	0,7367	0,7387	0,7239
2	0,7135	0,7524	0,7309	0,7135	0,7063	0,7044
3	0,7285	0,7359	0,7158	0,7285	0,7064	0,7169
4	0,732	0,732	0,726	0,732	0,7452	0,7094
5	0,7319	0,7527	0,7194	0,7319	0,7379	0,7126
6	0,7374	0,7095	0,7243	0,7374	0,7289	0,7401
7	0,7125	0,7076	0,7384	0,7125	0,7213	0,7547
8	0,7294	0,7275	0,7181	0,7294	0,7261	0,6986
9	0,712	0,7158	0,7422	0,712	0,708	0,7395
10	0,7311	0,7264	0,7001	0,7311	0,7196	0,7144
11	0,7124	0,7168	0,7345	0,7124	0,7244	0,7349
12	0,7196	0,7266	0,7388	0,7196	0,7092	0,7276
13	0,7069	0,7401	0,711	0,7069	0,7026	0,7394

14	0,7339	0,7303	0,726	0,7339	0,7256	0,7194
15	0,7224	0,7434	0,7393	0,7224	0,7219	0,7512
16	0,7105	0,717	0,7335	0,7105	0,7203	0,7177
17	0,7033	0,7351	0,7141	0,7033	0,7344	0,7205
18	0,7257	0,7256	0,725	0,7257	0,7273	0,7214
19	0,7281	0,74	0,7272	0,7281	0,7223	0,7145
20	0,7348	0,7376	0,7337	0,7348	0,733	0,7389
21	0,7344	0,7251	0,706	0,7344	0,7388	0,7233
22	0,7147	0,7239	0,7161	0,7147	0,7428	0,7324
23	0,7279	0,745	0,727	0,7279	0,7101	0,7192
24	0,7321	0,7032	0,735	0,7321	0,7016	0,7191
25	0,7364	0,7183	0,7357	0,7364	0,7374	0,7205
26	0,7272	0,7435	0,7116	0,7272	0,7285	0,7178
27	0,7153	0,7064	0,7208	0,7153	0,7021	0,7274
28	0,7057	0,7287	0,7125	0,7057	0,7452	0,719
29	0,7311	0,7153	0,71	0,7311	0,7137	0,7499
30	0,717	0,7424	0,7381	0,717	0,716	0,7144
31	0,7069	0,7418	0,7225	0,7069	0,7236	0,7447
32	0,7368	0,7096	0,7269	0,7368	0,7392	0,738
33	0,7231	0,7502	0,6942	0,7231	0,7282	0,7426
34	0,7216	0,7307	0,7215	0,7216	0,7328	0,7096
35	0,7101	0,7202	0,7305	0,7101	0,7197	0,7465
36	0,7425	0,7476	0,7369	0,7425	0,7179	0,7232
37	0,7183	0,7372	0,7419	0,7183	0,7142	0,7179
38	0,7331	0,7237	0,7377	0,7331	0,73	0,7127
39	0,6973	0,7013	0,7414	0,6973	0,7275	0,6909
40	0,7241	0,719	0,7213	0,7241	0,7441	0,7318
41	0,7115	0,7247	0,7335	0,7115	0,7356	0,717
42	0,7213	0,714	0,7076	0,7213	0,7199	0,7168
43	0,7293	0,7314	0,7463	0,7293	0,7556	0,7365
44	0,7303	0,7431	0,7367	0,7303	0,7309	0,713
45	0,7088	0,7162	0,722	0,7088	0,7248	0,7094
46	0,7234	0,7272	0,7081	0,7234	0,7635	0,7299
47	0,7132	0,7438	0,7106	0,7132	0,7421	0,7255
48	0,7232	0,7463	0,7325	0,7232	0,7296	0,7258
49	0,7053	0,7195	0,7261	0,7053	0,7286	0,7218
50	0,722	0,7229	0,7145	0,722	0,7241	0,738

51	0,714	0,716	0,7353	0,714	0,729	0,7375
52	0,7159	0,7284	0,7143	0,7159	0,7262	0,7301
53	0,7059	0,7267	0,7525	0,7059	0,7349	0,7256
54	0,7247	0,7352	0,7156	0,7247	0,7327	0,7393
55	0,715	0,7032	0,7289	0,715	0,7375	0,7134
56	0,7372	0,7306	0,7162	0,7372	0,7183	0,7135
57	0,7221	0,7357	0,7281	0,7221	0,7188	0,7429
58	0,7158	0,7408	0,7465	0,7158	0,7469	0,7168
59	0,72	0,7387	0,7221	0,72	0,7404	0,7374
60	0,7309	0,7197	0,7442	0,7309	0,727	0,7384
61	0,6965	0,7087	0,7143	0,6965	0,7276	0,7156
62	0,7184	0,7141	0,7122	0,7184	0,7083	0,7308
63	0,7181	0,7036	0,7461	0,7181	0,7437	0,7331
64	0,7212	0,6973	0,7478	0,7212	0,7306	0,7369
65	0,688	0,7189	0,7481	0,688	0,7333	0,7193
66	0,7345	0,7351	0,7357	0,7345	0,7167	0,7369
67	0,7445	0,7258	0,7419	0,7445	0,735	0,7062
68	0,7212	0,7362	0,7097	0,7212	0,7191	0,7455
69	0,7219	0,7339	0,7313	0,7219	0,6999	0,7374
70	0,7189	0,7218	0,7325	0,7189	0,7231	0,7461
71	0,7243	0,7326	0,7199	0,7243	0,7181	0,7313
72	0,7151	0,7318	0,7425	0,7151	0,7296	0,7209
73	0,7055	0,726	0,7163	0,7055	0,7179	0,7279
74	0,7174	0,7248	0,7479	0,7174	0,7174	0,7372
75	0,7303	0,7285	0,72	0,7303	0,7336	0,717
76	0,7201	0,7488	0,7398	0,7201	0,7406	0,7129
77	0,7114	0,7426	0,7113	0,7114	0,7039	0,7216
78	0,7394	0,7158	0,7279	0,7394	0,7236	0,7228
79	0,7099	0,7081	0,711	0,7099	0,7489	0,7301
80	0,716	0,7256	0,7262	0,716	0,7155	0,7109
81	0,715	0,7372	0,72	0,715	0,701	0,7315
82	0,7225	0,727	0,7331	0,7225	0,7182	0,7096
83	0,7198	0,7318	0,7225	0,7198	0,7269	0,7112
84	0,7285	0,7393	0,7091	0,7285	0,7232	0,7131
85	0,7115	0,7198	0,7246	0,7115	0,7378	0,7387
86	0,7068	0,7373	0,7223	0,7068	0,7219	0,7082
87	0,7079	0,7362	0,7152	0,7079	0,7161	0,7294

88	0,7318	0,7265	0,7445	0,7318	0,725	0,7551
89	0,702	0,737	0,7248	0,702	0,7312	0,7238
90	0,7363	0,7325	0,6978	0,7363	0,7281	0,7094
91	0,7091	0,7341	0,7235	0,7091	0,729	0,7299
92	0,7168	0,7227	0,7461	0,7168	0,7176	0,7064
93	0,7267	0,7218	0,7368	0,7267	0,7233	0,7174
94	0,7316	0,7349	0,7239	0,7316	0,7425	0,7374
95	0,7317	0,7287	0,7316	0,7317	0,7099	0,7077
96	0,7193	0,7156	0,7194	0,7193	0,7333	0,7245
97	0,7304	0,7366	0,7258	0,7304	0,7307	0,725
98	0,7266	0,712	0,7406	0,7266	0,7244	0,7247
99	0,6994	0,7153	0,7326	0,6994	0,7081	0,7278
100	0,7304	0,7104	0,7238	0,7304	0,7324	0,735
MÉDIA	0,7208	0,7271	0,7268	0,7208	0,7261	0,7254
DESVIO PADRÃO	0,0112	0,0127	0,0129	0,0112	0,0124	0,0129
VARIÂNCIA	0,0001	0,0002	0,0002	0,0001	0,0002	0,0002
INT. CONFIANÇA	0,0022	0,0025	0,0025	0,0022	0,0024	0,0025

Tabela 34 – Resultados das 100 execuções do IG-CEE com o *dataset* D1 nas configurações 3 e 4.

