

UNIVERSIDADE FEDERAL FLUMINENSE  
INSTITUTO DE COMPUTAÇÃO  
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

ELZENCLEVER FREITAS DE AGUIAR

DETERMINAÇÃO DA INFLUÊNCIA DA REATIVIDADE DOS GRUPOS SANGÜÍNEOS  
ABO E RH NA ETNIA SIMPLIFICADA ATRAVÉS DA UTILIZAÇÃO DE REDES  
NEURAS ARTIFICIAIS

NITERÓI

2004

ELZENCLEVER FREITAS DE AGUIAR

DETERMINAÇÃO DA INFLUÊNCIA DA REATIVIDADE DOS GRUPOS SANGÜÍNEOS ABO E RH NA ETNIA SIMPLIFICADA ATRAVÉS DA UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS

Dissertação apresentada ao Curso de Pós-Graduação em Computação da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Mestre em Computação.  
Área de Concentração: Inteligência Artificial.

Orientador: Prof<sup>o</sup> Dr<sup>o</sup> LUIZ SATORU OCHI

Niterói

2004

ELZENCLEVER FREITAS DE AGUIAR

DETERMINAÇÃO DA INFLUÊNCIA DA REATIVIDADE DOS GRUPOS SANGÜÍNEOS  
ABO E RH NA ETNIA SIMPLIFICADA ATRAVÉS DA UTILIZAÇÃO DE REDES  
NEURAS ARTIFICIAIS

Dissertação apresentada ao Curso de Pós-Graduação em Computação da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Mestre em Computação.  
Área de Concentração: Inteligência Artificial.

Aprovada em março de 2004

BANCA EXAMINADORA

---

Prof<sup>o</sup>. Dr. Luiz Satoru Ochi – UFF (Presidente)

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Priscila Machado V. Lima – UFRJ

---

Prof<sup>o</sup>. Dr. Felipe Maia Galvão França – UFRJ

---

Prof<sup>o</sup>. Dr. Ronaldo Curi Gismondi - UERJ

Niterói

2004

Aos meus pais, minha esposa e a minha família pelas constantes demonstrações de carinho, compreensão e amor.

## AGRADECIMENTOS

A Luiz Satoru Ochi - meu orientador, pela confiança e apoio.

A Priscila Machado V. Lima – pelas orientações, críticas e apoio.

A Ronaldo C. Gismondi – pelo apoio e confiança.

A Claudia Santos – pelo desafio de provar novas idéias.

A Rafael de O. V. dos Santos – pela ajuda inestimável.

Aos professores da UFF e colegas de estudo – pelo apoio e reflexões críticas.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>18</b>
1.1	MOTIVAÇÕES .....	18
1.2	OBJETIVOS E DESCRIÇÃO DO TRABALHO .....	21
1.3	ORGANIZAÇÃO DOS CAPÍTULOS .....	22
<b>2</b>	<b>CONCEITOS EM BIOLOGIA E ANTROPOLOGIA .....</b>	<b>23</b>
2.1	GENÉTICA, TIPAGEM SANGUÍNEA E COR .....	23
2.1.1	<i>Alguns Conceitos Iniciais em Genética</i> .....	24
2.1.2	<i>Sistema A B O</i> .....	25
2.1.3	<i>Aspectos da Herança do Sistema ABO</i> .....	26
2.1.4	<i>Interação Gênica</i> .....	26
2.1.4.1	<i>Epistasia</i> .....	27
2.2	COR .....	28
2.2.1	<i>História</i> .....	28
2.2.2	<i>Cor e subjetividade</i> .....	31
2.2.2.1	Quantas denominações de cor?.....	33
2.2.2.2	Pergunta aberta por pergunta fechada.....	34
2.2.2.3	Regionalização da Cor.....	35
2.3	COMENTÁRIOS DO CAPÍTULO .....	36
<b>3</b>	<b>CONCEITOS EM REDES NEURAIS ARTIFICIAIS (RNA) .....</b>	<b>38</b>
3.1	APRESENTAÇÃO.....	38
3.2	NEURÔNIO: BIOLÓGICO E ARTIFICIAL.....	38
3.3	REDES NEURAIS – COMPONENTES E TIPOS.....	42
3.3.1	<i>Padrões, funções, conexões e elementos de processamento</i> .....	42
3.3.2	<i>Redes Neurais - Classificação</i> .....	44
3.3.3	<i>Redes Neurais - Estrutura</i> .....	46
3.3.4	<i>Redes Neurais – Parâmetros</i> .....	47
3.4	REDES NEURAIS – O QUE REPRESENTAM .....	48
3.5	REDES NEURAIS – ALGORITMO DE APRENDIZADO .....	49
3.6	REDES NEURAIS – MODELOS .....	51
3.6.1	<i>Perceptron / Adaline</i> .....	51
3.6.2	<i>Backpropagation</i> .....	51
3.7	COMITÊS DE REDES NEURAIS .....	52
3.7.1	<i>Níveis de classificação</i> .....	53
3.7.2	<i>Métodos de combinação</i> .....	53
3.7.3	<i>Formação de comitês</i> .....	54
3.7.4	<i>Comentários do capítulo</i> .....	56
<b>4</b>	<b>DESCRIÇÃO DOS EXPERIMENTOS.....</b>	<b>58</b>
4.1	INTRODUÇÃO .....	58
4.2	EXPERIMENTOS.....	62
4.2.1	<i>Experimento A: Validação da Rede pela Tipagem Sangüínea</i> .....	62
4.2.1.1	Experimento A.1 .....	63
4.2.1.2	Experimento A.2 .....	65
4.2.1.3	Experimento A.3 .....	66
4.2.1.4	Experimento A.4 .....	68
4.2.1.5	Experimento A.5 .....	69
4.2.1.6	Experimento A.6 .....	71
4.2.2	<i>Experimento B: Determinação da Etnia simplificada</i> .....	73
4.2.2.1	Experimento B.1.....	73
4.2.2.2	Experimento B.2.....	75
4.2.2.3	Experimento B.3.....	76
4.2.2.4	Experimento B.4.....	78
4.2.2.5	Experimento B.5.....	79
4.2.2.6	Experimento B.6.....	81

4.2.3	<i>Experimento C: Estratificação do conjunto de dados</i> .....	82
4.2.3.1	Experimento C.1.....	82
4.2.3.2	Experimento C.2.....	84
4.2.3.3	Experimento C.3.....	86
4.2.3.4	Experimento C.4.....	87
4.2.3.5	Experimento C.5.....	89
4.2.3.6	Experimento C.6.....	90
4.2.4	<i>Experimento D: Estratificação do conjunto de dados</i> .....	92
4.2.4.1	Experimento D.1.....	92
4.2.4.2	Experimento D.2.....	94
4.2.4.3	Experimento D.3.....	95
4.2.4.4	Experimento D.4.....	97
4.2.4.5	Experimento D.5.....	98
4.2.4.6	Experimento D.6.....	100
4.2.5	<i>Experimento E: Utilização da técnica de RDP e Comitês de RNA's</i> .....	101
4.2.5.1	Experimento E.1.....	101
4.2.5.2	Experimento E.2.....	103
4.2.5.3	Experimento E.3.....	105
4.2.5.4	Experimento E.4.....	106
4.2.6	<i>Experimento F: Classificação sem a Classe Dominante</i> .....	108
4.2.6.1	Experimento F.1.....	108
4.2.6.2	Experimento F.2.....	110
4.2.6.3	Experimento F.3.....	111
4.2.6.4	Experimento F.4.....	112
4.2.6.5	Experimento F.5.....	114
4.2.6.6	Experimento F.6.....	115
4.2.7	<i>Experimento G: Separação dos Atributos de Entrada</i> .....	116
4.2.7.1	Experimento G.1.....	117
4.2.7.2	Experimento G.2.....	118
4.2.7.3	Experimento G.3.....	120
4.2.7.4	Experimento G.4.....	121
4.2.7.5	Experimento G.5.....	123
4.2.7.6	Experimento G.6.....	124
4.2.8	<i>Experimento H: Utilização de Conjuntos Disjuntos de Dados</i> .....	126
4.2.8.1	Experimento H.1.....	126
4.2.8.2	Experimento H.2.....	128
4.2.8.3	Experimento H.3.....	129
4.2.8.4	Experimento H.4.....	131
4.2.8.5	Experimento H.5.....	132
4.2.8.6	Experimento H.6.....	133
4.3	COMENTÁRIOS DOS EXPERIMENTOS.....	135
4.3.1	<i>Quadro Resumo dos Experimentos</i> .....	135
4.3.2	<i>Experimento A</i> .....	136
4.3.3	<i>Experimento B</i> .....	137
4.3.4	<i>Experimento C</i> .....	138
4.3.5	<i>Experimento D</i> .....	140
4.3.6	<i>Experimento E</i> .....	141
4.3.7	<i>Experimento F</i> .....	142
4.3.8	<i>Experimento G</i> .....	144
4.3.9	<i>Experimento H</i> .....	145
<b>5</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS</b> .....	<b>147</b>
5.1	CONCLUSÕES.....	147
5.1.1	TRABALHOS FUTUROS.....	149
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>150</b>
	<b>ANEXOS</b> .....	<b>154</b>
	ANEXO 1 - DISTRIBUIÇÃO DAS CATEGORIAS ESPONTÂNEAS DO QUESITO COR.....	154
	ANEXO 2 – CRITÉRIOS DE AGREGAÇÃO DA VARIÁVEL COR PARA AS CATEGORIAS AGRUPADAS.....	157
	ANEXO 3 – COR ABERTA POR COR FECHADA NA REGIÃO METROPOLITANA DO RIO DE JANEIRO.....	158
	ANEXO 4 - DISTRIBUIÇÃO DA POPULAÇÃO POR COR SEGUNDO AS REGIÕES METROPOLITANAS. (PETRUCCELLI, P. 54).....	159
	<b>APÊNDICE – TRABALHO PUBLICADO PELO AUTOR</b> .....	<b>160</b>

## LISTA DE FIGURAS

Figura 1 - A Redenção de Can, por Modesto Brocos y Gómez, 1895, pertencente ao acervo do Museu Nacional de Belas Artes, Rio de Janeiro. ....	19
Figura 2- Diagrama esquemático de um neurônio (ZURADA, 1995, p. 28).....	39
Figura 3 - Representação esquemática da integração temporal dos estímulos por um neurônio (KOVÁCS, 1996, p. 25).....	40
Figura 4 - Esquema de unidade McCullock – Pitts (ZURADA, 1995, p. 36).....	40
Figura 5 - Uma rede <i>feed-forward</i> simples com duas entradas, dois nós ocultos , e um nó de saída (RUSSEL & NORVIG, 1995, p. 572). ....	41
Figura 6 - Rede Neural com duas camadas (ZURADA, 1995, p. 40).....	42
Figura 7 - Características das Funções de Limiar (KOVÁCS, 1996, p. 29). ....	44
Figura 8 - Conexão com retro-alimentação (ZURADA, 1995, p. 42). ....	45
Figura 9 - (a) Supervisionado, (b) Não Supervisionado (ZURADA, 1995, p. 57). ....	46
Figura 10 - Exemplo de saída desejada no lado esquerdo e saída com <i>overfitting</i> do lado direito (MATLAB, 1994, p. 36).....	47
Figura 11 - Unidades com Função de ativação em degrau podem agir como portas lógicas, com pesos e <i>thresholds</i> apropriados (RUSSEL & NORVIG, 1995, p. 570).....	48
Figura 12 - Separação linear em um perceptron (RUSSEL & NORVIG, 1995, p. 575). ....	48
Figura 13 - (a) Hipótese consistente. (b) Falso negativo. (c) A hipótese é generalizada. (d) Um falso-positivo. (e) A hipótese é especializada (RUSSEL & NORVIG, 1995, p. 547).....	49
Figura 14 - Método geral de aprendizado de uma rede neural (RUSSEL & NORVIG, 1995, p. 577).....	50
Figura 15 - Máquina de Comitê (HAYKIN, 2001, p. 387).....	52
Figura 16 - Curva conceitual de aprendizado do Experimento A.1. ....	64
Figura 17 - Curva conceitual de aprendizado do Experimento A.2. ....	66
Figura 18 - Curva conceitual de aprendizado do Experimento A.3. ....	67
Figura 19 - Curva conceitual de aprendizado do Experimento A.4. ....	69
Figura 20 - Curva conceitual de aprendizado do Experimento A.5. ....	71
Figura 21 - Curva conceitual de aprendizado do Experimento A.6. ....	72
Figura 22 - Curva conceitual de aprendizado do Experimento B.1. ....	74
Figura 23 - Curva conceitual de aprendizado do Experimento B.2. ....	76
Figura 24 - Curva conceitual de aprendizado do Experimento B.3. ....	77
Figura 25- Curva conceitual de aprendizado do Experimento B.4. ....	79
Figura 26 - Curva conceitual de aprendizado do Experimento B.5. ....	80
Figura 27 - Curva conceitual de aprendizado do Experimento B.6. ....	82
Figura 28 - Curva conceitual de aprendizado do Experimento C.1. ....	84
Figura 29 - Curva conceitual de aprendizado do Experimento C.2. ....	85
Figura 30 - Curva conceitual de aprendizado do Experimento C.3. ....	87
Figura 31 - Curva conceitual de aprendizado do Experimento C.4. ....	88
Figura 32 - Curva conceitual de aprendizado do Experimento C.5. ....	90
Figura 33 - Curva conceitual de aprendizado do Experimento C.6. ....	91
Figura 34 - Curva conceitual de aprendizado do Experimento D.1. ....	93
Figura 35 - Curva conceitual de aprendizado do Experimento D.2. ....	95
Figura 36 - Curva conceitual de aprendizado do Experimento D.3. ....	96
Figura 37 - Curva conceitual de aprendizado do Experimento D.4. ....	98
Figura 38 - Curva conceitual de aprendizado do Experimento D.5. ....	99
Figura 39 - Curva conceitual de aprendizado do Experimento D.6. ....	101



Figura 40 - Curva conceitual de aprendizado do Experimento E.1. ....	103
Figura 41 - Curva conceitual de aprendizado do Experimento E.2. ....	104
Figura 42 - Curva conceitual de aprendizado do Experimento E.3. ....	106
Figura 43 - Curva conceitual de aprendizado do Experimento E.4. ....	107
Figura 44 - Curva conceitual de aprendizado do Experimento F.1.....	109
Figura 45 - Curva conceitual de aprendizado do Experimento F.2.....	111
Figura 46 - Curva conceitual de aprendizado do Experimento F.3.....	112
Figura 47 - Curva conceitual de aprendizado do Experimento F.4.....	113
Figura 48 - Curva conceitual de aprendizado do Experimento F.5.....	115
Figura 49 - Curva conceitual de aprendizado do Experimento F.6.....	116
Figura 50 - Curva conceitual de aprendizado do Experimento G.1.....	118
Figura 51 - Curva conceitual de aprendizado do Experimento G.2.....	120
Figura 52 - Curva conceitual de aprendizado do Experimento G.3.....	121
Figura 53 - Curva conceitual de aprendizado do Experimento G.4.....	123
Figura 54 - Curva conceitual de aprendizado do Experimento G.5.....	124
Figura 55 - Curva conceitual de aprendizado do Experimento G.6.....	126
Figura 56 - Curva conceitual de aprendizado do Experimento H.1.....	128
Figura 57 - Curva conceitual de aprendizado do Experimento H.2.....	129
Figura 58 - Curva conceitual de aprendizado do Experimento H.3.....	130
Figura 59 - Curva conceitual de aprendizado do Experimento H.4.....	132
Figura 60 - Curva conceitual de aprendizado do Experimento H.5.....	133
Figura 61 - Curva conceitual de aprendizado do Experimento H.6.....	134
Figura 62 - Gráfico de dispersão dos atributos da base de dados de etnia .....	146

## LISTA DE TABELAS

Tabela 1 - Tipos sanguíneos X Presença de Aglutinogênios e Aglutininas (GENÉTICA, s. Polialelismo A-B-O, p. 2). .....	25
Tabela 2 - Relação Fenótipo X Genótipo (GENÉTICA, s. Polialelismo A-B-O, p. 3). .....	26
Tabela 3 - Relação Fenótipo X Genótipo (GENÉTICA, s. Interação Gênica, p. 2). .....	27
Tabela 4 - Relação Fenótipo X Genótipo (GENÉTICA, 2003, s. Herança Quantitativa, p.1). .....	28
Tabela 5 - Localização da humanidade através dos anos (KENSKI, 2003). .....	29
Tabela 6 - Clima <i>versus</i> características corporais (KENSKI, 2003). .....	29
Tabela 7 - Subespécies <i>versus</i> Características (KENSKI, 2003). .....	30
Tabela 8 - Mistura entre etnias (PETRUCCELLI, 1998; PENA, 2000). .....	32
Tabela 9 - Denominação das etnias pelos respectivos censos demográficos (PETRUCCELLI, 2000). .....	33
Tabela 10 - Distribuição das categorias de cor por pergunta fechada <i>versus</i> aberta (PETRUCCELLI, 2000, p. 28). .....	34
Tabela 11 - Distribuição das categorias mais significativas de pergunta aberta por Região Metropolitana (PETRUCCELLI, 2000, p. 32). .....	35
Tabela 12 - Distribuição das denominações intermediárias de cor (PETRUCCELLI, 2000, p. 33). .....	35
Tabela 13 - Distribuição da população por cor segundo as Regiões Metropolitanas (PETRUCCELLI, 2000, p. 54). .....	36
Tabela 14 - Redes Neurais - Estrutura. ....	46
Tabela 15 - Conjunto de treinamento original onde A, B, AB, D, C, RA e RB são atributos e CLASSE representa as três classes existentes. ....	54
Tabela 16 - Conjunto de treinamento especializado na Classe 1. A, B, AB, D, C, RA e RB são atributos. ....	55
Tabela 17 - Conjunto de treinamento especializado na Classe 2. A, B, AB, D, C, RA e RB são atributos. ....	55
Tabela 18 - Conjunto de treinamento especializado na Classe 3. A, B, AB, D, C, RA e RB são atributos. ....	56
Tabela 19 - Reatividade dos grupos sanguíneos, segundo reagentes. ....	59
Tabela 20 - Notação. ....	59
Tabela 21 - Exemplo de Matriz de Confusão Percentual. ....	61
Tabela 22 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.1. ....	63
Tabela 23 - Matriz de confusão percentual do conjunto de teste do Experimento A.1. ....	64
Tabela 24 - Percentuais de reconhecimento do Experimento A.1. ....	64
Tabela 25 - Vetor de competência do Experimento A.1. ....	64
Tabela 26 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.2. ....	65
Tabela 27 - Matriz de confusão percentual do conjunto de teste do Experimento A.2. ....	65
Tabela 28 - Percentuais de reconhecimento do Experimento A.2. ....	65
Tabela 29 - Vetor de competência. ....	66
Tabela 30 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.3. ....	66
Tabela 31 - Matriz de confusão percentual do conjunto de teste do Experimento A.3. ....	67
Tabela 32 - Percentuais de reconhecimento do Experimento A.3. ....	67
Tabela 33 - Vetor de competência do Experimento A.3. ....	67
Tabela 34 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.4. ....	68
Tabela 35 - Matriz de confusão percentual do conjunto de teste do Experimento A.4. ....	68
Tabela 36 - Percentuais de reconhecimento do Experimento A.4. ....	68

Tabela 37 - Vetor de competência do Experimento A.4.....	69
Tabela 38 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.5.....	70
Tabela 39 - Matriz de confusão percentual do conjunto de teste do Experimento A.5.....	70
Tabela 40 - Percentuais de reconhecimento do Experimento A.5.....	70
Tabela 41 - Vetor de competência do Experimento A.5.....	70
Tabela 42 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.6.....	71
Tabela 43 - Matriz de confusão percentual do conjunto de teste do Experimento A.6.....	72
Tabela 44 - Percentuais de reconhecimento do Experimento A.6.....	72
Tabela 45 - Vetor de competência do Experimento A.6.....	72
Tabela 46 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.1.....	73
Tabela 47 - Matriz de confusão percentual do conjunto de teste do Experimento B.1.....	74
Tabela 48 - Percentuais de reconhecimento do Experimento B.1.....	74
Tabela 49 - Vetor de competência do Experimento B.1.....	74
Tabela 50 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.2.....	75
Tabela 51 - Matriz de confusão percentual do conjunto de teste do Experimento B.2.....	75
Tabela 52 - Percentuais de reconhecimento do Experimento B.2.....	75
Tabela 53 - Vetor de competência do Experimento B.2.....	75
Tabela 54 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.3.....	76
Tabela 55 - Matriz de confusão percentual do conjunto de teste do Experimento B.3.....	77
Tabela 56 - Percentuais de reconhecimento do Experimento B.3.....	77
Tabela 57 - Vetor de competência do Experimento B.3.....	77
Tabela 58 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.4.....	78
Tabela 59 - Matriz de confusão percentual do conjunto de teste do Experimento B.4.....	78
Tabela 60 - Percentuais de reconhecimento do Experimento B.4.....	78
Tabela 61 - Vetor de competência do Experimento B.4.....	78
Tabela 62 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.5.....	79
Tabela 63 - Matriz de confusão percentual do conjunto de teste do Experimento B.5.....	80
Tabela 64 - Percentuais de reconhecimento do Experimento B.5.....	80
Tabela 65 - Vetor de competência do Experimento B.5.....	80
Tabela 66 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.6.....	81
Tabela 67 - Matriz de confusão percentual do conjunto de teste do Experimento B.6.....	81
Tabela 68 - Percentuais de reconhecimento do Experimento B.6.....	81
Tabela 69 - Vetor de competência do Experimento B.6.....	82
Tabela 70 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.1.....	83
Tabela 71 - Matriz de confusão percentual do conjunto de teste do Experimento C.1.....	83
Tabela 72 - Percentuais de reconhecimento do Experimento C.1.....	83
Tabela 73 - Vetor de competência do Experimento C.1.....	84
Tabela 74 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.2.....	84
Tabela 75 - Matriz de confusão percentual do conjunto de teste do Experimento C.2.....	85
Tabela 76 - Percentuais de reconhecimento do Experimento C.2.....	85
Tabela 77 - Vetor de competência do Experimento C.2.....	85
Tabela 78 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.3.....	86
Tabela 79 - Matriz de confusão percentual do conjunto de teste do Experimento C.3.....	86
Tabela 80 - Percentuais de reconhecimento do Experimento C.3.....	86
Tabela 81 - Vetor de competência do Experimento C.3.....	87
Tabela 82 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.4.....	88
Tabela 83 - Matriz de confusão percentual do conjunto de teste do Experimento C.4.....	88
Tabela 84 - Percentuais de reconhecimento do Experimento C.4.....	88
Tabela 85 - Vetor de competência do Experimento C.4.....	88
Tabela 86 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.5.....	89

Tabela 87 - Matriz de confusão percentual do conjunto de teste do Experimento C.5.....	89
Tabela 88 - Percentuais de reconhecimento do Experimento C.5.....	89
Tabela 89 - Vetor de competência do Experimento C.5. ....	90
Tabela 90 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.6.....	91
Tabela 91 - Matriz de confusão percentual do conjunto de teste do Experimento C.6.....	91
Tabela 92 - Percentuais de reconhecimento do Experimento C.6.....	91
Tabela 93 - Vetor de competência do Experimento C.6. ....	91
Tabela 94 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.1.....	93
Tabela 95 - Matriz de confusão percentual do conjunto de teste do Experimento D.1. ....	93
Tabela 96 - Percentuais de reconhecimento do Experimento D.1. ....	93
Tabela 97 - Vetor de competência do Experimento D.1.....	93
Tabela 98 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.2.....	94
Tabela 99 - Matriz de confusão percentual do conjunto de teste do Experimento D.2. ....	94
Tabela 100 - Percentuais de reconhecimento do Experimento D.2. ....	94
Tabela 101 - Vetor de competência do Experimento D.2. ....	95
Tabela 102 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.3. .....	95
Tabela 103 - Matriz de confusão percentual do conjunto de teste do Experimento D.3. ....	96
Tabela 104 - Percentuais de reconhecimento do Experimento D.3. ....	96
Tabela 105 - Vetor de competência do Experimento D.3.....	96
Tabela 106 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.4. .....	97
Tabela 107 - Matriz de confusão percentual do conjunto de teste do Experimento D.4. ....	97
Tabela 108 - Percentuais de reconhecimento do Experimento D.4. ....	97
Tabela 109 - Vetor de competência do Experimento D.4.....	98
Tabela 110 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.5. .....	98
Tabela 111 - Matriz de confusão percentual do conjunto de teste do Experimento D.5. ....	99
Tabela 112 - Percentuais de reconhecimento do Experimento D.5. ....	99
Tabela 113 - Vetor de competência do Experimento D.5.....	99
Tabela 114 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.6. .....	100
Tabela 115 - Matriz de confusão percentual do conjunto de teste do Experimento D.6. ....	100
Tabela 116 - Percentuais de reconhecimento do Experimento D.6. ....	100
Tabela 117 - Vetor de competência do Experimento D.6.....	101
Tabela 118 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.1. .....	102
Tabela 119 - Matriz de confusão percentual do conjunto de teste do Experimento E.1.....	102
Tabela 120 - Percentuais de reconhecimento do Experimento E.1.....	102
Tabela 121 - Vetor de competência do Experimento E.1. ....	103
Tabela 122 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.2. .....	103
Tabela 123 - Matriz de confusão percentual do conjunto de teste do Experimento E.2.....	104
Tabela 124 - Percentuais de reconhecimento do Experimento E.2.....	104
Tabela 125 - Vetor de competência do Experimento E.2. ....	104
Tabela 126 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.3. .....	105
Tabela 127 - Matriz de confusão percentual do conjunto de teste do Experimento E.3.....	105
Tabela 128 - Percentuais de reconhecimento do Experimento E.3.....	105
Tabela 129 - Vetor de competência do Experimento E.3. ....	106