Sementes	Conf. 3			Conf. 4		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,7367	0,7129	0,7292	0,7367	0,73	0,7225
2	0,7135	0,7407	0,7242	0,7135	0,7375	0,7364
3	0,7285	0,71	0,7153	0,7285	0,7167	0,7184
4	0,732	0,7095	0,7285	0,732	0,7089	0,7232
5	0,7319	0,7174	0,7237	0,7319	0,7281	0,7352
6	0,7374	0,7229	0,7096	0,7374	0,7085	0,739
7	0,7125	0,7375	0,7276	0,7125	0,7237	0,7336
8	0,7294	0,7312	0,723	0,7294	0,715	0,7392
9	0,712	0,7439	0,745	0,712	0,7337	0,7245
10	0,7311	0,7222	0,7239	0,7311	0,7312	0,7117
11	0,7124	0,7175	0,7161	0,7124	0,7387	0,7339
12	0,7196	0,7538	0,7109	0,7196	0,7288	0,7198
13	0,7069	0,7083	0,7273	0,7069	0,6991	0,7209
14	0,7339	0,718	0,713	0,7339	0,7211	0,7242

15	0,7224	0,7232	0,7288	0,7224	0,7253	0,7049
16	0,7105	0,7237	0,7131	0,7105	0,6947	0,7295
17	0,7033	0,7201	0,7356	0,7033	0,6908	0,7437
18	0,7257	0,7185	0,7432	0,7257	0,7242	0,7323
19	0,7281	0,7215	0,7136	0,7281	0,7146	0,7118
20	0,7348	0,7031	0,7215	0,7348	0,7256	0,7223
21	0,7344	0,743	0,7119	0,7344	0,7219	0,751
22	0,7147	0,7275	0,7348	0,7147	0,7143	0,745
23	0,7279	0,7323	0,721	0,7279	0,7095	0,7264
24	0,7321	0,7282	0,7219	0,7321	0,7162	0,711
25	0,7364	0,7033	0,7408	0,7364	0,7526	0,7293
26	0,7272	0,7337	0,7242	0,7272	0,7405	0,7327
27	0,7153	0,727	0,7414	0,7153	0,7255	0,714
28	0,7057	0,7122	0,7119	0,7057	0,7315	0,7615
29	0,7311	0,7192	0,7214	0,7311	0,7489	0,7224
30	0,717	0,7414	0,7301	0,717	0,7213	0,7299
31	0,7069	0,7594	0,7104	0,7069	0,7191	0,7387
32	0,7368	0,7132	0,7363	0,7368	0,7217	0,7229
33	0,7231	0,7204	0,7176	0,7231	0,7409	0,7511
34	0,7216	0,7313	0,7367	0,7216	0,7118	0,7304
35	0,7101	0,7155	0,7244	0,7101	0,7384	0,7433
36	0,7425	0,7245	0,7244	0,7425	0,7282	0,7238
37	0,7183	0,7289	0,7262	0,7183	0,7361	0,7183
38	0,7331	0,7226	0,719	0,7331	0,7385	0,7331
39	0,6973	0,7328	0,7241	0,6973	0,7448	0,7332
40	0,7241	0,7313	0,7371	0,7241	0,7537	0,7325
41	0,7115	0,6947	0,7557	0,7115	0,7019	0,7362
42	0,7213	0,7236	0,74	0,7213	0,7511	0,7332
43	0,7293	0,713	0,7378	0,7293	0,7219	0,7394
44	0,7303	0,7218	0,7211	0,7303	0,7282	0,7361
45	0,7088	0,7306	0,7226	0,7088	0,724	0,7246
46	0,7234	0,7152	0,7339	0,7234	0,7125	0,7023
47	0,7132	0,7207	0,7236	0,7132	0,7267	0,723
48	0,7232	0,7316	0,7071	0,7232	0,7144	0,7215
49	0,7053	0,7282	0,7149	0,7053	0,7364	0,7199
50	0,722	0,7367	0,7133	0,722	0,7269	0,7224
51	0,714	0,7299	0,7168	0,714	0,7397	0,738

52	0,7159	0,7114	0,7399	0,7159	0,7272	0,7165
53	0,7059	0,728	0,7213	0,7059	0,6946	0,7425
54	0,7247	0,7175	0,726	0,7247	0,7211	0,738
55	0,715	0,7118	0,721	0,715	0,7232	0,711
56	0,7372	0,7285	0,7305	0,7372	0,7456	0,7395
57	0,7221	0,7129	0,6997	0,7221	0,7334	0,7214
58	0,7158	0,732	0,7343	0,7158	0,7315	0,7182
59	0,72	0,7343	0,7165	0,72	0,7061	0,7282
60	0,7309	0,7266	0,728	0,7309	0,7134	0,7075
61	0,6965	0,7288	0,7318	0,6965	0,7249	0,7301
62	0,7184	0,7271	0,7313	0,7184	0,7341	0,7083
63	0,7181	0,6984	0,7534	0,7181	0,7303	0,7602
64	0,7212	0,7112	0,7153	0,7212	0,7191	0,7525
65	0,688	0,7312	0,7215	0,688	0,7264	0,7273
66	0,7345	0,7152	0,726	0,7345	0,7175	0,7068
67	0,7445	0,7264	0,7185	0,7445	0,7284	0,7187
68	0,7212	0,7321	0,7237	0,7212	0,7132	0,7247
69	0,7219	0,7163	0,7238	0,7219	0,7261	0,7146
70	0,7189	0,7341	0,7289	0,7189	0,727	0,7281
71	0,7243	0,7219	0,737	0,7243	0,7073	0,715
72	0,7151	0,7426	0,7555	0,7151	0,7213	0,7396
73	0,7055	0,7281	0,7131	0,7055	0,7305	0,7288
74	0,7174	0,7283	0,7317	0,7174	0,7137	0,7173
75	0,7303	0,7313	0,7662	0,7303	0,7346	0,7263
76	0,7201	0,7124	0,7482	0,7201	0,7069	0,7431
77	0,7114	0,7005	0,7311	0,7114	0,7281	0,7179
78	0,7394	0,7125	0,7273	0,7394	0,7361	0,7256
79	0,7099	0,7482	0,7179	0,7099	0,7208	0,711
80	0,716	0,727	0,7234	0,716	0,7125	0,7343
81	0,715	0,7401	0,7519	0,715	0,7281	0,7355
82	0,7225	0,7294	0,7369	0,7225	0,7232	0,7291
83	0,7198	0,7239	0,7238	0,7198	0,7431	0,7173
84	0,7285	0,7258	0,7283	0,7285	0,7356	0,7049
85	0,7115	0,7292	0,7424	0,7115	0,6849	0,7155
86	0,7068	0,7391	0,7349	0,7068	0,7241	0,7352
87	0,7079	0,725	0,7153	0,7079	0,7179	0,735
88	0,7318	0,725	0,7242	0,7318	0,7357	0,7182

89	0,702	0,7368	0,7115	0,702	0,7375	0,7195
90	0,7363	0,7496	0,7225	0,7363	0,7332	0,7238
91	0,7091	0,7258	0,7329	0,7091	0,735	0,7259
92	0,7168	0,7323	0,7039	0,7168	0,7472	0,7231
93	0,7267	0,7263	0,7379	0,7267	0,7263	0,7277
94	0,7316	0,7272	0,7247	0,7316	0,7231	0,7398
95	0,7317	0,7198	0,7173	0,7317	0,7294	0,7346
96	0,7193	0,719	0,7305	0,7193	0,7305	0,7337
97	0,7304	0,718	0,747	0,7304	0,7356	0,7169
98	0,7266	0,723	0,719	0,7266	0,72	0,707
99	0,6994	0,7158	0,7047	0,6994	0,7333	0,7063
100	0,7304	0,7251	0,733	0,7304	0,7119	0,7117
MÉDIA	0,7208	0,7248	0,7266	0,7208	0,7251	0,7269
DESVIO PADRÃO	0,0112	0,0115	0,0121	0,0112	0,0132	0,0122
VARIÂNCIA	0,0001	0,0001	0,0001	0,0001	0,0002	0,0001
INT. CONFIANÇA	0,0022	0,0022	0,0024	0,0022	0,0026	0,0024

Tabela 35 – Resultados das 100 execuções do IG-CEE com o *dataset* D1 nas configurações 5 e 6.

Sementes	Conf. 5			Conf. 6		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,7367	0,724	0,7532	0,7367	0,7275	0,7039
2	0,7135	0,7244	0,7231	0,7135	0,7262	0,721
3	0,7285	0,7117	0,7222	0,7285	0,7315	0,7349
4	0,732	0,7234	0,7312	0,732	0,7182	0,7318
5	0,7319	0,7256	0,7306	0,7319	0,7152	0,7139
6	0,7374	0,7297	0,733	0,7374	0,7254	0,7348
7	0,7125	0,736	0,7307	0,7125	0,7231	0,7346
8	0,7294	0,736	0,7394	0,7294	0,7113	0,7077
9	0,712	0,7273	0,7161	0,712	0,7342	0,7157
10	0,7311	0,7332	0,7107	0,7311	0,695	0,7351
11	0,7124	0,7255	0,7181	0,7124	0,7226	0,7069
12	0,7196	0,7393	0,7191	0,7196	0,7172	0,7509
13	0,7069	0,7212	0,7185	0,7069	0,7119	0,7136
14	0,7339	0,7538	0,7141	0,7339	0,7237	0,7271

15	0,7224	0,7203	0,72	0,7224	0,7256	0,7323
16	0,7105	0,7282	0,7386	0,7105	0,7346	0,7481
17	0,7033	0,7368	0,7087	0,7033	0,7227	0,735
18	0,7257	0,7343	0,7184	0,7257	0,7163	0,7481
19	0,7281	0,727	0,7431	0,7281	0,7326	0,7125
20	0,7348	0,7259	0,7186	0,7348	0,7325	0,7159
21	0,7344	0,7248	0,7264	0,7344	0,7266	0,7314
22	0,7147	0,7363	0,7294	0,7147	0,7361	0,7131
23	0,7279	0,7317	0,7392	0,7279	0,7343	0,7218
24	0,7321	0,7351	0,7294	0,7321	0,7361	0,7231
25	0,7364	0,7093	0,7213	0,7364	0,7345	0,7329
26	0,7272	0,7042	0,732	0,7272	0,7443	0,721
27	0,7153	0,7244	0,71	0,7153	0,7296	0,7267
28	0,7057	0,7151	0,7408	0,7057	0,7109	0,7253
29	0,7311	0,7237	0,7161	0,7311	0,7197	0,7326
30	0,717	0,7284	0,7231	0,717	0,7315	0,7142
31	0,7069	0,7282	0,7188	0,7069	0,7291	0,7151
32	0,7368	0,7512	0,7313	0,7368	0,7359	0,7302
33	0,7231	0,7081	0,7278	0,7231	0,7131	0,7023
34	0,7216	0,7258	0,7275	0,7216	0,7437	0,7204
35	0,7101	0,7425	0,7076	0,7101	0,7297	0,7317
36	0,7425	0,7165	0,7453	0,7425	0,7268	0,7406
37	0,7183	0,7268	0,7388	0,7183	0,7173	0,7156
38	0,7331	0,7331	0,7179	0,7331	0,727	0,7425
39	0,6973	0,7136	0,7237	0,6973	0,7194	0,7163
40	0,7241	0,7295	0,7122	0,7241	0,6972	0,7181
41	0,7115	0,7358	0,734	0,7115	0,7372	0,7344
42	0,7213	0,7109	0,7387	0,7213	0,7418	0,7212
43	0,7293	0,738	0,7205	0,7293	0,7169	0,6888
44	0,7303	0,7375	0,7226	0,7303	0,7388	0,7267
45	0,7088	0,724	0,7255	0,7088	0,7285	0,7427
46	0,7234	0,7187	0,7275	0,7234	0,7012	0,7275
47	0,7132	0,7186	0,725	0,7132	0,712	0,7246
48	0,7232	0,7197	0,7065	0,7232	0,7337	0,737
49	0,7053	0,7226	0,7456	0,7053	0,7262	0,7321
50	0,722	0,7317	0,7286	0,722	0,7351	0,6968
51	0,714	0,721	0,7346	0,714	0,7249	0,72

52	0,7159	0,7341	0,7172	0,7159	0,7292	0,7237
53	0,7059	0,7157	0,7183	0,7059	0,7343	0,7274
54	0,7247	0,7344	0,7162	0,7247	0,718	0,7272
55	0,715	0,7141	0,7227	0,715	0,7418	0,7247
56	0,7372	0,7254	0,7251	0,7372	0,7257	0,7274
57	0,7221	0,7032	0,7469	0,7221	0,724	0,7181
58	0,7158	0,7215	0,7338	0,7158	0,7419	0,7461
59	0,72	0,7349	0,7418	0,72	0,7425	0,7362
60	0,7309	0,7137	0,7369	0,7309	0,7256	0,7026
61	0,6965	0,7071	0,7247	0,6965	0,7275	0,7109
62	0,7184	0,7237	0,7325	0,7184	0,7081	0,7287
63	0,7181	0,7025	0,7462	0,7181	0,7243	0,7275
64	0,7212	0,7441	0,7213	0,7212	0,7426	0,7113
65	0,688	0,7245	0,7113	0,688	0,725	0,7169
66	0,7345	0,7198	0,7188	0,7345	0,7269	0,7505
67	0,7445	0,7503	0,7319	0,7445	0,7205	0,7256
68	0,7212	0,7294	0,7184	0,7212	0,7364	0,7104
69	0,7219	0,7258	0,7197	0,7219	0,7225	0,7303
70	0,7189	0,7137	0,7167	0,7189	0,7255	0,7313
71	0,7243	0,73	0,7239	0,7243	0,7115	0,7243
72	0,7151	0,7393	0,7265	0,7151	0,7235	0,7305
73	0,7055	0,7148	0,7185	0,7055	0,7242	0,7326
74	0,7174	0,7358	0,7352	0,7174	0,6996	0,7493
75	0,7303	0,7309	0,7232	0,7303	0,7216	0,7324
76	0,7201	0,7382	0,7369	0,7201	0,7236	0,7344
77	0,7114	0,7138	0,7331	0,7114	0,71	0,7337
78	0,7394	0,717	0,732	0,7394	0,709	0,7319
79	0,7099	0,709	0,7206	0,7099	0,7143	0,7354
80	0,716	0,7355	0,7114	0,716	0,7286	0,7141
81	0,715	0,7129	0,7255	0,715	0,7238	0,7261
82	0,7225	0,746	0,7147	0,7225	0,7364	0,7374
83	0,7198	0,745	0,7127	0,7198	0,7034	0,7125
84	0,7285	0,7182	0,7196	0,7285	0,7359	0,7112
85	0,7115	0,7165	0,7307	0,7115	0,7099	0,7124
86	0,7068	0,7159	0,7206	0,7068	0,7384	0,7237
87	0,7079	0,7491	0,7251	0,7079	0,7333	0,7221
88	0,7318	0,7317	0,7282	0,7318	0,721	0,7067

89	0,702	0,7103	0,7313	0,702	0,7328	0,7367
90	0,7363	0,7232	0,7212	0,7363	0,7053	0,7169
91	0,7091	0,7105	0,7099	0,7091	0,7405	0,7117
92	0,7168	0,7306	0,7341	0,7168	0,7312	0,7344
93	0,7267	0,7432	0,7027	0,7267	0,7345	0,732
94	0,7316	0,7057	0,7112	0,7316	0,7358	0,7301
95	0,7317	0,7321	0,7137	0,7317	0,725	0,701
96	0,7193	0,7235	0,7371	0,7193	0,7165	0,7282
97	0,7304	0,7149	0,737	0,7304	0,7135	0,719
98	0,7266	0,7356	0,7093	0,7266	0,7335	0,7218
99	0,6994	0,7387	0,7306	0,6994	0,7021	0,7218
100	0,7304	0,7318	0,7227	0,7304	0,7547	0,7265
MÉDIA	0,7208	0,7261	0,7253	0,7208	0,7252	0,7248
DESVIO PADRÃO	0,0112	0,0114	0,0104	0,0112	0,0115	0,0121
VARIÂNCIA	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
INT. CONFIANÇA	0,0022	0,0022	0,0020	0,0022	0,0023	0,0024

Tabela 36 – Resultados das 100 execuções do IG-CEE com o *dataset* D2 nas configurações 1 e 2.

Sementes	Conf. 1			Conf. 2		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,6898	0,696	0,733	0,6898	0,6955	0,7199
2	0,6748	0,6998	0,7055	0,6748	0,7267	0,6998
3	0,6971	0,6936	0,7073	0,6971	0,7006	0,7224
4	0,6855	0,6943	0,7309	0,6855	0,6987	0,6885
5	0,6948	0,7237	0,7312	0,6948	0,698	0,706
6	0,72	0,6916	0,6951	0,72	0,7004	0,6977
7	0,7053	0,7071	0,7062	0,7053	0,6874	0,7272
8	0,6939	0,7107	0,7061	0,6939	0,7142	0,7134
9	0,7027	0,6962	0,7027	0,7027	0,694	0,7148
10	0,6883	0,6873	0,7224	0,6883	0,6923	0,6947
11	0,6981	0,682	0,7175	0,6981	0,6777	0,7067
12	0,7064	0,702	0,6932	0,7064	0,7082	0,7089
13	0,6805	0,7094	0,7238	0,6805	0,68	0,6892
14	0,6878	0,7036	0,6936	0,6878	0,695	0,7186

15	0,7269	0,6898	0,7	0,7269	0,683	0,7364
16	0,6935	0,6764	0,71	0,6935	0,7097	0,7117
17	0,7016	0,6842	0,6939	0,7016	0,712	0,6907
18	0,7061	0,6967	0,7005	0,7061	0,7208	0,6896
19	0,7037	0,7053	0,7009	0,7037	0,6821	0,6885
20	0,6989	0,7108	0,7172	0,6989	0,6937	0,6957
21	0,7006	0,7082	0,7043	0,7006	0,6929	0,7115
22	0,6899	0,7071	0,7207	0,6899	0,6814	0,726
23	0,6955	0,703	0,7096	0,6955	0,7118	0,7498
24	0,6943	0,6901	0,6935	0,6943	0,6902	0,714
25	0,7138	0,7013	0,7071	0,7138	0,6824	0,6856
26	0,7147	0,6885	0,6824	0,7147	0,7084	0,7218
27	0,7067	0,6893	0,7129	0,7067	0,6883	0,7026
28	0,7166	0,7095	0,7134	0,7166	0,6795	0,7158
29	0,681	0,6735	0,7224	0,681	0,6906	0,7192
30	0,6888	0,7105	0,6955	0,6888	0,6782	0,7291
31	0,7174	0,6938	0,7114	0,7174	0,6906	0,7141
32	0,6969	0,7201	0,6927	0,6969	0,6875	0,7069
33	0,7031	0,7139	0,7092	0,7031	0,6904	0,6864
34	0,7064	0,6864	0,7069	0,7064	0,71	0,7084
35	0,7155	0,6887	0,7041	0,7155	0,6768	0,7156
36	0,705	0,6825	0,6976	0,705	0,7142	0,6779
37	0,7329	0,6853	0,7103	0,7329	0,703	0,7304
38	0,7115	0,6816	0,6993	0,7115	0,6849	0,7222
39	0,6952	0,6831	0,7028	0,6952	0,6912	0,6871
40	0,7119	0,6671	0,7082	0,7119	0,714	0,7019
41	0,7034	0,7032	0,6783	0,7034	0,7073	0,7002
42	0,7045	0,6852	0,7029	0,7045	0,6971	0,7167
43	0,7114	0,7043	0,7232	0,7114	0,6928	0,7102
44	0,71	0,7046	0,7509	0,71	0,6867	0,7101
45	0,701	0,7015	0,707	0,701	0,7	0,7203
46	0,7062	0,7072	0,7214	0,7062	0,6936	0,7362
47	0,7218	0,7046	0,6985	0,7218	0,7155	0,702
48	0,6811	0,6911	0,7069	0,6811	0,6995	0,7383
49	0,6903	0,6945	0,7045	0,6903	0,7035	0,7223
50	0,6964	0,7073	0,7204	0,6964	0,6995	0,7062
51	0,7144	0,6979	0,6888	0,7144	0,7137	0,7252