Tabela 130 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.4. .....	106
Tabela 131 - Matriz de confusão percentual do conjunto de teste do Experimento E.4.....	107
Tabela 132 - Percentuais de reconhecimento do Experimento E.4.....	107
Tabela 133 - Vetor de competência do Experimento E.4. ....	107
Tabela 134- - Matriz de confusão percentual do conjunto de treinamento do Experimento F.1. .....	108
Tabela 135 - Matriz de confusão percentual do conjunto de teste do Experimento F.1. ....	109
Tabela 136 - Percentuais de reconhecimento do Experimento F.1. ....	109
Tabela 137 - Vetor de competência do Experimento F.1.....	109
Tabela 138 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.2. .....	110
Tabela 139 - Matriz de confusão percentual do conjunto de teste do Experimento F.2. ....	110
Tabela 140 - Percentuais de reconhecimento do Experimento F.2. ....	110
Tabela 141 - Vetor de competência do Experimento F.2.....	110
Tabela 142 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.3. .....	111
Tabela 143 - Matriz de confusão percentual do conjunto de teste do Experimento F.3. ....	111
Tabela 144 - Percentuais de reconhecimento do Experimento F.3. ....	112
Tabela 145 - Vetor de competência do Experimento F.3.....	112
Tabela 146 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.4. .....	113
Tabela 147 - Matriz de confusão percentual do conjunto de teste do Experimento F.4. ....	113
Tabela 148 - Percentuais de reconhecimento do Experimento F.4. ....	113
Tabela 149 - Vetor de competência do Experimento F.4.....	113
Tabela 150 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.5. .....	114
Tabela 151 - Matriz de confusão percentual do conjunto de teste do Experimento F.5. ....	114
Tabela 152 - Percentuais de reconhecimento do Experimento F.5. ....	114
Tabela 153 - Vetor de competência do Experimento F.5.....	114
Tabela 154 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.6. .....	115
Tabela 155 - Matriz de confusão percentual do conjunto de teste do Experimento F.6. ....	115
Tabela 156 - Percentuais de reconhecimento do Experimento F.6. ....	116
Tabela 157 - Vetor de competência do Experimento F.6.....	116
Tabela 158 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.1. .....	117
Tabela 159 - Matriz de confusão percentual do conjunto de teste do Experimento G.1. ....	118
Tabela 160 - Percentuais de reconhecimento do Experimento G.1. ....	118
Tabela 161 - Vetor de competência do Experimento G.1.....	118
Tabela 162 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.2. .....	119
Tabela 163 - Matriz de confusão percentual do conjunto de teste do Experimento G.2. ....	119
Tabela 164 - Percentuais de reconhecimento do Experimento G.2. ....	119
Tabela 165 - Vetor de competência do Experimento G.2.....	119
Tabela 166 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.3. .....	120
Tabela 167 - Matriz de confusão percentual do conjunto de teste do Experimento G.3. ....	121
Tabela 168 - Percentuais de reconhecimento do Experimento G.3. ....	121
Tabela 169 - Vetor de competência do Experimento G.3.....	121

Tabela 170 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.4. .....	122
Tabela 171 - Matriz de confusão percentual do conjunto de teste do Experimento G.4. ....	122
Tabela 172 - Percentuais de reconhecimento do Experimento G.4. ....	122
Tabela 173 - Vetor de competência do Experimento G.4. ....	122
Tabela 174 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.5. .....	123
Tabela 175 - Matriz de confusão percentual do conjunto de teste do Experimento G.5. ....	124
Tabela 176 - Percentuais de reconhecimento do Experimento G.5. ....	124
Tabela 177 - Vetor de competência do Experimento G.5. ....	124
Tabela 178 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.6. .....	125
Tabela 179 - Matriz de confusão percentual do conjunto de teste do Experimento G.6. ....	125
Tabela 180 - Percentuais de reconhecimento do Experimento G.6. ....	125
Tabela 181 - Vetor de competência do Experimento G.6. ....	125
Tabela 182 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.1. .....	127
Tabela 183 - Matriz de confusão percentual do conjunto de teste do Experimento H.1. ....	127
Tabela 184 - Percentuais de reconhecimento do Experimento H.1. ....	127
Tabela 185 - Vetor de competência do Experimento H.1. ....	127
Tabela 186 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.2. .....	128
Tabela 187 - Matriz de confusão percentual do conjunto de teste do Experimento H.2. ....	128
Tabela 188 - Percentuais de reconhecimento do Experimento H.2. ....	129
Tabela 189 - Vetor de competência do Experimento H.2. ....	129
Tabela 190 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.3. .....	130
Tabela 191 - Matriz de confusão percentual do conjunto de teste do Experimento H.3. ....	130
Tabela 192 - Percentuais de reconhecimento do Experimento H.3. ....	130
Tabela 193 - Vetor de competência do Experimento H.3. ....	130
Tabela 194 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.4. .....	131
Tabela 195 - Matriz de confusão percentual do conjunto de teste do Experimento H.4. ....	131
Tabela 196 - Percentuais de reconhecimento do Experimento H.4. ....	131
Tabela 197 - Vetor de competência do Experimento H.4. ....	131
Tabela 198 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.5. .....	132
Tabela 199 - Matriz de confusão percentual do conjunto de teste do Experimento H.5. ....	132
Tabela 200 - Percentuais de reconhecimento do Experimento H.5. ....	133
Tabela 201 - Vetor de competência do Experimento H.5. ....	133
Tabela 202 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.6. .....	134
Tabela 203 - Matriz de confusão percentual do conjunto de teste do Experimento H.6. ....	134
Tabela 204 - Percentuais de reconhecimento do Experimento H.6. ....	134
Tabela 205 - Vetor de competência do Experimento H.6. ....	134
Tabela 206 - Resumo das variações de técnicas aplicadas aos experimentos. ....	135
Tabela 207 - Comparação dos resultados do Experimento A. Percentuais de reconhecimento. .....	136
Tabela 208 - Comparação dos resultados do Experimento A. Vetor de Competência. ....	137

Tabela 209 - Comparação dos resultados do Experimento <i>B</i> . Percentuais de reconhecimento. .....	138
Tabela 210 - Comparação dos resultados do Experimento <i>B</i> . Vetor de Competência.....	138
Tabela 211 - Comparação dos resultados do Experimento <i>C</i> . Percentuais de reconhecimento. .....	139
Tabela 212 - Comparação dos resultados do Experimento <i>C</i> . Vetor de Competência. ....	139
Tabela 213 - Comparação dos resultados do Experimento <i>D</i> . Percentuais de reconhecimento. .....	140
Tabela 214 - Comparação dos resultados do Experimento <i>D</i> . Vetor de Competência. ....	140
Tabela 215 - Comparação dos resultados do Experimento <i>E</i> . Percentuais de reconhecimento. .....	142
Tabela 216 - Comparação dos resultados do Experimento <i>E</i> . Vetor de Competência.....	142
Tabela 217 - Comparação dos resultados do Experimento <i>F</i> . Percentuais de reconhecimento. .....	143
Tabela 218 - Comparação dos resultados do Experimento <i>F</i> . Vetor de Competência.....	143
Tabela 219 - Comparação dos resultados do Experimento <i>G</i> . Percentuais de reconhecimento. .....	144
Tabela 220 - Comparação dos resultados do Experimento <i>G</i> . Vetor de Competência. ....	145
Tabela 221 - Comparação dos resultados do <i>Experimento H</i> . Percentuais de reconhecimento. .....	146
Tabela 222 - Comparação dos resultados do Experimento <i>H</i> . Vetor de Competência. ....	146
Tabela 223 - Distribuição das categorias espontâneas do quesito Cor (PETRUCCELLI, p. 46) - (parte 1) .....	154
Tabela 224 - Distribuição das categorias espontâneas do quesito Cor (PETRUCCELLI, p. 46) - (parte 2) .....	155
Tabela 225 - Distribuição das categ. espontâneas do quesito Cor (PETRUCCELLI, 2000, p. 46) - (parte 3) .....	156
Tabela 226 - Agregação da variável Cor por Categoria (PETRUCCELLI, 2000, p. 47). ....	157
Tabela 227 - Cor Aberta X Cor Fechada - Rio de Janeiro (PETRUCCELLI, 2000, p. 50)...	158
Tabela 228 - Cor X Região Metropolitana (PETRUCCELLI, 2000, P. 54). ....	159

## RESUMO

A Epidemiologia é o ramo da Medicina relacionado ao estudo da incidência e prevalência das doenças em uma dada população. Cada vez mais há interesse em se estudar a susceptibilidade étnica em variadas doenças, tais como a anemia falciforme, que é mais comum em negros e a esclerose múltipla, em caucasianos. Em populações heterogêneas, onde é freqüente o casamento inter-racial, a determinação da etnia é particularmente difícil. Como definir a etnia em uma população altamente miscigenada, como a do Brasil, e quantificar a incidência de doenças e o acesso escolar e aos serviços de saúde por cada grupo étnico? Objetivando contribuir para o estudo da determinação da etnia simplificada através da reatividade individual aos reagentes utilizados na aferição do grupo sanguíneo ABO e Rh, este trabalho propõe um modelo conexionista, através de redes MLP. Para isto, utiliza-se o conceito de comitês de redes neurais com resultados sendo obtidos pelo método de combinação por média e especialização do conjunto de treinamento através da técnica de Replicação Dirigida de Padrões (RDP). Pelos resultados apresentados, verificou-se que, com a base de dados utilizada na pesquisa, não se obteve resultados significativos que abonem a utilização desta técnica, para a determinação, através do conjunto de reagentes apresentados, da etnia simplificada dos indivíduos. Uma grande variação de esquemas classificadores, e de configurações e arquiteturas de rede, foram testadas. Contudo, um resultado significativo foi apresentado em relação à determinação dos respectivos tipos sanguíneos, utilizando-se o mesmo conjunto de reagentes.

Palavras-chave: sangue, etnia, redes neurais artificiais, classificadores neurais, comitês de redes neurais.



## ABSTRACT

The Epidemiology is the branch of the Medicine related to the study of the incidence and prevalence of the illnesses in a given population. There is growing interest in studying the ethnic susceptibility in varied illnesses, such as the falciforme anemy, that is more common in blacks and the multiple sclerosis, in caucasians. In heterogeneous populations, where the Inter-racial marriage is frequent, the determination of the ethnicity is particularly difficult. How to define the ethnicity in a population highly mixed, as of Brazil, and to quantify the incidence of illnesses and the pertaining to school access and to the services of health for each ethnic group? Aiming at contributing to the determination of the simplified ethnicity through the individual reactivity to the reagents used in the gauging of blood group ABO and Rh, this work considers a connexionist model, through MLP nets. For this, are used concepts of neural network committees with results being acquired by the method of average and specialization of the training set through conduct pattern replication (RDP). For the presented results, it was verified that, with the database used in the research, it did not get resulted significant that warrants the use of this technique, for the determination, through the set of presented reagents, of the simplified ethnicity of the individuals. A great variation of classification projects, and configurations and architectures of net, had been tested. However, a significant result was presented regarding to the determination of the respective sanguine types, using itself the same joint of reagents.

Key-words: neural, classifying blood, ethnicity, artificial neural nets, neural classifiers, committees of neural nets.

## 1 INTRODUÇÃO

### 1.1 MOTIVAÇÕES

Em Casa-Grande e Senzala (1933), Gilberto Freire apresenta a mestiçagem como um elemento crucial da formação nacional e como um símbolo da democrática e flexível cultura brasileira. Esta visão não é compartilhada por todos; por exemplo, Florestan Fernandes, em sua reflexão “A Integração do Negro à Sociedade de Classes” (FERNANDES, 1965, v. II, p. 393-394 apud MAIO & SANTOS, 2002), diz que mantemos o negro à margem de uma sociedade que ele ajudou a levantar (MAIO & SANTOS, 2002).

No decorrer da história encontramos a discriminação racial nas mais diversas formas. Como ilustração, pode-se citar a obra de Modesto Brocos y Gómez, intitulada “A Redenção de Can” na qual se observa uma negra com os braços estendidos aos céus em agradecimento pela filha mulata que casou com um europeu e teve um filho branco. Este quadro (Figura 1) retrata claramente uma tendência antiga de se tentar clarear a população a fim de se “purificar” uma “raça”. Haveria necessidade de se tentar estudar o problema de encontrar uma forma justa, precisa, correta, não tendenciosa e, principalmente não discriminante de se determinar a etnia da população em nosso país.

Apesar da classificação étnica, em alguns casos, ser utilizada com objetivos negativos, podemos identificar motivações para o uso, possivelmente positivas, de tal classificação. Neste sentido temos as tentativas de correção dos abismos (diferenças) sociais, econômicos,

educacionais e até mesmo de saúde dos indivíduos oriundos predominantemente de determinadas etnias.



**Figura 1 - A Redenção de Can, por Modesto Brocos y Gómez, 1895, pertencente ao acervo do Museu Nacional de Belas Artes, Rio de Janeiro.**

Com séculos de considerações subjetivas acerca deste tema, chegamos aos tempos atuais em que geneticistas, com novas ferramentas e técnicas, a partir da década de 50, começam a estudar as relações *intra* e *interétnicas*. No Brasil foram utilizados marcadores genéticos clássicos, como sistema de grupos sanguíneos e proteínas séricas nos estudos da mistura racial. Salzano, em 1986, escreveu:

“A análise das características hematológicas permite não só identificar genes típicos de outras raças em pessoas de grupos aparentemente não miscigenados, como quantificar o grau de mistura presente em determinado grupo ao cabo de séculos de convivência entre membros de duas ou mais raças” (SALZANO, 1986, p. 52 apud MAIO & SANTOS, 2002).

Genética, talvez nesta área do conhecimento esteja a resposta para as indagações das sociedades de diversos países sobre questões polêmicas, tais como as Origens do Homem Moderno, e qual grupo étnico ou racial predomina em um determinado indivíduo. No livro “Homo Brasilis” (PENA, 2002), Sergio Danilo Pena demonstra resultados alcançados em seu trabalho denominado “Retrato Molecular do Brasil” (PENA et al, 2000). Na referida obra, foram utilizados marcadores genéticos do DNA<sup>1</sup> *mitochondrial* em uma amostra de indivíduos

<sup>1</sup> Ac. Nucléico formado por 2 fitas de nucleotídeos unidas por pontes de hidrogênio.

*autoclassificados* como brancos e ficou demonstrado que a amostra apresentava maior frequência de indivíduos de origem africana e/ou ameríndia (MAIO & SANTOS, 2002).

Mencionou-se nos parágrafos acima a eficácia dos testes genéticos para a determinação da origem étnica de um povo/indivíduo. Contudo, vale a pena ressaltar que os testes genéticos são relativamente caros, tanto para a maioria dos indivíduos da população de uma nação, quanto para estudos populacionais. Sabe-se que o custo atual para o teste de DNA que determina a ancestralidade de um ser humano é de aproximadamente U\$ 160.00, feito por FRUDAKIS (2003).