52	0,7033	0,6955	0,7003	0,7033	0,6718	0,6947
53	0,6884	0,6827	0,7281	0,6884	0,7035	0,718
54	0,6994	0,6966	0,6945	0,6994	0,6936	0,7045
55	0,6983	0,6937	0,7108	0,6983	0,699	0,7081
56	0,7114	0,6968	0,7007	0,7114	0,7074	0,6886
57	0,6891	0,6826	0,7008	0,6891	0,6896	0,7101
58	0,7025	0,7054	0,7191	0,7025	0,7212	0,7183
59	0,7164	0,7002	0,6934	0,7164	0,7068	0,7074
60	0,6982	0,6783	0,7036	0,6982	0,6971	0,6873
61	0,7003	0,6805	0,7048	0,7003	0,6944	0,7011
62	0,7161	0,696	0,7049	0,7161	0,6829	0,6933
63	0,6925	0,7038	0,731	0,6925	0,7144	0,6896
64	0,6906	0,6971	0,7019	0,6906	0,6937	0,7218
65	0,7048	0,6932	0,7091	0,7048	0,6968	0,6901
66	0,7021	0,7143	0,7076	0,7021	0,7042	0,7171
67	0,6964	0,6955	0,6947	0,6964	0,7029	0,7016
68	0,6879	0,6755	0,6993	0,6879	0,7047	0,7134
69	0,7162	0,7004	0,7087	0,7162	0,686	0,6982
70	0,7161	0,6916	0,7259	0,7161	0,6802	0,7015
71	0,6991	0,6957	0,7176	0,6991	0,6923	0,7051
72	0,72	0,6822	0,6774	0,72	0,7112	0,7403
73	0,6846	0,6954	0,7156	0,6846	0,6928	0,7149
74	0,6845	0,6797	0,6977	0,6845	0,7023	0,7165
75	0,7309	0,6934	0,6949	0,7309	0,6971	0,6922
76	0,7124	0,6784	0,6999	0,7124	0,6819	0,7138
77	0,716	0,7245	0,692	0,716	0,7178	0,7144
78	0,7019	0,6744	0,7208	0,7019	0,691	0,6922
79	0,7129	0,692	0,6845	0,7129	0,655	0,7158
80	0,6981	0,6843	0,7132	0,6981	0,6878	0,6921
81	0,6901	0,7063	0,7047	0,6901	0,7024	0,6948
82	0,6934	0,6937	0,7096	0,6934	0,6996	0,6901
83	0,7104	0,668	0,7265	0,7104	0,7113	0,6933
84	0,6968	0,7024	0,7141	0,6968	0,6819	0,714
85	0,6846	0,6826	0,7007	0,6846	0,688	0,7216
86	0,7273	0,6975	0,6837	0,7273	0,7086	0,7137
87	0,7195	0,6985	0,7184	0,7195	0,6981	0,7101
88	0,7041	0,7142	0,7086	0,7041	0,7029	0,7186

89	0,6577	0,7015	0,7105	0,6577	0,6711	0,7156
90	0,7181	0,7108	0,7087	0,7181	0,7007	0,7248
91	0,7042	0,6826	0,7061	0,7042	0,6932	0,7102
92	0,7016	0,6963	0,7295	0,7016	0,6997	0,6971
93	0,7109	0,6951	0,7111	0,7109	0,6996	0,7064
94	0,6959	0,7001	0,7086	0,6959	0,7025	0,7064
95	0,7064	0,7046	0,6968	0,7064	0,6764	0,7156
96	0,7042	0,6879	0,6817	0,7042	0,6793	0,7149
97	0,6896	0,7075	0,7309	0,6896	0,7042	0,6971
98	0,7215	0,7085	0,6909	0,7215	0,6978	0,6952
99	0,7085	0,6802	0,7097	0,7085	0,6832	0,7278
100	0,6976	0,6776	0,701	0,6976	0,7	0,6923
MÉDIA	0,7023	0,6954	0,7070	0,7023	0,6963	0,7086
DESVIO PADRÃO	0,0128	0,0120	0,0131	0,0128	0,0125	0,0142
VARIÂNCIA	0,0002	0,0001	0,0002	0,0002	0,0002	0,0002
INT. CONFIANÇA	0,0025	0,0023	0,0026	0,0025	0,0024	0,0028

Tabela 37 – Resultados das 100 execuções do IG-CEE com o *dataset* D2 nas configurações 3 e 4.

Sementes	Conf. 3			Conf. 4		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,6898	0,6929	0,7279	0,6898	0,6789	0,7168
2	0,6748	0,7047	0,6837	0,6748	0,7124	0,7226
3	0,6971	0,6776	0,7113	0,6971	0,6804	0,6837
4	0,6855	0,7093	0,7038	0,6855	0,6899	0,7163
5	0,6948	0,6722	0,7299	0,6948	0,707	0,7007
6	0,72	0,6865	0,7031	0,72	0,6782	0,6922
7	0,7053	0,6939	0,7054	0,7053	0,7013	0,7164
8	0,6939	0,6896	0,7236	0,6939	0,6969	0,7034
9	0,7027	0,6762	0,6815	0,7027	0,7014	0,7172
10	0,6883	0,6975	0,7219	0,6883	0,6936	0,7152
11	0,6981	0,695	0,7105	0,6981	0,6867	0,7
12	0,7064	0,6823	0,7064	0,7064	0,6768	0,7012
13	0,6805	0,707	0,7093	0,6805	0,7044	0,7181
14	0,6878	0,6968	0,7069	0,6878	0,6823	0,692

15	0,7269	0,695	0,7247	0,7269	0,6787	0,7185
16	0,6935	0,7013	0,697	0,6935	0,6952	0,7029
17	0,7016	0,6725	0,7205	0,7016	0,7255	0,7057
18	0,7061	0,6842	0,7128	0,7061	0,7019	0,7065
19	0,7037	0,6763	0,7201	0,7037	0,6719	0,7104
20	0,6989	0,6886	0,6972	0,6989	0,6717	0,7036
21	0,7006	0,68	0,7054	0,7006	0,6863	0,7183
22	0,6899	0,7005	0,7138	0,6899	0,6979	0,7149
23	0,6955	0,6913	0,7233	0,6955	0,7162	0,7169
24	0,6943	0,6984	0,7109	0,6943	0,7189	0,7038
25	0,7138	0,6953	0,7217	0,7138	0,6863	0,7033
26	0,7147	0,693	0,7053	0,7147	0,6891	0,6993
27	0,7067	0,7084	0,6935	0,7067	0,6988	0,7068
28	0,7166	0,6991	0,7011	0,7166	0,7022	0,7195
29	0,681	0,7087	0,7	0,681	0,6878	0,7206
30	0,6888	0,7033	0,685	0,6888	0,7188	0,7272
31	0,7174	0,7187	0,7282	0,7174	0,7204	0,695
32	0,6969	0,6886	0,7314	0,6969	0,6968	0,6952
33	0,7031	0,6833	0,7052	0,7031	0,6945	0,6868
34	0,7064	0,6751	0,719	0,7064	0,7032	0,6946
35	0,7155	0,6903	0,6955	0,7155	0,7082	0,7154
36	0,705	0,7194	0,707	0,705	0,6723	0,7074
37	0,7329	0,7097	0,7026	0,7329	0,698	0,7064
38	0,7115	0,6975	0,717	0,7115	0,6857	0,6973
39	0,6952	0,7132	0,7305	0,6952	0,689	0,7081
40	0,7119	0,705	0,7229	0,7119	0,7121	0,6949
41	0,7034	0,7124	0,6983	0,7034	0,6825	0,6942
42	0,7045	0,6891	0,7066	0,7045	0,6878	0,7006
43	0,7114	0,6861	0,6999	0,7114	0,6854	0,6959
44	0,71	0,6851	0,7126	0,71	0,7085	0,7255
45	0,701	0,6962	0,7115	0,701	0,6987	0,6852
46	0,7062	0,6983	0,7073	0,7062	0,6942	0,7146
47	0,7218	0,7009	0,7083	0,7218	0,7162	0,6961
48	0,6811	0,6937	0,6818	0,6811	0,6939	0,7222
49	0,6903	0,6839	0,7078	0,6903	0,6961	0,7066
50	0,6964	0,6794	0,6988	0,6964	0,7005	0,722
51	0,7144	0,7033	0,6993	0,7144	0,6908	0,716

52	0,7033	0,7031	0,6892	0,7033	0,6923	0,7055
53	0,6884	0,6919	0,7134	0,6884	0,687	0,7311
54	0,6994	0,6934	0,6993	0,6994	0,6978	0,6893
55	0,6983	0,7025	0,7017	0,6983	0,6962	0,718
56	0,7114	0,7082	0,7045	0,7114	0,6917	0,7038
57	0,6891	0,6673	0,7014	0,6891	0,693	0,6814
58	0,7025	0,6989	0,7022	0,7025	0,7091	0,7091
59	0,7164	0,7129	0,717	0,7164	0,6804	0,7087
60	0,6982	0,6934	0,7093	0,6982	0,6909	0,7019
61	0,7003	0,6931	0,7001	0,7003	0,7076	0,7189
62	0,7161	0,6867	0,6874	0,7161	0,6903	0,6927
63	0,6925	0,7201	0,7023	0,6925	0,6802	0,7072
64	0,6906	0,6667	0,7177	0,6906	0,6849	0,7037
65	0,7048	0,6909	0,731	0,7048	0,6729	0,6794
66	0,7021	0,7112	0,7025	0,7021	0,685	0,6819
67	0,6964	0,7009	0,711	0,6964	0,6871	0,7003
68	0,6879	0,7094	0,7035	0,6879	0,6957	0,7297
69	0,7162	0,6782	0,7117	0,7162	0,7021	0,726
70	0,7161	0,6812	0,7057	0,7161	0,7077	0,7195
71	0,6991	0,6921	0,7063	0,6991	0,6913	0,6842
72	0,72	0,6968	0,7173	0,72	0,6607	0,7131
73	0,6846	0,6928	0,7026	0,6846	0,7033	0,7094
74	0,6845	0,6839	0,7139	0,6845	0,6693	0,7304
75	0,7309	0,709	0,7254	0,7309	0,6879	0,7223
76	0,7124	0,6815	0,7154	0,7124	0,7012	0,6945
77	0,716	0,6841	0,7079	0,716	0,7018	0,7168
78	0,7019	0,7227	0,7139	0,7019	0,6905	0,717
79	0,7129	0,6854	0,7289	0,7129	0,7062	0,7149
80	0,6981	0,6988	0,7043	0,6981	0,6616	0,7064
81	0,6901	0,666	0,6983	0,6901	0,6841	0,7156
82	0,6934	0,6717	0,6965	0,6934	0,7055	0,7321
83	0,7104	0,6836	0,7159	0,7104	0,6792	0,7115
84	0,6968	0,7151	0,6998	0,6968	0,7247	0,6988
85	0,6846	0,6997	0,7068	0,6846	0,6965	0,7226
86	0,7273	0,7237	0,7085	0,7273	0,7101	0,7098
87	0,7195	0,7039	0,707	0,7195	0,7076	0,7192
88	0,7041	0,6914	0,7145	0,7041	0,6843	0,7203

89	0,6577	0,704	0,7054	0,6577	0,706	0,6738
90	0,7181	0,7327	0,6965	0,7181	0,6748	0,7054
91	0,7042	0,6988	0,7242	0,7042	0,7181	0,7107
92	0,7016	0,7051	0,6908	0,7016	0,6865	0,722
93	0,7109	0,6538	0,6979	0,7109	0,668	0,6909
94	0,6959	0,6985	0,714	0,6959	0,6933	0,7084
95	0,7064	0,7024	0,717	0,7064	0,722	0,6994
96	0,7042	0,7074	0,6719	0,7042	0,6747	0,7073
97	0,6896	0,6787	0,6927	0,6896	0,7163	0,7163
98	0,7215	0,6873	0,712	0,7215	0,7088	0,707
99	0,7085	0,7028	0,7093	0,7085	0,7047	0,7002
100	0,6976	0,7031	0,7058	0,6976	0,6746	0,7004
MÉDIA	0,7023	0,6949	0,7078	0,7023	0,6944	0,7074
DESVIO PADRÃO	0,0128	0,0140	0,0118	0,0128	0,0143	0,0126
VARIÂNCIA	0,0002	0,0002	0,0001	0,0002	0,0002	0,0002
INT. CONFIANÇA	0,0025	0,0027	0,0023	0,0025	0,0028	0,0025

Tabela 38 – Resultados das 100 execuções do IG-CEE com o *dataset* D2 nas configurações 5 e 6.

Sementes	Conf. 5			Conf. 6		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,6898	0,6903	0,6872	0,6898	0,6852	0,7027
2	0,6748	0,7039	0,6999	0,6748	0,7108	0,7095
3	0,6971	0,6998	0,706	0,6971	0,7141	0,7222
4	0,6855	0,713	0,6937	0,6855	0,6983	0,7152
5	0,6948	0,6751	0,7201	0,6948	0,703	0,6959
6	0,72	0,6783	0,6981	0,72	0,6914	0,7192
7	0,7053	0,6935	0,7038	0,7053	0,6813	0,7072
8	0,6939	0,6936	0,6987	0,6939	0,6991	0,7014
9	0,7027	0,6886	0,7068	0,7027	0,6926	0,6947
10	0,6883	0,6911	0,6826	0,6883	0,7207	0,7147
11	0,6981	0,6728	0,73	0,6981	0,6957	0,7082
12	0,7064	0,6909	0,714	0,7064	0,6882	0,7194
13	0,6805	0,7144	0,7253	0,6805	0,7006	0,6973
14	0,6878	0,7018	0,6933	0,6878	0,6863	0,7091

15	0,7269	0,6855	0,7081	0,7269	0,6788	0,6897
16	0,6935	0,705	0,6862	0,6935	0,6859	0,7095
17	0,7016	0,6767	0,7266	0,7016	0,6809	0,7089
18	0,7061	0,7088	0,7244	0,7061	0,6964	0,6685
19	0,7037	0,7006	0,7161	0,7037	0,701	0,6969
20	0,6989	0,678	0,7109	0,6989	0,7148	0,7059
21	0,7006	0,6774	0,6959	0,7006	0,6752	0,7149
22	0,6899	0,7039	0,7097	0,6899	0,7193	0,706
23	0,6955	0,6827	0,6815	0,6955	0,6897	0,7095
24	0,6943	0,6932	0,7085	0,6943	0,656	0,7245
25	0,7138	0,6993	0,7117	0,7138	0,7109	0,7316
26	0,7147	0,6823	0,7121	0,7147	0,6837	0,6975
27	0,7067	0,7177	0,7102	0,7067	0,7323	0,7058
28	0,7166	0,6921	0,7077	0,7166	0,7029	0,6966
29	0,681	0,6876	0,715	0,681	0,6844	0,7026
30	0,6888	0,7493	0,7159	0,6888	0,7058	0,7162
31	0,7174	0,7116	0,685	0,7174	0,6684	0,6946
32	0,6969	0,7009	0,7089	0,6969	0,6807	0,6948
33	0,7031	0,6928	0,7034	0,7031	0,6898	0,7167
34	0,7064	0,6938	0,7001	0,7064	0,6817	0,7218
35	0,7155	0,6793	0,7404	0,7155	0,7037	0,7065
36	0,705	0,6669	0,6829	0,705	0,6955	0,723
37	0,7329	0,7096	0,7072	0,7329	0,7025	0,7059
38	0,7115	0,7026	0,6917	0,7115	0,6902	0,6645
39	0,6952	0,7121	0,7161	0,6952	0,6979	0,7128
40	0,7119	0,6987	0,6907	0,7119	0,6707	0,7035
41	0,7034	0,6891	0,7081	0,7034	0,7117	0,7108
42	0,7045	0,7002	0,7011	0,7045	0,6865	0,7069
43	0,7114	0,6726	0,7105	0,7114	0,69	0,6926
44	0,71	0,716	0,7227	0,71	0,692	0,6997
45	0,701	0,6962	0,6913	0,701	0,6921	0,6951
46	0,7062	0,6993	0,7187	0,7062	0,6989	0,7083
47	0,7218	0,7207	0,7062	0,7218	0,69	0,6952
48	0,6811	0,7114	0,7195	0,6811	0,6816	0,704
49	0,6903	0,6838	0,7195	0,6903	0,7131	0,7138
50	0,6964	0,6989	0,7195	0,6964	0,66	0,6984
51	0,7144	0,6825	0,6866	0,7144	0,6791	0,6928