É interessante, então, buscar-se outra maneira de conseguir uma classificação étnica de um indivíduo que seja barata, relativamente precisa e eficaz. Então, este trabalho de dissertação procura contribuir para o estudo de uma forma mais barata de determinação da etnia predominante em um indivíduo; ou melhor, busca utilizar-se de um fenômeno chamado *interação gênica*, para tentar correlacionar sangue com etnia. Para a determinação do tipo sanguíneo de uma pessoa, costuma-se usar um conjunto de reagentes, denominados *A*, *B*, *C*, *D*, *RA* (reverso de *A*) e *RB* (reverso de *B*). Através deles são determinados os grupos sanguíneos em *A*, *B*, *AB* e *O*. Neste trabalho, estes conjuntos serão utilizados numa tentativa de se classificar corretamente a etnia simplificada dos indivíduos.

A relação entre grupos sanguíneos e etnia já é conhecida na literatura, (ZAGO, 1996). Um recente estudo, realizado em Cuba (HERNANDEZ, 1997), mostrou que a classificação do grupo sanguíneo *ABO* em subtipos poderá ser de grande utilidade para o conhecimento da etnia predominante em um indivíduo pertencente a uma população miscigenada, como a dos cubanos, e, por extensão, a de brasileiros.

Além dos fatores já citados para a importância de uma correta classificação étnica de um grupo, pode-se mencionar a influência da etnia na manifestação de quadros clínicos mais ou menos graves, como, por exemplo, a esclerose múltipla. Faltam, na literatura, estudos suficientes para classificar a etnia do brasileiro segundo a tipagem sanguínea, o que poderia diminuir a possibilidade de erros cometidos em classificação, baseados somente no fenótipo de cor de pele, tipo de cabelo e feições. Atualmente, mesmo no censo realizado pelo IBGE, o método ainda é a autodenominação da cor (PETRUCCELLI, 2000).

Podemos, então, dizer que o estudo da correta classificação étnica de indivíduos possui aplicabilidade social, podendo vir a corrigir estudos estatísticos, quanto à comparação entre brancos, pardos, negros e outras etnias no acesso aos serviços públicos, escolares e hospitalares (PENA, 2002; PETRUCCELLI, 2000; LEITE, 2002). Pode-se citar, por exemplo, o problema do uso de cotas para acesso de alunos às faculdades públicas, pelo Vestibular, no Estado do Rio de Janeiro, mais precisamente na UERJ, aonde uma percentagem das vagas foi reservada a alunos de etnias desprestigiadas (negros, pardos e índios) e alunos advindos de escolas públicas (LIMA, 2002; BRANCOS, 2003; CENSO, 2000; GTI, 2004; ABRANCHES, 2003; SISTEMA, 2003; MEROLA, 2003).

## 1.2 OBJETIVOS E DESCRIÇÃO DO TRABALHO

Este trabalho tem por objetivo utilizar técnicas de Redes Neurais Artificiais (Redes MLP – Seção 3) para tentar, através do conjunto de reagentes que determinam os grupos sanguíneos em  $A$ ,  $B$ ,  $AB$  e  $O$ , classificar corretamente a etnia simplificada dos indivíduos de uma população. Caso isto seja possível, teremos obtido uma forma eficaz e a um custo bem mais baixo que os modernos testes de DNA.

Com este fim, será pesquisada a influência da reatividade com os reagentes  $A$ ,  $B$ ,  $AB$ ,  $C$ ,  $D$ ,  $RA$  e  $RB$  na etnia, e através de redes neurais determinar se é adequada a utilização de um classificador neural com alto índice de confiabilidade e baixa probabilidade de erro. Para tal, os dados serão testados nas seguintes etapas: treinamento da rede; aferição da rede, e correlação de dados com a etnia simplificada.

Os estudos para a confecção deste trabalho envolveram as seguintes áreas: i) pesquisa bibliográfica, a fim de se investigar experimentos que permitissem correlacionar esses conjuntos de reagentes à etnia simplificada; ii) estudo de Redes Neurais Artificiais, formas de treinamento, performance, formação de comitês (Seção 3.7.3) e métodos de combinação de classificadores. Com base nestes estudos, foi proposta e realizada uma série de experimentos, cada um deles contendo variantes, a fim de investigar uma possível correlação entre os fatores sanguíneos mencionados e a etnia simplificada (predominante) de um indivíduo. As variações

da arquitetura das redes, bem como dos métodos de combinação de classificadores visavam melhorar a capacidade de classificação das mesmas, sem induzir resultados.

### **1.3 ORGANIZAÇÃO DOS CAPÍTULOS**

Esta dissertação é composta, além do capítulo introdutório, de mais quatro capítulos, quatro anexos e um apêndice.

O Capítulo 2 detalha tópicos relacionados à genética, cor (subjetividade) e tipagem sangüínea.

O Capítulo 3 apresenta conceitos de Redes Neurais Artificiais, comitês e o método de variação do conjunto de treinamento.

O Capítulo 4 apresenta um conjunto de experimentos utilizando-se dos conceitos estudados, e uma aplicação prática – Classificação Étnica.

O Capítulo 5 apresenta as principais conclusões desta tese, além de discussões sobre trabalhos futuros.

Os Anexos de 1 a 4 trazem a relação do quesito Cor com diversos fatores de distribuição.

O Apêndice apresenta o trabalho publicado pelo autor no decorrer desta pesquisa.

## 2 CONCEITOS EM BIOLOGIA E ANTROPOLOGIA

### 2.1 GENÉTICA, TIPAGEM SANGUINEA E COR

Ao discorrer sobre os fatores que influenciam a etnia das pessoas, principalmente no que diz respeito a este trabalho, torna-se relevante rever conceitos relativos à transmissão de caracteres dos seres vivos. Para isto, é necessário levar em consideração a transmissão genética de informações. Inicialmente, podemos citar a *Teoria da Transmissão dos Caracteres Adquiridos*, do naturalista francês Jean Baptiste Lamarck, em 1801 (GENÉTICA, 2003, s. Histórico, p. 1; BERKELEY, 2003) - “as variações dos fatores ambientais induziriam os seres vivos a sofrerem modificações estruturais necessárias a uma melhor adaptação, aumentando as chances de sobrevivência. Tais modificações seriam transmitidas aos descendentes”.

De 1800 Em diante, várias outras teorias sobre a transmissão de caracteres foram lançadas. Dos trabalhos pioneiros, o mais relevante foi o de Gregor Mendel (GENÉTICA, 2003, s. Histórico), em 1865, que realizou diversos cruzamentos com ervilhas e chegou à conclusão de que existiam fatores hereditários transmitidos através das gerações. Estes fatores são conhecidos como genes<sup>2</sup>. Além disso, ele também chegou à conclusão de que os genes ocorreriam aos pares (alelos<sup>3</sup>) nas células somáticas<sup>4</sup>, sendo individualizados nos gametas<sup>5</sup> e

---

<sup>2</sup> Fragmento de DNA cromossômico capaz de determinar a síntese de uma proteína.

<sup>3</sup> São genes situados no mesmo locus de cromossomos homólogos, responsáveis pela determinação de um mesmo caráter.

<sup>4</sup> Todas as células corporais (2N) que possuem o cariótipo completo.

<sup>5</sup> Células que sofreram meiose e possuem apenas metade da carga genética do indivíduo.

reunindo-se de novo, aos pares, no ato da fecundação. Outros estudos corroboraram as conclusões de Mendel; dentre eles, pode-se citar os de Cuénot e Batson (GENÉTICA, 2003), que em 1902 mostraram que os mesmos princípios utilizados na transmissão das características das ervilhas de Mendel seriam válidos para animais.

### 2.1.1 Alguns Conceitos Iniciais em Genética

As características dos seres humanos, como, por exemplo, a cor da pele, ou até mesmo o tipo sanguíneo, são determinadas pelos genes. Genes que determinam variedades diferentes do mesmo caráter são denominados alelos e ocupam um local específico no cromossomo<sup>6</sup>, chamado locus gênico. A constituição genética para uma determinada característica expressa o genótipo de um indivíduo, que é determinado pelos genes alelos. O genótipo, por sua vez, pode ser influenciado pelo meio ambiente, acarretando mudanças expressas no fenótipo, que é o somatório de todas as características que são observáveis em um determinado ser. Quando temos características de um indivíduo sendo definidas por alelos iguais, o indivíduo denomina-se homocigoto; caso contrário, denomina-se heterocigoto em relação à característica em questão. Os alelos podem ser dominantes (normalmente representados por letras maiúsculas) ou recessivos (letras minúsculas) (GENÉTICA, 2003). No cruzamento entre dois seres, pode ocorrer um efeito chamado de *herança sem dominância*, que consiste no fato de dois alelos interagirem de modo que o heterocigoto (no caso o filho) apresente características fenotípicas intermediárias aos apresentados pelos pais; por exemplo, quando os pais têm cores diferentes e o filho apresenta um tom de pele intermediário (GENÉTICA, 2003, s. 1ª Lei de Mendel).

Os estudos de Mendel prosseguiram e ele passou a estudar, ao invés de um par de cada vez, dois pares de caracteres por vez; isto é, duas características simultaneamente. Tem-se, então, a 2ª Lei de Mendel, ou Lei da Segregação Independente, que consiste em aplicar a Lei de Mendel para o estudo de duas, três ou mais características, simultaneamente, determinadas por alelos situados em pares de cromossomos homólogos<sup>7</sup> diferentes (GENÉTICA, s. 2ª Lei de Mendel). Quando temos características determinadas por três ou mais genes alelos

---

<sup>6</sup> Estrutura nuclear formada pela molécula de DNA mais proteínas, de forma espiralizada, contendo uma sucessão linear de genes.

<sup>7</sup> Cromossomos que apresentam o mesmo tamanho, mesma posição do centrômero e a mesma seqüência gênica (um de origem materna, e outro paterna).



contrastantes, dá-se o nome de *polialelia*. Neste caso, para a determinação de uma característica, os alelos interagem dois a dois (GENÉTICA, s. Polialelismo A-B-O).

Este trabalho desenvolve experimentos computacionais, buscando evidenciar a existência de relacionamento entre características independentes, mais especificamente sangue e cor da pele (etnia simplificada).

### 2.1.2 Sistema A B O

A espécie humana tem quatro tipos sangüíneos básicos, que são: **A**, **B**, **AB** e **O**. O organismo, quando toma contato com *antígenos*<sup>8</sup>, produz *anticorpos*<sup>9</sup>, neutralizando o efeito dos mesmos. Este processo é chamado de *defesa humoral*. Entre os componentes do sangue, temos as hemácias e o plasma. Nas hemácias, temos dois tipos de antígenos, denominados *aglutinogênios A* e *B*, que são responsáveis pela determinação do fenótipo sangüíneo. O plasma tem dois anticorpos, denominados *aglutininas anti-A* e *anti-B*. Contatos de tipos sangüíneos diferentes desencadeiam reações de antígenos versus anticorpos. O sangue de tipo **A** contém *aglutinogênio A* e *aglutinina anti-B*. O sangue do tipo **B** contém *aglutinogênio B* e *aglutinina anti-A*. O sangue **AB**, por sua vez, possui *aglutinogênios A* e *B*, e não possui *aglutininas*. Por último, o sangue tipo **O** não possui *aglutinogênios* e possui *aglutininas anti-A* e *anti-B*. Um resumo pode ser visto na Tabela 1 (GENÉTICA, 2003, s. Polialelismo A-B-O).

TIPO SANGÜÍNEO	AGLUTINOGÊNIOS NAS HEMÁCIAS	AGLUTININAS NO PLASMA
<b>A</b>	A	anti-B
<b>B</b>	B	anti-A
<b>AB</b>	A e B	-
<b>O</b>	-	anti-A e anti-B

Tabela 1 - Tipos sangüíneos X Presença de Aglutinogênios e Aglutininas (GENÉTICA, s. Polialelismo A-B-O, p. 2).

<sup>8</sup> Proteínas estranhas a um determinado organismo.

<sup>9</sup> Proteína.

### 2.1.3 Aspectos da Herança do Sistema ABO

Tem-se no sistema ABO quatro diferentes fenótipos, determinados por três diferentes alelos:  $I^A$ ,  $I^B$  e  $i$ . Eles apresentam a seguinte relação de dominância:  $I^A = I^B$ ,  $I^A > i$ ,  $I^B > i$ . A Tabela 2, a seguir, apresenta a relação entre fenótipo e genótipo (GENÉTICA, 2003, s. Polialelismo A-B-O, p. 3).

FENÓTIPOS	GENÓTIPOS	
<b>A</b>	$I^A I^A$	$I^A i$
<b>B</b>	$I^B I^B$	$I^B i$
<b>AB</b>	$I^A I^B$	
<b>O</b>	$ii$	

Tabela 2 - Relação Fenótipo X Genótipo (GENÉTICA, s. Polialelismo A-B-O, p. 3).

### 2.1.4 Interação Gênica

Um importante fenômeno a se mencionar neste trabalho é a *interação gênica*, que consiste no processo pelo qual dois ou mais pares de genes, com distribuições independentes, afetam uma única característica. Como exemplo, pode-se citar a interação gênica não-epistática no cruzamento entre aves: a forma das cristas em galinhas, que podem apresentar os seguintes fenótipos: ervilha, rosa, noz e simples. Esse caráter depende da interação entre dois pares alelos:  $R$  e  $E$ . Cada par apresenta um gene que atua como dominante ( $R$  ou  $E$ ) em relação ao outro, recessivo ( $r$  ou  $e$ ), capazes de determinar: 1) Crista ervilha: na presença do gene  $E$ , desde que não ocorra o gene  $R$ ; 2) Crista rosa: na presença do gene  $R$ , desde que não ocorra o gene  $E$ ; 3) Crista noz: quando ocorrem os genes  $E$  e  $R$ ; 4) Crista simples: manifesta-se na ausência dos genes  $E$  e  $R$ .