52	0,7033	0,713	0,6992	0,7033	0,6891	0,7063
53	0,6884	0,7139	0,7155	0,6884	0,7006	0,6971
54	0,6994	0,7046	0,692	0,6994	0,6991	0,6841
55	0,6983	0,689	0,7426	0,6983	0,6727	0,6999
56	0,7114	0,7051	0,7119	0,7114	0,6918	0,7192
57	0,6891	0,6809	0,7053	0,6891	0,7164	0,7071
58	0,7025	0,7	0,717	0,7025	0,7094	0,7153
59	0,7164	0,6952	0,7088	0,7164	0,6847	0,703
60	0,6982	0,6992	0,7113	0,6982	0,6799	0,7025
61	0,7003	0,701	0,6942	0,7003	0,6662	0,6891
62	0,7161	0,6918	0,7161	0,7161	0,6656	0,7063
63	0,6925	0,7049	0,7061	0,6925	0,6798	0,692
64	0,6906	0,6927	0,6977	0,6906	0,6768	0,7087
65	0,7048	0,6987	0,7153	0,7048	0,7074	0,7125
66	0,7021	0,6924	0,6881	0,7021	0,696	0,7011
67	0,6964	0,7052	0,7229	0,6964	0,6887	0,6984
68	0,6879	0,7021	0,7038	0,6879	0,68	0,7007
69	0,7162	0,6918	0,7085	0,7162	0,706	0,6961
70	0,7161	0,7065	0,6984	0,7161	0,6837	0,7018
71	0,6991	0,7021	0,7476	0,6991	0,6967	0,7033
72	0,72	0,6872	0,7093	0,72	0,7098	0,7145
73	0,6846	0,6886	0,718	0,6846	0,6994	0,6985
74	0,6845	0,677	0,7139	0,6845	0,6824	0,6894
75	0,7309	0,7011	0,7037	0,7309	0,6968	0,7028
76	0,7124	0,699	0,7057	0,7124	0,68	0,7169
77	0,716	0,6938	0,7016	0,716	0,7176	0,6986
78	0,7019	0,6839	0,6875	0,7019	0,6974	0,6893
79	0,7129	0,7158	0,714	0,7129	0,7078	0,6732
80	0,6981	0,702	0,712	0,6981	0,6737	0,7003
81	0,6901	0,6961	0,7021	0,6901	0,7128	0,6957
82	0,6934	0,6927	0,7323	0,6934	0,705	0,7176
83	0,7104	0,7026	0,7119	0,7104	0,703	0,6949
84	0,6968	0,6819	0,6862	0,6968	0,6915	0,7174
85	0,6846	0,6903	0,6995	0,6846	0,7016	0,7257
86	0,7273	0,7157	0,7009	0,7273	0,6936	0,7051
87	0,7195	0,6973	0,7046	0,7195	0,6994	0,7063
88	0,7041	0,709	0,6972	0,7041	0,6764	0,7059

89	0,6577	0,7229	0,7213	0,6577	0,7037	0,7053
90	0,7181	0,6772	0,7232	0,7181	0,6958	0,7107
91	0,7042	0,696	0,7184	0,7042	0,6962	0,7178
92	0,7016	0,697	0,7048	0,7016	0,6832	0,7197
93	0,7109	0,7032	0,7089	0,7109	0,6912	0,6895
94	0,6959	0,6803	0,6982	0,6959	0,7291	0,6983
95	0,7064	0,69	0,7107	0,7064	0,6925	0,7098
96	0,7042	0,6962	0,6931	0,7042	0,7124	0,7347
97	0,6896	0,7075	0,7206	0,6896	0,6741	0,689
98	0,7215	0,6737	0,718	0,7215	0,6934	0,7092
99	0,7085	0,7101	0,7227	0,7085	0,6902	0,72
100	0,6976	0,6946	0,7103	0,6976	0,6958	0,6716
MÉDIA	0,7023	0,6966	0,7078	0,7023	0,6936	0,7044
DESVIO PADRÃO	0,0128	0,0132	0,0132	0,0128	0,0145	0,0124
VARIÂNCIA	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
INT. CONFIANÇA	0,0025	0,0026	0,0026	0,0025	0,0028	0,0024

Tabela 39 – Resultados das 100 execuções do IG-CEE com o *dataset* D3 nas configurações 1 e 2.

Sementes	Conf1.			Conf 2.		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,635	0,6419	0,6276	0,635	0,6375	0,6205
2	0,625	0,6095	0,6305	0,625	0,6306	0,659
3	0,6359	0,6344	0,652	0,6359	0,6422	0,6413
4	0,6238	0,631	0,6321	0,6238	0,6347	0,6394
5	0,618	0,6201	0,6459	0,618	0,6148	0,6155
6	0,6213	0,6257	0,6561	0,6213	0,6272	0,635
7	0,6311	0,626	0,6353	0,6311	0,6169	0,6319
8	0,6306	0,6402	0,6316	0,6306	0,6384	0,6399
9	0,6268	0,6056	0,6387	0,6268	0,6281	0,6161
10	0,6302	0,6268	0,6095	0,6302	0,6365	0,6241
11	0,6176	0,6348	0,6469	0,6176	0,638	0,649
12	0,6308	0,6325	0,631	0,6308	0,6249	0,6275
13	0,6258	0,6405	0,6232	0,6258	0,6261	0,6475
14	0,6248	0,6126	0,6244	0,6248	0,6281	0,6307

15	0,6431	0,6553	0,6453	0,6431	0,604	0,6226
16	0,6367	0,6359	0,6275	0,6367	0,6548	0,6198
17	0,6201	0,6109	0,6319	0,6201	0,6451	0,6389
18	0,6358	0,6175	0,6344	0,6358	0,6422	0,6383
19	0,6383	0,6463	0,6204	0,6383	0,6289	0,644
20	0,615	0,6202	0,6502	0,615	0,6245	0,6381
21	0,6419	0,6253	0,6198	0,6419	0,6306	0,6317
22	0,6256	0,6386	0,6174	0,6256	0,6033	0,6243
23	0,6307	0,6231	0,623	0,6307	0,6322	0,6418
24	0,6296	0,6284	0,6249	0,6296	0,6238	0,6168
25	0,6213	0,6412	0,6243	0,6213	0,6348	0,6189
26	0,6451	0,6085	0,6249	0,6451	0,6409	0,6168
27	0,6228	0,6289	0,6329	0,6228	0,6269	0,6278
28	0,632	0,6194	0,6389	0,632	0,6205	0,6106
29	0,6196	0,6314	0,6117	0,6196	0,6158	0,6385
30	0,6287	0,6406	0,6395	0,6287	0,6276	0,6369
31	0,6146	0,6396	0,619	0,6146	0,6391	0,6387
32	0,6214	0,6445	0,6099	0,6214	0,6219	0,6339
33	0,6434	0,6165	0,6158	0,6434	0,6262	0,6202
34	0,6466	0,6343	0,6253	0,6466	0,612	0,6299
35	0,6125	0,6134	0,6006	0,6125	0,6245	0,6085
36	0,6389	0,6183	0,6354	0,6389	0,6154	0,6308
37	0,6154	0,6273	0,6198	0,6154	0,63	0,6203
38	0,6284	0,6423	0,6413	0,6284	0,6433	0,608
39	0,6327	0,6209	0,6248	0,6327	0,6211	0,6295
40	0,6376	0,6338	0,6248	0,6376	0,6495	0,6123
41	0,6302	0,6147	0,6191	0,6302	0,6303	0,6272
42	0,6281	0,6307	0,6516	0,6281	0,6376	0,6342
43	0,6254	0,6221	0,6295	0,6254	0,633	0,6557
44	0,6296	0,6349	0,6396	0,6296	0,6207	0,6231
45	0,6268	0,6357	0,6558	0,6268	0,625	0,6356
46	0,637	0,637	0,6239	0,637	0,6369	0,6242
47	0,612	0,6526	0,6185	0,612	0,6476	0,6331
48	0,6198	0,623	0,6284	0,6198	0,625	0,6297
49	0,6148	0,607	0,6297	0,6148	0,6302	0,6492
50	0,6249	0,6059	0,6415	0,6249	0,6224	0,614
51	0,6445	0,6167	0,6179	0,6445	0,626	0,618

52	0,6453	0,6414	0,6471	0,6453	0,6214	0,6177
53	0,6354	0,6262	0,6341	0,6354	0,6452	0,6217
54	0,6262	0,626	0,6323	0,6262	0,6305	0,6289
55	0,6297	0,6401	0,6258	0,6297	0,6363	0,6297
56	0,6329	0,6146	0,6229	0,6329	0,6397	0,6347
57	0,6354	0,627	0,6325	0,6354	0,6259	0,6245
58	0,6286	0,6183	0,6178	0,6286	0,6036	0,6256
59	0,625	0,618	0,6251	0,625	0,6365	0,6501
60	0,6331	0,6199	0,6269	0,6331	0,6183	0,6225
61	0,6232	0,6084	0,6225	0,6232	0,6243	0,6277
62	0,6348	0,6265	0,6429	0,6348	0,6302	0,6266
63	0,6362	0,66	0,6287	0,6362	0,6311	0,6202
64	0,6155	0,6358	0,6123	0,6155	0,609	0,6163
65	0,6152	0,6419	0,6276	0,6152	0,6207	0,6386
66	0,6306	0,6279	0,635	0,6306	0,621	0,6046
67	0,6296	0,6233	0,6296	0,6296	0,6498	0,6167
68	0,6361	0,6115	0,6364	0,6361	0,6394	0,6443
69	0,6114	0,623	0,6427	0,6114	0,6098	0,6299
70	0,6272	0,6372	0,6192	0,6272	0,6166	0,6489
71	0,6441	0,6133	0,6509	0,6441	0,6166	0,6282
72	0,6321	0,6314	0,6225	0,6321	0,6286	0,6267
73	0,6327	0,6294	0,6406	0,6327	0,6331	0,6409
74	0,6256	0,626	0,6264	0,6256	0,6282	0,6465
75	0,6304	0,6228	0,6299	0,6304	0,6221	0,6239
76	0,6264	0,6271	0,6282	0,6264	0,6381	0,6308
77	0,6221	0,6229	0,6382	0,6221	0,6204	0,6263
78	0,6102	0,6305	0,6263	0,6102	0,6178	0,6362
79	0,6297	0,6374	0,6125	0,6297	0,6372	0,6402
80	0,6211	0,6429	0,6137	0,6211	0,622	0,6223
81	0,6215	0,6154	0,6199	0,6215	0,631	0,6236
82	0,6285	0,6324	0,6292	0,6285	0,6132	0,6462
83	0,6333	0,6388	0,6277	0,6333	0,6377	0,6192
84	0,6293	0,6313	0,6316	0,6293	0,6352	0,6306
85	0,6471	0,6244	0,6237	0,6471	0,6358	0,6419
86	0,6189	0,6234	0,6224	0,6189	0,6245	0,6177
87	0,6187	0,6377	0,6211	0,6187	0,6284	0,6109
88	0,6263	0,6448	0,6178	0,6263	0,6294	0,6334