### 2.1.4.1 Epistasia

Uma das modalidades de interação gênica, chamada *epistasia*, consiste no fato de um par de alelos inibir a manifestação de genes de outros pares. Aos genes que impedem a atuação de outros (chamados *hipoestáticos*), dá-se o nome de genes *epistáticos*. Como exemplo, pode-se citar o caso das galinhas da raça *Leghorn*. Elas possuem um gene C, dominante, que caracteriza plumagem colorida. Aves coloridas devem ter genótipos CC ou Cc. Se a ave tiver um gene dominante I, epistático em relação a C, não terá plumagem colorida. A relação fenótipo *versus* genótipo pode ser vista na Tabela 3 (GENÉTICA, 2003, s. Interação Gênica).

FENÓTIPOS	GENÓTIPOS
<b>Plumagem colorida</b>	CCii ou Ccii
<b>Plumagem branca</b>	CCII, CCIi, CcII, ccII, ccIi, ccii

Tabela 3 - Relação Fenótipo X Genótipo (GENÉTICA, s. Interação Gênica, p. 2).

Um fenômeno inverso à interação gênica é chamado de *pleiotropia* e consiste no fato de um único par de genes atuar em várias características. Como exemplo, pode-se citar a *fenilcetonúria* (MEDLINEPLUS, 2003). A criança que tem essa doença é portadora de um par de alelos recessivos, que causam uma ausência de uma enzima chamada *fenilalanina hidroxilase*, que, por sua vez, está indiretamente ligada à produção de melanina. As crianças acometidas dessa doença apresentam pele mais clara do que deveriam ter. Aqui, nesse caso, um único par de genes atua em dois caracteres diferentes, que são: cor da pele e capacidade de metabolização da enzima (GENÉTICA, 2003, s. Interação Gênica).

Existe ainda um outro tipo de interação gênica, chamada *herança quantitativa*, que consiste no fato de termos dois ou mais pares de genes apresentando seus efeitos somados, em relação a um mesmo caractere, resultando diferentes intensidades de um mesmo fenótipo. Como exemplo de herança quantitativa, pode-se citar a coloração da pele humana. Simplificadamente, pode-se considerar que dois pares de genes condicionam a produção de **melanina**<sup>10</sup>. Dois genes dominantes A e B (tem-se os respectivos alelos a e b, recessivos) condicionam a síntese de melanina. “Quanto maior o número de genes dominantes em um

<sup>10</sup> Pigmento escuro, cuja quantidade determina a coloração da pele.

indivíduo, mais escura será a sua pele” (GENÉTICA, s. Herança Qualitativa, p. 1). A relação fenótipo *versus* genótipo, em relação à cor da pele, pode ser vista na Tabela 4.

FENÓTIPOS	GENÓTIPOS
Negros	AABB
Mulatos	AABb, AaBB
Mulatos Médios	AAbb, AaBb, aaBB
Mulatos Claros	Aabb, aaBb
Branco	aabb

**Tabela 4 - Relação Fenótipo X Genótipo (GENÉTICA, 2003, s. Herança Quantitativa, p.1).**

Pode-se, então, observar que o que determina se a cor da pele é mais escura, ou mais clara, é a quantidade de genes dominantes, e não a qualidade dos mesmos. Quanto mais genes dominantes, mais escura é a cor da pele, e quanto menor a quantidade, mais clara o é.

## 2.2 COR

### 2.2.1 História

Há mais de um milhão de anos atrás os ancestrais do homem perderam uma grande quantidade de pelos que cobria a pele, justamente para esfriar o corpo que era , então, submetido a longas caminhadas. Nas partes descobertas, foi necessária, então, a produção de melanina (que torna a pele escura), para absorver os raios ultravioletas do sol e fazer com que os mesmos perdessem energia. Em regiões menos ensolaradas, a pele escura impediria a formação de vitamina D (importante para desenvolvimento ósseo e sistema imunológico). O processo evolutivo baseado em *Seleção Natural* levou, então, à seguinte solução: região menos ensolarada = pele mais clara = absorção maior de raios ultravioleta; região intermediária= bronzamento = protege o *folato* em épocas de sol e produz vitamina D quando o sol não é tão forte (KENSKI, 2003).

Segundo estudos encontrados em 1991 pela antropóloga Nina Joblonski (KENSKI, p. 44, 2003), pessoas de pele clara expostas à luz forte produzem níveis baixos de *folato*. Esta substância é importante, para que, na gravidez as crianças não apresentem problemas de coluna, assim como, mais tarde, para a produção de células do sangue e também de espermatozóides. A cor da pele, aqui, pode ser vista como um moderador de nutrientes.

Na Tabela 5, pode-se observar a localização e distribuição da humanidade, através dos anos:

<b>Idade em relação a data atual (anos)</b>	<b>Localização</b>
200000	África
100000	Ásia, depois Oceania e Europa
15000	América

**Tabela 5 - Localização da humanidade através dos anos (KENSKI, 2003).**

A humanidade, quando submetida às intempéries dos diversos ambientes, desenvolveu características próprias de proteção em cada clima. Por exemplo, pode-se observar, na Tabela 6, características de algumas etnias em relação às diferentes regiões:

<b>Regiões quentes</b>	Tamanho do corpo: é vantajoso ser baixo, como os pigmeus, ou alongados, como os quenianos, com a superfície do corpo grande em relação ao volume, para facilitar a evaporação do suor Cabelo encarapinhado: ajuda a reter o suor no couro cabeludo e resfria a cabeça
<b>Regiões frias</b>	Cabeça e corpo: cabeça e corpo tendem a ser arredondados, para guardar calor; Nariz: pequeno, para não congelar; Narinas estreitas: para aquecer o ar que chega aos pulmões; Olhos: alongados e protegidos do vento, por dobras de pele

**Tabela 6 - Clima *versus* características corporais (KENSKI, 2003).**

Como a “espécie humana“ surgiu na África e com um pequeno número de indivíduos, não houve, nesse meio tempo, um grande número de gerações e, como também ocorreu uma grande miscigenação entre as populações de diferentes lugares, ocorre que, hoje em dia, as pessoas acabam por trazer características individuais de diversas etnias.

Em 1758, o botânico sueco Carolus Linnaeus (KENSKI, 2003), criou um sistema de classificação de todos os seres vivos; chamou os humanos de *Homo sapiens* e os dividiu em quatro subespécies:

Subespécie	Características
Vermelhos americanos	Geniosos, despreocupados e livres
Amarelos asiáticos	Severos e ambiciosos
Negros africanos	Ardilosos e irrefletidos
Branços europeus	Ativos, inteligentes e engenhosos

**Tabela 7 - Subespécies versus Características (KENSKI, 2003).**

Esta subdivisão da raça humana em subespécies e suas características acirrou os embates étnicos e serviu para justificar a escravidão, por diversos anos, embora nunca tenha se comprovado que determinada etnia seja mais inteligente, ou mais capaz, do que outra. Uma pesquisa efetuada em 2002, por sete pesquisadores, dos Estados Unidos, França e Rússia, comparou 377 partes do DNA de 1056 pessoas, de 52 populações, de todos os continentes. A *diferença genética* foi encontrada entre 93 e 95 % dos humanos pertencentes ao mesmo grupo. A *diversidade* entre as populações só foi encontrada entre 3 e 5 % dos indivíduos. Chegou-se à conclusão de que não existem genes exclusivos de uma população, mas também se chegou à conclusão de que a ancestralidade declarada reflete uma diferença genética, mesmo que seja 3 a 5 % da diferença genética que existe entre os humanos (KENSKI, 2003). Segundo PENA (2000), “Cada país tem uma composição genética diferente, que varia de acordo com a história e a interação entre os grupos que para lá migraram”.

Segundo o biólogo Noah Rosemberg (KENSKI, 2003), um dos autores do trabalho citado acima, “Se você permitir que as pessoas declarem múltiplas ancestralidades, terá boas chances de determinar as diferenças genéticas”. E, segundo o geneticista David Goldenstein,

pode-se tentar reagrupar os indivíduos de acordo com semelhanças genéticas. Então, para cada característica, ter-se-ia um novo agrupamento. Ambas são abordagens que podem ser levadas em conta em estudos relacionados a classificação dos seres humanos segundo diferentes características (KENSKI, 2003).

### **2.2.2 Cor e subjetividade**

O debate sobre a definição da etnia de uma determinada população é bem antigo, como pôde ser visto na seção acima, e já serviu como pano de fundo no processo de tomada de decisões em diversas áreas de nosso país. Como exemplo, pode-se citar a atual polêmica sobre cotas raciais para vagas em universidades públicas no estado do Rio de Janeiro e, até mesmo, a reserva de vagas para negros em concursos de juízes. A classificação de cor na sociedade brasileira é bem difícil, mesmo para o etnólogo ou antropólogo, visto que, em geral, a exata classificação dependeria de exames morfológicos, que uma pessoa leiga não poderia executar. Além dessas dificuldades, conforme PETRUCCELLI (2000, p. 9), “a percepção de determinados traços físicos só se constitui como uma cor, e se reveste de significado, em um contexto histórico-cultural”. Como exemplo, pode-se citar o fato de os índios, no censo de 1970, serem enquadrados como pardos (SUBCOMISSÃO DO CENSO DEMOGRÁFICO, 1970).

A mistura entre as raças observadas em nosso país deu origem a uma série de denominações. Algumas delas levam em conta o período histórico, político, cultural e até mesmo regional. Como exemplo, pode-se observar, na Tabela 8:

<b>Etnia</b>	<b>Denominação</b>	<b>Descrição</b>
Branco + Negro	Pardo	Denominação dada desde o início da colonização até os dias atuais. Gradativamente substituído pela denominação “morena”.
	Mulato	Pai branco e mãe negra ou vice-versa.
	Mestiço	No início: europeus com ameríndios.
	Moreno	Os pardos atualmente gostam de se denominar morenos.
Branco + Índio	Caboclo	
Negro + Índio	Cafuzo	

**Tabela 8 - Mistura entre etnias (PETRUCCELLI, 1998; PENA, 2000).**

Em alguns censos, o quesito cor foi mantido como sendo a única indicação disponível sobre as raças que formam a etnia brasileira. Em outros, a decisão baseou-se em fatores puramente políticos. Na Tabela 9, a seguir, pode-se observar a coleta do quesito cor nos respectivos censos e, também, que algumas denominações sofreram mudanças (foram acrescentadas, ou excluídas). No censo de 1890, pode-se observar características sobre a cor do indivíduo (branco e negro) e outras que remetem a uma ascendência racial, como, por exemplo, cabloco para ameríndios e descendentes, e mestiço (no lugar de pardo), para classificar descendentes de negros e brancos (PETRUCCELLI, 2000). Ainda falando sobre as dificuldades relacionadas a este tema, no censo de 1940, segundo CARVALHO (1970), este quesito, apesar de constar dos censos de 1872 e 1892, fora omitido nos de 1900, 1920 e 1930 e manteve-se no de 1940, provavelmente por ser de importância para a Figura política do Brasil no exterior. O quesito amarelo foi utilizado, então, para dar conta da classificação dos imigrantes japoneses. No censo de 1970, volta a ser descartado, depois de diversas reuniões, debates e controvérsias, que envolveram especialistas de diversas áreas (SUBCOMISSÃO DO CENSO DEMOGRÁFICO, 1970).



<b>Censos</b>	<b>Denominações dadas</b>
<b>1872</b>	BRANCO, NEGRO, PARDO E CABOCLO
<b>1890</b>	BRANCO, NEGRO, MESTIÇO E CABOCLO
<b>1900</b>	NÃO FOI COLETADO O QUESITO COR
<b>1920</b>	NÃO FOI COLETADO O QUESITO COR
<b>1930</b>	NÃO FOI COLETADO O QUESITO COR
<b>1940</b>	BRANCO, NEGRO, PARDO E AMARELO
<b>1970</b>	NÃO FOI COLETADO O QUESITO COR
<b>1980</b>	BRANCO, NEGRO, PARDO, AMARELO E ÍNDIO

**Tabela 9 - Denominação das etnias pelos respectivos censos demográficos (PETRUCCELLI, 2000).**

### **2.2.2.1 Quantas denominações de cor?**

É muito interessante a autoclassificação realizada pelos indivíduos de nosso País. Contudo, só algumas têm representatividade, real significação de cor e importância em relação ao número de casos na amostra do censo. Pode-se ver, como exemplo, respostas como *moreno-pálida*, *branquinha*, *sarará*, *negrinha*, *branquela*, *tiziu* e muitas outras. A Tabela completa pode ser vista no Anexo 1, com as 143 autoclassificações de cor encontradas. Esta amostra consta da Pesquisa Mensal de Emprego de Julho de 1998 (PETRUCCELLI, 2000). Tem-se 53 % das denominações só aparecendo uma única vez, outras várias que com certeza foram dadas erradamente, pois indicavam nomes de países ou estados de origem. Uma ampla maioria dessas denominações foi agrupada, sem prejudicar o contexto da pesquisa, em 27 grupos de categorias. Por exemplo: *morena clara*, *morena escura*, *branco-morena* e *branco-morena-clara*, podem ser agrupadas na denominação *morena*. A Tabela completa, com os 27 grupos, pode ser vista no Anexo 2.

### 2.2.2.2 Pergunta aberta por pergunta fechada.

É interessante comparar os resultados das respostas às perguntas abertas, como citado, com as respostas às perguntas fechadas. Como pergunta fechada, entenda-se dar como opção ao indivíduo, no caso da Pesquisa Mensal de Emprego de 1998, as alternativas: branco, preto, pardo, amarelo e indígena. Como exemplo, 90 % dos que se declararam brancos na pergunta fechada se autotransferiram da mesma cor na pergunta aberta. A Tabela 10 contém a distribuição das categorias de cor na pergunta fechada, segundo a pergunta aberta (PETRUCCELLI, 2000).