89	0,6217	0,6202	0,6495	0,6217	0,6236	0,6302
90	0,6359	0,6248	0,6325	0,6359	0,6313	0,6248
91	0,6216	0,6337	0,6453	0,6216	0,6324	0,6167
92	0,6244	0,614	0,633	0,6244	0,6213	0,6259
93	0,6287	0,6508	0,6268	0,6287	0,6399	0,6202
94	0,6091	0,6333	0,6481	0,6091	0,64	0,625
95	0,6299	0,6314	0,6363	0,6299	0,6377	0,653
96	0,6172	0,6367	0,6247	0,6172	0,619	0,6206
97	0,6101	0,6352	0,6418	0,6101	0,624	0,6321
98	0,6162	0,6295	0,638	0,6162	0,637	0,6063
99	0,6232	0,6234	0,6445	0,6232	0,6267	0,6176
100	0,6349	0,6152	0,6207	0,6349	0,6114	0,6278
MÉDIA	0,6277	0,6283	0,6301	0,6277	0,6285	0,6289
DESVIO PADRÃO	0,0089	0,0114	0,0112	0,0089	0,0104	0,0115
VARIÂNCIA	8,00E-05	1,30E-04	1,30E-04	8,00E-05	1,10E-04	1,30E-04
INT. CONFIANÇA	0,0017	0,0022	0,0022	0,0017	0,0020	0,0023

Tabela 40 – Resultados das 100 execuções do IG-CEE com o *dataset* D3 nas configurações 3 e 4.

Sementes	Conf. 3			Conf. 4		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,635	0,6459	0,6104	0,635	0,6223	0,6256
2	0,625	0,6533	0,6219	0,625	0,6276	0,6341
3	0,6359	0,6261	0,6429	0,6359	0,6286	0,6269
4	0,6238	0,632	0,628	0,6238	0,6244	0,605
5	0,618	0,6186	0,6084	0,618	0,6274	0,6275
6	0,6213	0,6221	0,6146	0,6213	0,6333	0,6227
7	0,6311	0,628	0,6467	0,6311	0,6374	0,637
8	0,6306	0,6298	0,6231	0,6306	0,6053	0,639
9	0,6268	0,6227	0,6468	0,6268	0,6225	0,6201
10	0,6302	0,628	0,6551	0,6302	0,6369	0,6347
11	0,6176	0,6509	0,6014	0,6176	0,6333	0,6484
12	0,6308	0,639	0,623	0,6308	0,6344	0,6296
13	0,6258	0,6476	0,6416	0,6258	0,6315	0,6146
14	0,6248	0,6231	0,611	0,6248	0,623	0,6156

15	0,6431	0,6378	0,6316	0,6431	0,624	0,6322
16	0,6367	0,6329	0,6477	0,6367	0,6135	0,6165
17	0,6201	0,6344	0,6219	0,6201	0,6355	0,6222
18	0,6358	0,6246	0,6471	0,6358	0,6093	0,6217
19	0,6383	0,6075	0,6064	0,6383	0,6141	0,6256
20	0,615	0,6352	0,6321	0,615	0,6328	0,6336
21	0,6419	0,6407	0,6289	0,6419	0,624	0,6291
22	0,6256	0,6255	0,634	0,6256	0,6271	0,6319
23	0,6307	0,6221	0,6243	0,6307	0,6369	0,6553
24	0,6296	0,6301	0,624	0,6296	0,6292	0,634
25	0,6213	0,6408	0,6274	0,6213	0,6104	0,66
26	0,6451	0,6095	0,6155	0,6451	0,6184	0,6195
27	0,6228	0,6218	0,6115	0,6228	0,6208	0,6316
28	0,632	0,6426	0,6118	0,632	0,6241	0,6333
29	0,6196	0,6369	0,618	0,6196	0,6303	0,6345
30	0,6287	0,6251	0,6254	0,6287	0,6293	0,6289
31	0,6146	0,6052	0,63	0,6146	0,6264	0,6175
32	0,6214	0,6341	0,6009	0,6214	0,6244	0,6259
33	0,6434	0,6334	0,6299	0,6434	0,6159	0,6403
34	0,6466	0,6201	0,624	0,6466	0,6307	0,6305
35	0,6125	0,6429	0,6225	0,6125	0,6381	0,6387
36	0,6389	0,624	0,6415	0,6389	0,6408	0,6213
37	0,6154	0,6232	0,6445	0,6154	0,6456	0,6484
38	0,6284	0,6286	0,6289	0,6284	0,6209	0,6177
39	0,6327	0,6449	0,6071	0,6327	0,6269	0,6416
40	0,6376	0,6371	0,6289	0,6376	0,6275	0,6599
41	0,6302	0,6163	0,629	0,6302	0,6195	0,6305
42	0,6281	0,6385	0,6446	0,6281	0,6484	0,6228
43	0,6254	0,6207	0,6488	0,6254	0,6237	0,6394
44	0,6296	0,6201	0,6397	0,6296	0,6309	0,6251
45	0,6268	0,6312	0,6192	0,6268	0,6186	0,6335
46	0,637	0,5986	0,63	0,637	0,6175	0,623
47	0,612	0,6374	0,6258	0,612	0,6297	0,6233
48	0,6198	0,6389	0,6322	0,6198	0,6431	0,6302
49	0,6148	0,6186	0,6264	0,6148	0,6268	0,6265
50	0,6249	0,6207	0,6238	0,6249	0,6204	0,6298
51	0,6445	0,64	0,61	0,6445	0,6209	0,6401

52	0,6453	0,6366	0,6185	0,6453	0,607	0,6445
53	0,6354	0,6417	0,6369	0,6354	0,6111	0,6344
54	0,6262	0,6474	0,6338	0,6262	0,6325	0,6436
55	0,6297	0,627	0,6045	0,6297	0,617	0,6304
56	0,6329	0,6265	0,6701	0,6329	0,6202	0,6418
57	0,6354	0,6346	0,6337	0,6354	0,6216	0,6269
58	0,6286	0,645	0,6421	0,6286	0,6297	0,6379
59	0,625	0,6253	0,6453	0,625	0,6248	0,6182
60	0,6331	0,6279	0,6236	0,6331	0,6377	0,6248
61	0,6232	0,607	0,6398	0,6232	0,6141	0,6303
62	0,6348	0,6011	0,6618	0,6348	0,6321	0,6395
63	0,6362	0,6221	0,6476	0,6362	0,6116	0,6327
64	0,6155	0,6176	0,6229	0,6155	0,6422	0,631
65	0,6152	0,6381	0,6356	0,6152	0,6235	0,6442
66	0,6306	0,6291	0,631	0,6306	0,6199	0,6279
67	0,6296	0,6133	0,6249	0,6296	0,6487	0,6314
68	0,6361	0,6428	0,6248	0,6361	0,6198	0,6152
69	0,6114	0,6361	0,6167	0,6114	0,6269	0,6211
70	0,6272	0,6261	0,6399	0,6272	0,646	0,6268
71	0,6441	0,6379	0,6241	0,6441	0,6066	0,6347
72	0,6321	0,6259	0,6325	0,6321	0,6405	0,6467
73	0,6327	0,6264	0,6297	0,6327	0,6347	0,6381
74	0,6256	0,6266	0,6225	0,6256	0,6398	0,6342
75	0,6304	0,6342	0,6326	0,6304	0,6397	0,6593
76	0,6264	0,6117	0,6228	0,6264	0,6271	0,6442
77	0,6221	0,6304	0,6238	0,6221	0,6283	0,6367
78	0,6102	0,6381	0,6299	0,6102	0,6333	0,6166
79	0,6297	0,641	0,6282	0,6297	0,6476	0,6153
80	0,6211	0,6223	0,6301	0,6211	0,6371	0,6198
81	0,6215	0,6276	0,6212	0,6215	0,6191	0,6385
82	0,6285	0,6199	0,6429	0,6285	0,633	0,6215
83	0,6333	0,6197	0,6248	0,6333	0,6157	0,6165
84	0,6293	0,6356	0,6459	0,6293	0,6423	0,625
85	0,6471	0,6412	0,6449	0,6471	0,6369	0,6457
86	0,6189	0,6369	0,6154	0,6189	0,6206	0,6261
87	0,6187	0,6288	0,6267	0,6187	0,6167	0,6254
88	0,6263	0,6207	0,6464	0,6263	0,6156	0,629

89	0,6217	0,6509	0,6219	0,6217	0,6316	0,625
90	0,6359	0,6227	0,6267	0,6359	0,6108	0,6356
91	0,6216	0,6258	0,635	0,6216	0,6262	0,6321
92	0,6244	0,6246	0,6414	0,6244	0,6398	0,6225
93	0,6287	0,6273	0,649	0,6287	0,6332	0,6272
94	0,6091	0,6491	0,6175	0,6091	0,6292	0,6186
95	0,6299	0,6177	0,6273	0,6299	0,6363	0,6286
96	0,6172	0,6293	0,6269	0,6172	0,6381	0,6324
97	0,6101	0,6228	0,6122	0,6101	0,6265	0,6287
98	0,6162	0,6056	0,6194	0,6162	0,6454	0,6398
99	0,6232	0,6441	0,6517	0,6232	0,6034	0,6273
100	0,6349	0,6354	0,6219	0,6349	0,6171	0,6475
MÉDIA	0,6277	0,6293	0,6292	0,6277	0,6272	0,6310
DESVIO PADRÃO	0,0089	0,0114	0,0132	0,0089	0,0104	0,0104
VARIÂNCIA	8,00E-05	1,30E-04	1,70E-04	8,00E-05	1,10E-04	1,10E-04
INT. CONFIANÇA	0,0017	0,0022	0,0026	0,0017	0,0020	0,0020

Tabela 41 – Resultados das 100 execuções do IG-CEE com o *dataset* D3 nas configurações 5 e 6.