PERGUNTA ABERTA	PERGUNTA FECHADA					TOTAL
	BRANCA	PRETA	AMARELA	PARDA	INDÍGENA	
AMARELA	0,06	0,03	83,36	0,04	0,19	1,11
BRANCA	91,30	0,65	5,92	1,33	4,08	54,24
BRANCA* <sup>11</sup>	0,06	0,00	0,00	0,01	0,18	0,04
BRASILEIRA	0,19	0,03	0,04	0,02	0,00	0,12
BUGRE	0,00	0,00	0,00	0,01	0,26	0,00
CABOCLA	0,00	0,02	0,00	0,03	0,65	0,02
CABOVERDE	0,00	0,05	0,00	0,03	0,67	0,02
CAFUSO	0,00	0,02	0,00	0,02	0,07	0,01
CANELA	0,00	0,03	0,00	0,02	0,06	0,01
CASTANHO	0,01	0,00	0,00	0,02	0,00	0,01
CLARA	1,15	0,03	0,73	0,32	0,13	0,78
ESCURA	0,00	3,29	0,00	0,21	0,61	0,38
GALEGO	0,01	0,00	0,08	0,01	0,00	0,01
INDÍGENA	0,00	0,00	0,13	0,01	13,92	0,13
JAMBO	0,01	0,01	0,00	0,06	0,14	0,02
LOIRA	0,08	0,02	0,00	0,00	0,00	0,05
MAROMCHOCO	0,00	0,10	0,00	0,05	0,00	0,03
MESTIÇAMISTA	0,02	0,13	0,04	0,16	0,69	0,08
MORENA	4,89	13,97	6,23	54,16	61,94	20,88
MORENACLARA	1,91	0,45	1,85	5,62	7,49	2,92
MORENAESC	0,02	1,83	0,08	0,83	2,12	0,45
MULATA	0,03	2,17	0,00	1,94	1,25	0,81
NEGRA	0,02	31,00	0,04	0,69	1,76	3,14
PARDA	0,18	1,56	0,86	34,07	2,50	10,40
PRETA	0,03	44,52	0,22	0,25	0,80	4,27
SARARA	0,01	0,09	0,23	0,08	0,00	0,04
VERMELHA	0,02	0,00	0,18	0,01	0,50	0,02
TOTAL	100,00	100,00	100,00	100,00	100,00	100,00

**Tabela 10 - Distribuição das categorias de cor por pergunta fechada versus aberta (PETRUCCELLI, 2000, p. 28).**

<sup>11</sup> Qualificação Branca com algum termo complementar.

### 2.2.2.3 Regionalização da Cor

Segundo dados analisados por PETRUCCELLI (2000), na Pesquisa Mensal de Empregos de 1998, deve-se levar em conta o fator regional, quando em se analisar denominações de cor de uma população. Como exemplo, 5 % das pessoas se identificaram como morenas em Porto Alegre, enquanto que em Recife 52 % o fizeram. Na Tabela 11, pode-se ter uma idéia melhor da distribuição da autoclassificação pelas 6 maiores Regiões metropolitanas.

Cor/Resp. Esp.	Recife	Salvador	Belo Horiz.	Rio de Janeiro	São Paulo	Porto Alegre
AMARELA	0,4	0,4	0,2	0,1	2,4	0,1
BRANCA	32,5	19,5	37,0	56,1	61,6	80,1
CLARA	0,2	0,7	3,8	0,5	0,5	0,4
ESCURA	0,0	1,7	0,8	0,6	0,0	0,0
MORENA	52,0	25,7	35,5	14,1	18,4	5,2
MORENACLARA	8,0	3,9	8,8	0,8	2,1	1,4
MORENAESCURA	0,8	1,5	1,8	0,1	0,2	0,0
MULATA	0,2	1,5	0,4	2,0	0,3	0,2
NEGRA	1,6	7,6	4,2	2,8	2,9	1,9
PARDA	1,9	26,7	3,8	16,0	8,6	4,0
PRETA	1,9	9,4	3,3	6,5	2,7	4,7
OUTRAS	0,5	1,4	0,4	0,4	0,4	1,9
TOTAL	100,0	100,0	100,0	100,0	100,0	100,0

**Tabela 11 - Distribuição das categorias mais significativas de pergunta aberta por Região Metropolitana (PETRUCCELLI, 2000, p. 32).**

Pode-se observar, na Tabela 12, a distribuição relativa das denominações intermediárias de cor por Região Metropolitana. Nela, pode-se observar a utilização da denominação parda com maior percentual no Rio de Janeiro e a denominação morena com maior percentual em Recife.

COR/RESP. ESP.	RECIFE	SALVADOR	BELO HORIZ.	RIO DE JANEIRO	SÃO PAULO	PORTO ALEGRE
ESCURA	0,0	2,8	1,6	1,8	0,0	0,1
MORENA	82,6	42,1	69,5	42,0	62,1	48,1
MORENACLARA	12,7	6,4	17,2	2,4	7,1	13,0
MORENAESCURA	1,3	2,5	3,5	0,3	0,7	0,0
MULATA	0,3	2,5	0,8	6,0	1,0	1,9
PARDA	3,1	43,8	7,4	47,6	29,0	37,0
TOTAL	100,0	100,0	100,0	100,0	100,0	100,0

**Tabela 12 - Distribuição das denominações intermediárias de cor (PETRUCCELLI, 2000, p. 33).**

Ainda, falando-se em termos de regionalização, pode-se ver, na Tabela 13, a distribuição de cor segundo as Regiões Metropolitanas.

<b>RM</b>	<b>Branca</b>	<b>Preta</b>	<b>Amarela</b>	<b>Parda</b>	<b>Indígena</b>	<b>Total</b>
Recife	41.83	10.83	1.06	43.94	2.34	100.00
Salvador	21.26	20.62	0.53	56.56	1.03	100.00
Belo Horizonte	48.78	11.59	0.37	37.89	1.36	100.00
Rio de Janeiro	58.63	10.72	0.23	29.65	0.77	100.00
São Paulo	65.43	6.34	2.47	25.19	0.57	100.00
Porto Alegre	83.76	7.56	0.18	7.66	0.85	100.00
<b>TOTAL</b>	<b>54.60</b>	<b>10.90</b>	<b>0.85</b>	<b>32.54</b>	<b>1.10</b>	<b>100.00</b>

**Tabela 13 - Distribuição da população por cor segundo as Regiões Metropolitanas (PETRUCCELLI,2000, p. 54).**

### **2.3 COMENTÁRIOS DO CAPÍTULO**

As principais linhas de pensamento atual, decorrentes de diversos estudos em genética, argumentam que o homem originou-se na África e espalhou-se por outros continentes. Isto põe por terra a teoria da superioridade racial que muitos, através dos séculos, defenderam. Como exemplo de pesquisas nessa área, pode-se relatar o trabalho do pesquisador brasileiro Sergio Danilo Pena (PENA, 2002), que afirmou não haver sentido falar-se em raças humanas, pois biologicamente não existem raças humanas. Segundo PENA (2002), o homem moderno teve sua origem na África, há mais de 100.000 anos atrás. Teve início, então, uma migração progressiva, que fez com que o mesmo ocupasse Ásia, Europa, Oceania e as Américas. Durante todo esse processo, o homem sofreu modificações morfológicas, para se adaptar às diversas condições climáticas e ambientais às quais se sujeitou. Todavia, PENA (2002) afirma, através de estudos de genética molecular moderna, que 90 % da diversidade genômica humana ocorre entre indivíduos e não entre populações de diferentes continentes. Outros estudos nessa área corroboram essa conclusão, tais como o de TEMPLETON (1999) e de FRUDAKIS (2003).

Devido a esses resultados, surge a dúvida de como podemos nos referenciar a determinados grupos da sociedade, por exemplo, aos índios. Uma nomenclatura que vem

sendo muito utilizada é a Etnia, que, segundo PENA (2002), poderia ser definida como grupo populacional que possui características físicas e/ou culturais em comum. Ele discorda da definição dada pelo “Novo Dicionário Aurélio”, 1ª edição, que dá a definição de etnia como “um grupo biológico e culturalmente homogêneo”, pois, segundo o resultado de seu trabalho, chegou-se à conclusão de que não existe, na face da terra, nenhum grupo humano biologicamente (nem culturalmente) homogêneo (PENA, 2002).

Além das dificuldades para a determinação da cor, devido à regionalização, o erro de avaliação, por parte do recenseador, também não pode ser descartado, seja por mau treinamento na tarefa de correta classificação, sobre em qual grupo uma determinada pessoa se enquadra, seja pelo cansaço que pode levar a erros de interpretação. A adequada classificação da cor na sociedade brasileira se torna bastante complexa, ainda, devido à grande miscigenação das etnias que formaram este povo. Sendo assim, exames morfológicos complexos, ou até mesmo exames de DNA poderiam ser utilizados, mas ambos mostram-se, ainda, de custo muito elevado, o que pode ser constatado em FRUDAKIS (2003) o qual cita que um teste para se definir a ancestralidade de uma pessoa custa, em média, US\$ 160.00. Tal custo configura-se bastante elevado, não apenas para a realidade da maior parte da população brasileira, mas também para a viabilização de reserva de cotas étnicas em concursos com grande número de candidatos.

Como visto nas seções acima, problemas para a determinação étnica de um povo vão desde a denominação das categorias de cor, até problemas relacionados à coleta dos dados, interpretação e utilização dessas informações. A Organização das Nações Unidas (ONU), nas recomendações para a realização de censos de população de 1959, “reconhece estar o levantamento de características étnicas, raciais e de nacionalidade sujeito a condições e necessidades nacionais; e, portanto, não recomenda critérios de aceitação universal” (COSTA, 2000). O trabalho de PETRUCCELLI (2000) também corrobora, através da análise dos resultados da pesquisa, a importância de se regionalizar, quando do estudo de um quesito tão importante como a cor de uma população. Por isso, neste trabalho levou-se em consideração uma região específica do estado do Rio de Janeiro, que é a cidade de Niterói.

### **3 CONCEITOS EM REDES NEURAIS ARTIFICIAIS (RNA)**

#### **3.1 APRESENTAÇÃO**

Há muito tempo o homem tenta simular o comportamento do corpo humano, ou de, ao menos, partes do corpo, de uma forma artificial. Seja para substituir partes do corpo lesionadas por um acidente, ou para substituir órgãos que estejam irremediavelmente doentes, ou, porque não, para tentar alcançar, de alguma forma, a imortalidade. Além disso, o homem também busca uma forma mais eficiente de, em determinadas situações, ter ajuda em processos de tomada de decisão. Então, é necessário que tenha ajuda de algo ou alguma ferramenta que possa raciocinar de uma forma inteligente, lógica e coerente.

Com isso, na tentativa de modelar matematicamente o funcionamento do cérebro humano, criaram-se as chamadas Redes Neurais Artificiais (RNA). Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência (ZURADA, 1995).

#### **3.2 NEURÔNIO: BIOLÓGICO E ARTIFICIAL**

O neurônio biológico, como pode ser visto na Figura 2, é constituído de três regiões principais: o soma (corpo da célula), o axônio e os dendritos.

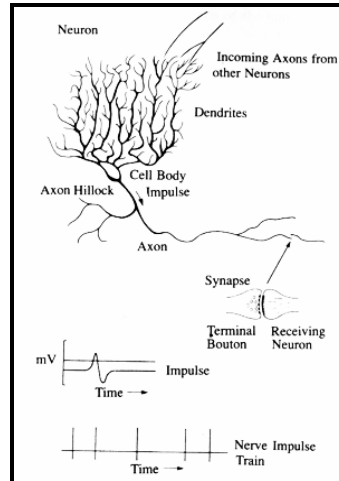


Figura 2- Diagrama esquemático de um neurônio (ZURADA, 1995, p. 28).

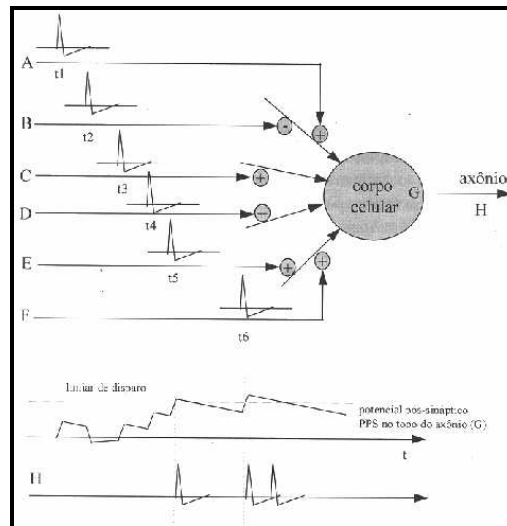
Resumindo o funcionamento do neurônio biológico, temos que sinais elétricos são gerados nos sensores disponíveis em nosso corpo, tais como a retina ocular<sup>12</sup> (aonde a informação visual é transformada em sinais elétricos) e as papilas gustativas<sup>13</sup> (a informação do paladar é transformada em sinais, para ser levada ao cérebro), são processados no soma, e caminham pela parte do neurônio chamada axônio. Antes de o sinal elétrico entrar no próximo neurônio, deve passar por uma sinapse, que é o processo de ligação entre o axônio e o dendrito de outro neurônio. Como um exemplo, de um dos tipos de função neuronal, pode-se citar o fato, de quando o sinal for superior a um certo limite (*threshold*), vai em frente; caso contrário, é bloqueado e não segue. Um neurônio recebe sinais através de inúmeros dendritos. Tais sinais são ponderados e, enviados para o axônio, podendo, ou não, seguir adiante. Na passagem por um neurônio, um sinal pode ser amplificado ou atenuado, dependendo do dendrito de origem, pois a cada condutor está associado um **peso** (*weight*), pelo qual o sinal é multiplicado. Este *peso* pode mudar mediante o treinamento do cérebro, formando o que chamamos de “memória”. Este processo vigora durante toda a vida útil do cérebro e é chamado de **memorização** (KOVÁCS, 1996; ZURADA, 1995).

Na Figura 3, pode-se ver uma representação esquemática da integração temporal de estímulos realizada por um neurônio. Tem-se um neurônio com 6 entradas *A, B, C, D, E* e *F* e uma saída *H*. A entrada “*B*” é inibitória e as demais são excitatórias. Os impulsos nervosos

<sup>12</sup> Membrana sensível do olho e a mais interna, na qual se formam as imagens.

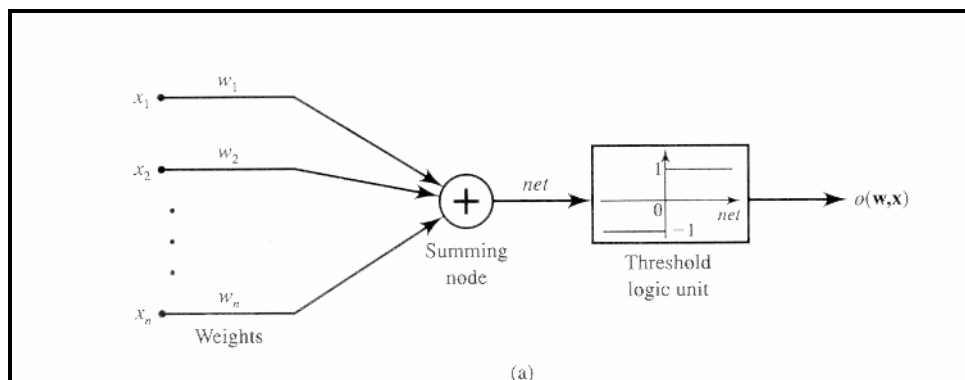
<sup>13</sup> Pequena saliência na mucosa, formada por ramificações nervosas.

chegam nestas entradas nos instantes  $t_1$ ,  $t_2$ ,  $t_3$ ,  $t_4$ ,  $t_5$  e  $t_6$ . Quando a tensão em G passa de um determinado potencial (limiar de disparo), o axônio responde com um potencial na sua saída (KOVÁCS, 1996).



**Figura 3 - Representação esquemática da integração temporal dos estímulos por um neurônio (KOVÁCS, 1996, p. 25).**

O funcionamento de um neurônio artificial pode ser descrito nos seguintes passos: 1) temos que sinais são apresentados à entrada; 2) cada sinal é multiplicado por um número, ou peso, que indica a sua influência na saída da unidade; 3) depois, é feita a soma ponderada dos sinais que produz um nível de atividade; 4) se este nível de atividade exceder um certo limite (*threshold*) a unidade produz uma determinada resposta de saída. O modelo esquemático pode ser visto na Figura 4 (ZURADA, 1995).



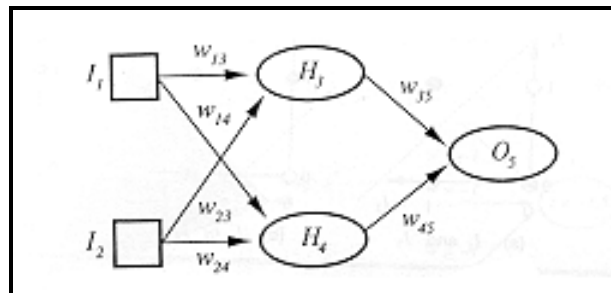
**Figura 4 - Esquema de unidade McCulloch – Pitts (ZURADA, 1995, p. 36).**



No modelo apresentado na Figura 4 (supondo que tenhamos  $p$  sinais de entrada  $X_1, X_2, \dots, X_p$  e pesos  $w_1, w_2, \dots, w_p$  e limitador  $t$ ; com sinais assumindo valores booleanos (0 ou 1) e pesos valores reais) tem-se que o nível de atividade  $a$  é dado por:  $a = w_1X_1 + w_2X_2 + \dots + w_pX_p$ .

Já no caso da Figura 5 abaixo, o nível de atividade  $A_5$  é dado pela equação (1):

$$A_5 = g(W_{3,5}a_3 + W_{4,5}a_4) = g(W_{3,5}g(W_{1,3}a_1 + W_{2,3}a_2) + W_{4,5}g(W_{1,4}a_1 + W_{2,4}a_2)) \quad (1)$$



**Figura 5 - Uma rede *feed-forward* simples com duas entradas, dois nós ocultos, e um nó de saída (RUSSEL & NORVIG, 1995, p. 572).**

Para o processo de memorização, ou treinamento, a maioria dos modelos de redes neurais possui alguma espécie de regra de aprendizado, onde os pesos das conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem através de exemplos.

Arquiteturas neurais são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades da camada posterior. Um exemplo pode ser visto na Figura 6 (ZURADA, 1995).

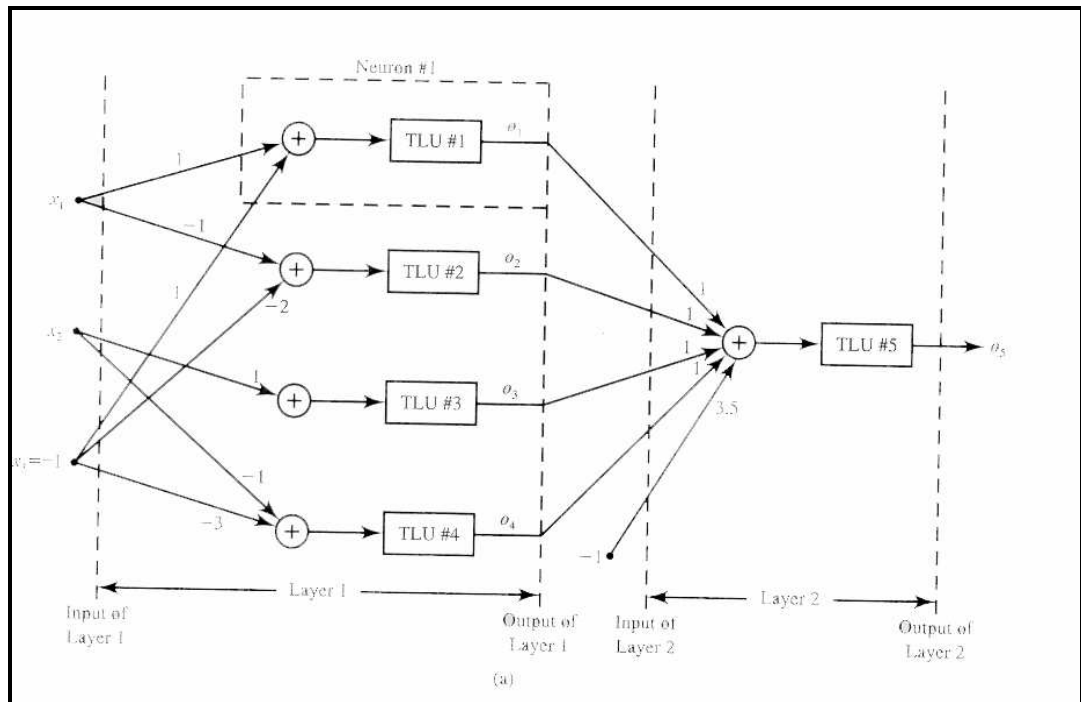


Figura 6 - Rede Neural com duas camadas (ZURADA, 1995, p. 40).

### 3.3 REDES NEURAIIS – COMPONENTES E TIPOS

Uma RNA se caracteriza por dois componentes físicos: **conexões e elementos de processamento** e, por dois componentes não-físicos: **padrões e funções** (ZURADA, 1995).

#### 3.3.1 Padrões, funções, conexões e elementos de processamento

Padrões são dados de entrada da rede. Eles constituem uma codificação que corresponde a certa modelagem da realidade de uma aplicação definida para a RNA. Podem ser valores numéricos ou mesmo caracteres que serão, posteriormente, transformados em números. Na modelagem, a um determinado padrão de entrada corresponde um sinal de saída, sendo que a dimensão, ou o tipo do padrão de entrada, pode ser diferente do padrão de saída (ZURADA, 1995; DEMUTH & BEALE, 1994).

Uma das principais aplicações de RNA's é o reconhecimento (classificação) de padrões. Por exemplo, em reconhecimento de caracteres (escritos) existem tipos diferentes de características envolvidas. Nesta aplicação, o padrão de entrada pode ser uma matriz *booleana* que contém o caractere, e a saída apenas um número que indica se o caractere foi reconhecido, ou não. Assim, temos um padrão de entrada com dimensão diferente do padrão de saída. Além disso, neste caso torna-se essencial um *pré-processamento* de informações. Isto significa que a seleção e a representação correta das características do padrão a ser aplicado, ou seja, da modelagem, pode afetar o desempenho da rede. Criar o melhor conjunto possível de características como padrão é o primeiro passo para o sucesso de qualquer aplicação em RNA's (DEMUTH & BEALE, 1997; MATLAB, 1997).

*Funções de transferência* ou *funções de limiar* (Figura 7) são responsáveis por determinar a forma e a intensidade de alteração dos valores transmitidos de um neurônio a outro. As principais funções de limiar são: *linear*, *hard-limiter* ou *step*, *em rampa*, *sigmoid* e *gaussiana*, e podem, variar, de neurônio para neurônio. (KOVÁCS, 1996; ZURADA, 1995; RUSSEL & NORVIG, 1995).

Em relação às conexões, muitas RNA's têm semelhanças com os grafos (grafos orientados). Um grafo tem arestas (conexões) entre nodos (elementos de processamento) com um só sentido. Isso faz com que a informação flua numa direção definida. Esta informação flui através das arestas e é coletada pelos nodos. RNA's estendem a representação de grafo por incluir um peso em cada aresta (conexão), que modula o sinal de saída, que é passado para o nodo adjacente.

Os pesos são os responsáveis pela memorização do padrão, pois são ajustados durante o processo de aprendizado, que captura as informações. Desta forma, uma conexão define tanto o fluxo quanto a modulação da informação que é passada. Além disso, pesos e conexões positivos (possuem um valor numérico positivo) são excitatórios, sendo os negativos inibitórios (ZURADA, 1995).

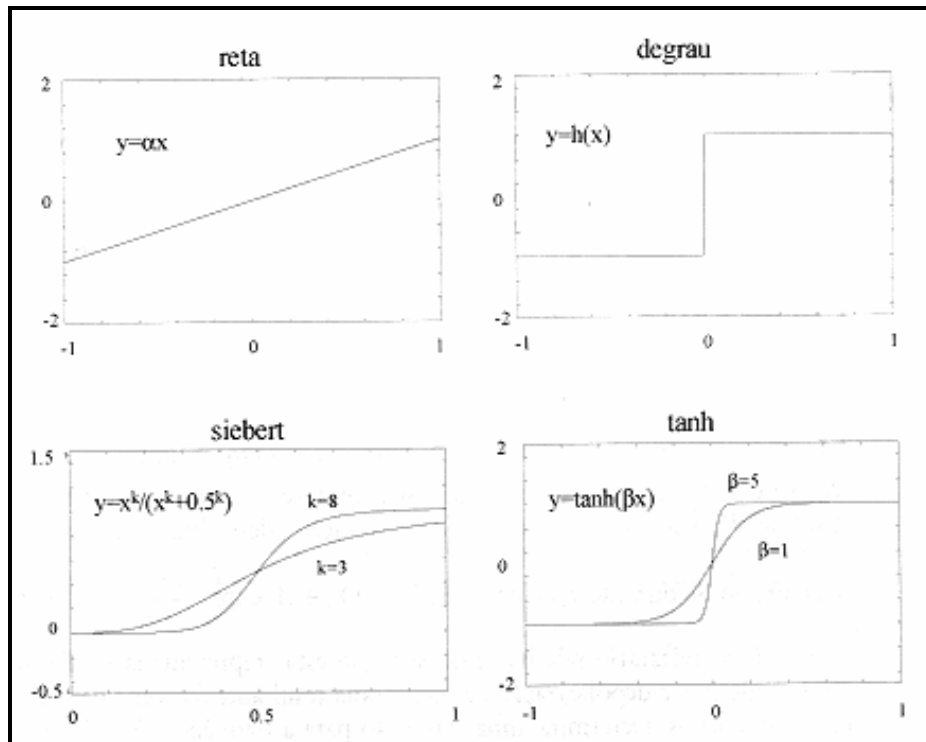


Figura 7 - Características das Funções de Limiar (KOVÁCS, 1996, p. 29).

O neurônio (elemento de processamento) é a parte da RNA onde é realizado todo o processamento. O neurônio da camada de entrada (que é um nodo no dígrafo), recebe apenas um valor do padrão de entrada correspondente, mas possui diversas conexões com os neurônios das camadas seguintes (que podem ter várias entradas) (ZURADA, 1995). Os neurônios nas camadas intermediárias coletam as informações que foram mandadas para eles (cada um deles só necessita de informações locais) e produzem um único valor de saída (que é uma função dos pesos e das entradas).

### 3.3.2 Redes Neurais - Classificação

Pode-se realizar uma distinção entre as redes neurais, através de suas características básicas, como o *tipo de entrada*, a *forma de conexão* e o *tipo de aprendizado*.

Por *tipos de entrada* entendemos a mesma ser binária ou intervalar. Modelos que aceitam entrada binária admitem também discretas, ou seja, somente na forma de 0 e 1. Modelos que permitem entrada intervalar são capazes de aceitar qualquer valor numérico como entrada (forma contínua).