Sementes	Conf. 5			Conf. 6		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
1	0,635	0,6082	0,622	0,635	0,6146	0,6176
2	0,625	0,616	0,6199	0,625	0,6344	0,6172
3	0,6359	0,6275	0,6324	0,6359	0,6416	0,6414
4	0,6238	0,6409	0,6335	0,6238	0,6263	0,6339
5	0,618	0,6304	0,6339	0,618	0,6205	0,6445
6	0,6213	0,6398	0,6312	0,6213	0,6338	0,6169
7	0,6311	0,6376	0,6349	0,6311	0,635	0,6301
8	0,6306	0,6315	0,6457	0,6306	0,6183	0,6296
9	0,6268	0,6087	0,6156	0,6268	0,6372	0,6175
10	0,6302	0,6011	0,6411	0,6302	0,6487	0,6127
11	0,6176	0,6174	0,6291	0,6176	0,635	0,637
12	0,6308	0,6349	0,6198	0,6308	0,6249	0,6251
13	0,6258	0,6119	0,6115	0,6258	0,6383	0,6281
14	0,6248	0,6238	0,6347	0,6248	0,6357	0,6413

15	0,6431	0,6236	0,6385	0,6431	0,6113	0,6209
16	0,6367	0,6383	0,6239	0,6367	0,6413	0,6181
17	0,6201	0,6359	0,6416	0,6201	0,6234	0,6258
18	0,6358	0,6444	0,6353	0,6358	0,6221	0,647
19	0,6383	0,6372	0,6442	0,6383	0,6254	0,6315
20	0,615	0,6126	0,6281	0,615	0,6298	0,6394
21	0,6419	0,6405	0,6408	0,6419	0,6331	0,6272
22	0,6256	0,6208	0,6308	0,6256	0,634	0,627
23	0,6307	0,6494	0,6346	0,6307	0,6285	0,6508
24	0,6296	0,6295	0,6393	0,6296	0,6145	0,6358
25	0,6213	0,6394	0,6303	0,6213	0,6194	0,6199
26	0,6451	0,631	0,6224	0,6451	0,6343	0,6235
27	0,6228	0,617	0,6446	0,6228	0,6276	0,6357
28	0,632	0,632	0,6144	0,632	0,6275	0,6329
29	0,6196	0,6228	0,6231	0,6196	0,6113	0,6373
30	0,6287	0,6242	0,6302	0,6287	0,6262	0,6279
31	0,6146	0,6061	0,6373	0,6146	0,6184	0,6385
32	0,6214	0,6335	0,6305	0,6214	0,6098	0,6161
33	0,6434	0,6335	0,6447	0,6434	0,6322	0,6203
34	0,6466	0,6413	0,622	0,6466	0,6406	0,62
35	0,6125	0,6241	0,6188	0,6125	0,6145	0,6443
36	0,6389	0,6345	0,6238	0,6389	0,6498	0,6319
37	0,6154	0,6362	0,6211	0,6154	0,616	0,6245
38	0,6284	0,6274	0,6319	0,6284	0,6286	0,6149
39	0,6327	0,6203	0,6401	0,6327	0,6345	0,6447
40	0,6376	0,6192	0,6331	0,6376	0,6317	0,6357
41	0,6302	0,6273	0,6345	0,6302	0,6151	0,6206
42	0,6281	0,6188	0,636	0,6281	0,641	0,6144
43	0,6254	0,6259	0,6155	0,6254	0,6252	0,6241
44	0,6296	0,6085	0,6277	0,6296	0,6281	0,6302
45	0,6268	0,6325	0,6124	0,6268	0,6341	0,6425
46	0,637	0,6349	0,6313	0,637	0,6309	0,629
47	0,612	0,6162	0,6348	0,612	0,6359	0,6524
48	0,6198	0,6145	0,6296	0,6198	0,6253	0,6371
49	0,6148	0,6177	0,623	0,6148	0,6218	0,6449
50	0,6249	0,6267	0,637	0,6249	0,6333	0,6255
51	0,6445	0,6283	0,6201	0,6445	0,6539	0,626

52	0,6453	0,631	0,6221	0,6453	0,6081	0,6314
53	0,6354	0,6159	0,6297	0,6354	0,6224	0,6243
54	0,6262	0,6339	0,6262	0,6262	0,6213	0,622
55	0,6297	0,6202	0,6309	0,6297	0,6338	0,6267
56	0,6329	0,6283	0,6263	0,6329	0,6221	0,6448
57	0,6354	0,639	0,6151	0,6354	0,6426	0,6207
58	0,6286	0,6286	0,6322	0,6286	0,6252	0,6372
59	0,625	0,6165	0,6269	0,625	0,6509	0,6109
60	0,6331	0,6365	0,6397	0,6331	0,6427	0,6411
61	0,6232	0,6215	0,616	0,6232	0,6153	0,643
62	0,6348	0,6352	0,6286	0,6348	0,623	0,6259
63	0,6362	0,61	0,6034	0,6362	0,6218	0,6203
64	0,6155	0,6258	0,6444	0,6155	0,6243	0,6252
65	0,6152	0,6516	0,635	0,6152	0,6212	0,6302
66	0,6306	0,6358	0,6295	0,6306	0,6265	0,6245
67	0,6296	0,6299	0,6328	0,6296	0,6316	0,622
68	0,6361	0,6254	0,6359	0,6361	0,6206	0,6408
69	0,6114	0,617	0,6407	0,6114	0,6486	0,617
70	0,6272	0,6172	0,6378	0,6272	0,6389	0,6343
71	0,6441	0,6421	0,6543	0,6441	0,632	0,6234
72	0,6321	0,6235	0,6304	0,6321	0,6153	0,6337
73	0,6327	0,6302	0,636	0,6327	0,6335	0,6335
74	0,6256	0,6229	0,6216	0,6256	0,6391	0,6337
75	0,6304	0,6432	0,6311	0,6304	0,6115	0,6114
76	0,6264	0,6436	0,6253	0,6264	0,6425	0,6043
77	0,6221	0,6222	0,6285	0,6221	0,6106	0,6531
78	0,6102	0,6399	0,6316	0,6102	0,6182	0,6348
79	0,6297	0,6391	0,6252	0,6297	0,6196	0,6373
80	0,6211	0,6294	0,6232	0,6211	0,6318	0,6307
81	0,6215	0,6433	0,6405	0,6215	0,6321	0,6292
82	0,6285	0,6203	0,6125	0,6285	0,6238	0,608
83	0,6333	0,643	0,6434	0,6333	0,6325	0,6249
84	0,6293	0,6189	0,6228	0,6293	0,631	0,635
85	0,6471	0,6379	0,6269	0,6471	0,6185	0,6254
86	0,6189	0,617	0,6361	0,6189	0,6506	0,642
87	0,6187	0,6181	0,6482	0,6187	0,6257	0,6663
88	0,6263	0,6324	0,6331	0,6263	0,6349	0,6251

89	0,6217	0,6304	0,6283	0,6217	0,636	0,6243
90	0,6359	0,6261	0,6142	0,6359	0,6285	0,6227
91	0,6216	0,6081	0,625	0,6216	0,6277	0,6197
92	0,6244	0,6108	0,6181	0,6244	0,6206	0,63
93	0,6287	0,6125	0,6288	0,6287	0,6107	0,608
94	0,6091	0,6205	0,6394	0,6091	0,636	0,6389
95	0,6299	0,6333	0,5948	0,6299	0,6098	0,6278
96	0,6346	0,6263	0,6212	0,6172	0,6278	0,6388
97	0,6363	0,6207	0,6272	0,6213	0,6202	0,6378
98	0,6205	0,6314	0,6211	0,6295	0,6315	0,633
99	0,6246	0,6409	0,625	0,6124	0,6212	0,6213
100	0,6247	0,6211	0,6278	0,6326	0,6322	0,62
MÉDIA	0,6281	0,6272	0,6293	0,6278	0,6282	0,6295
DESVIO PADRÃO	0,0087	0,0107	0,0098	0,0088	0,0103	0,0109
VARIÂNCIA	7,50E-05	1,10E-04	9,60E-05	7,70E-05	1,10E-04	1,20E-04
INT. CONFIANÇA	0,0017	0,0021	0,0019	0,0017	0,0020	0,0021