Por *forma de conexão* (entre os elementos processadores) definimos a maneira como os elementos da rede estão conectados e, por conseguinte, como o modelo matemático representa a transmissão dos sinais na rede. Existem três formas de *conexão*: em primeiro lugar, pode-se citar a *alimentação à frente*, onde os sinais de entrada são simplesmente transformados em sinais de saída; outra forma de conexão é a *retro-alimentação*, no qual os sinais ficam sendo alterados em diversas transições de estado, sendo a saída também alimentadora da entrada, como pode ser visto na Figura 8; finalmente, tem-se a *competitiva*, que realiza a interação lateral dos sinais recebidos na entrada, entre os elementos dentro de uma zona de vizinhança (ZURADA, 1995).

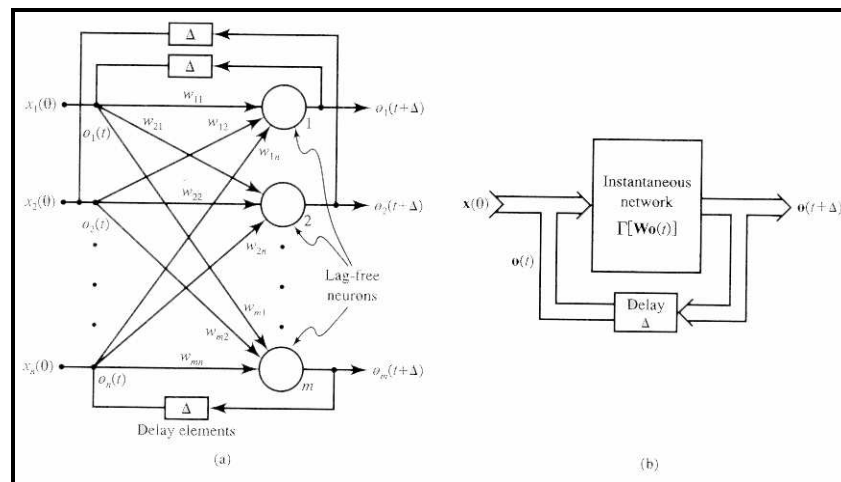


Figura 8 - Conexão com retro-alimentação (ZURADA, 1995, p. 42).

O *tipo de aprendizado* (treinamento) refere-se à existência, ou não, de um sinal de saída pré-definido para a rede. No aprendizado **supervisionado**, há uma definição sobre qual a saída que se deseja para a rede, o que leva a forçar o ajuste dos pesos, de modo a representar o sinal desejado. Por outro lado, há o **auto-aprendizado** (não-supervisionado), que se limita a

fazer uma representação da distribuição de probabilidade dos padrões de entrada na rede. Na Figura 9 pode-se observar a representação desses tipos de aprendizado (ZURADA, 1995).

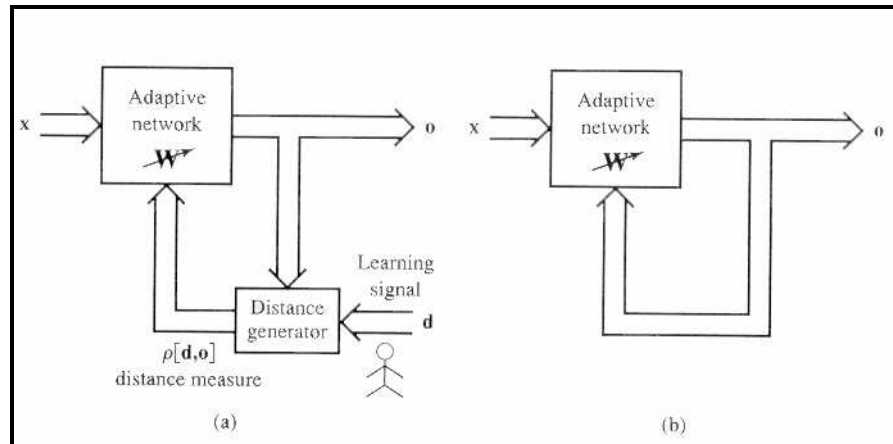


Figura 9 - (a) Supervisionado, (b) Não Supervisionado (ZURADA, 1995, p. 57).

### 3.3.3 Redes Neurais - Estrutura

Pode-se realizar uma distinção entre as redes neurais, através de suas características básicas, como pode ser visto na Tabela 14:

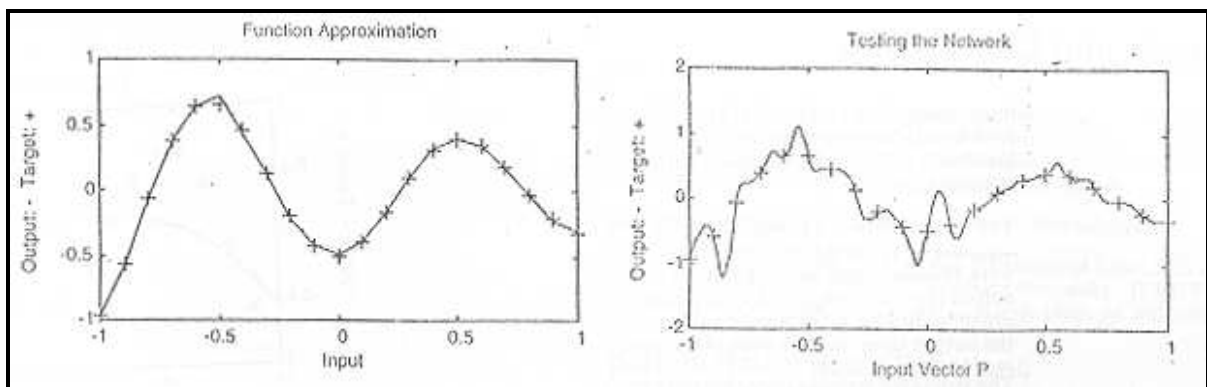
	<i>Feed-Forward</i>	<i>Recurrent</i>
<b>Links</b>	Links unidirecionais; não há ciclos; cada unidade é ligada só com unidades da próxima camada.	Links formam topologias arbitrárias.
<b>Computação</b>	A computação prossegue uniformemente entre as camadas.	Podem se tornar instáveis, oscilar ou exibir comportamento caótico.
<b>Aplicação</b>	Para problemas mais simples ou fazer parte de redes mais complexas. O problema de aprendizado é mais simples. <i>Perceptrons</i> são muito limitados no que podem representar.	Podem implementar problemas mais complexos.
<b>Redes</b>	<i>Perceptron, Backpropagation</i>	<i>Hopfield</i> (memória associativa, conexões bidirecionais)

Tabela 14 - Redes Neurais - Estrutura.

### 3.3.4 Redes Neurais – Parâmetros

A escolha de parâmetros de uma RNA é um quesito muito importante, pois dependendo da escolha do projetista da rede, ela pode não atingir os resultados esperados, ou, até mesmo, levar a resultados errôneos.

Escolhendo-se uma rede muito pequena, o modelo pode ser incapaz de representar a função desejada. Escolhendo-se uma rede muito grande, ela poderá memorizar todos os exemplos e não será capaz de fazer uma boa generalização para novas entradas que não foram vistas ainda. Este efeito é chamado de *overfitting*. E causa a geração de resultados que não são razoáveis apesar do treinamento ser bem simples. Por isso, é muito importante testar a rede com dados diferentes dos que foram utilizados no treinamento. Um exemplo deste efeito pode ser visto na Figura 10.

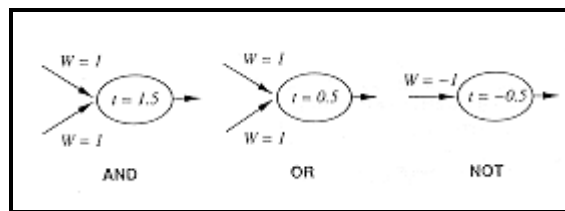


**Figura 10 - Exemplo de saída desejada no lado esquerdo e saída com *overfitting* do lado direito (MATLAB, 1994, p. 36).**

Para melhorar o problema de *overfitting*, pode-se utilizar a técnica de *cross-validation*, que consiste em separar um pequeno conjunto de dados não utilizados no treinamento e testar o resultado da rede para esse conjunto. Este processo pode ser executado repetidamente, após cada novo treinamento, até se ter um resultado satisfatório para a rede (mínimo erro e máxima capacidade de generalização).

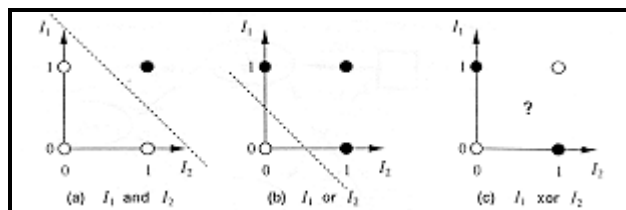
### 3.4 REDES NEURAIIS – O QUE REPRESENTAM

Pode-se comparar o funcionamento de elementos de processamento a portas lógicas. Na verdade, isto foi o que originalmente motivou os estudos dos pesquisadores McCulloch e Pitts, em 1943. Eles tentaram modelar o funcionamento das funções booleanas AND, OR e NOT. Atingiram seu objetivo, através da utilização de pesos e funções de limiares adequados ao problema (Figura 11).



**Figura 11 - Unidades com Função de ativação em degrau podem agir como portas lógicas, com pesos e thresholds apropriados (RUSSEL & NORVIG, 1995, p. 570).**

O *perceptron* só pode representar funções que são linearmente separáveis. Isto é dado pela equação:  $O = W.I$ , onde  $O$  é igual ao produto escalar entre vetor de pesos  $W$  e o vetor de entradas  $I$ . Um *perceptron* dá uma saída 1 somente se  $W.I > 0$ ; o espaço de entrada é dividido em dois pela fronteira de  $W.I = 0$ . A representação gráfica pode ser vista na Figura 12 (RUSSEL & NORVIG, 1995).



**Figura 12 - Separação linear em um perceptron (RUSSEL & NORVIG, 1995, p. 575).**

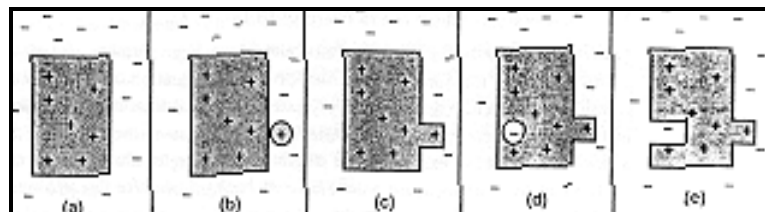


### 3.5 REDES NEURAIIS – ALGORITMO DE APRENDIZADO

A idéia por trás da busca *Current-Best-Hypothesis* (CBH) é manter uma hipótese simples e ajustá-la com a chegada de novos exemplos. Para se ter uma idéia da hipótese, basta fazer um paralelo com a medicina: quando se tem um resultado de laboratório positivo para uma determinada doença e o paciente, na verdade, não está doente. Isto é chamado de falso positivo. No caso contrário, é chamado de falso-negativo. Um exemplo pode ser visto na Figura 13 (RUSSEL & NORVIG, 1995).

Para se resolver esse tipo de problema, os dois métodos abaixo são utilizados:

- Generalização - a hipótese deveria ser negativa, mas é positiva, e deve ser acrescida no conjunto;
- Especialização - a hipótese deveria ser positiva, mas é negativa, e deve ser removida do conjunto.



**Figura 13 - (a) Hipótese consistente. (b) Falso negativo. (c) A hipótese é generalizada. (d) Um falso-positivo. (e) A hipótese é especializada (RUSSEL & NORVIG, 1995, p. 547).**

No caso do *Perceptron*, a hipótese é uma rede definida pelos valores correntes dos pesos. A rede é inicializada com valores randômicos de pesos ou coeficientes. Depois, deve ser atualizada, para ter consistência com os exemplos. Os pesos sofrem pequenos ajustes, para diminuir a diferença da saída real para a saída desejada. O processo de atualização dos pesos é dividido em épocas ou ciclos de treinamento. Cada época envolve o ajuste de todos os pesos, para todos os exemplos. Um exemplo de um algoritmo geral de aprendizado é mostrado na Figura 14 (RUSSEL & NORVIG, 1995).

```

Function NEURAL NETWORK LEARNING(examples) returns network
  network - a network with randomly assigned weights
  repeat
    for each e in examples do
      O - NEURAL NETWORK OUTPUT(network, e)
      T - the observed output values from e
      update the weights in network based on e, O, and T
    end
  until all examples correctly predicted or stopping criterion is reached
  Return network

```

Figura 14 - Método geral de aprendizado de uma rede neural (RUSSEL & NORVIG, 1995, p. 577).

O erro é dado por:  $Err = T - O$ , onde  $T$  é a saída correta. Se o erro é positivo, deve-se aumentar  $O$ ; se o erro é negativo, deve-se diminuir  $O$ . Cada unidade de entrada contribui com  $W_j I_j$  para a entrada total; então, se  $I_j$  é positivo, um aumento em  $W_j$  tenderá a aumentar  $O$ ; se  $I_j$  for negativo, um aumento em  $W_j$  tenderá a decrescer  $O$ . Temos, então, a seguinte regra:  $W_j - W_j + \alpha * I_j * Err$ , onde  $\alpha$  é chamado de *learning rate*.

O *perceptron* faz busca por *gradient descent* (ou *Hill-climbing*) no espaço de pesos. Ele sempre tenta fazer mudanças que melhorem o estado atual. Contudo, com o *Hill-climbing* podem ocorrer os problemas característicos desse algoritmo: *mínimo local*, *platô* e a *cadeia de montanhas*. Cada um deles pode interromper ou até mesmo atrasar o processo de encontrar a melhor solução.

Através da utilização do *simulated annealing*, pode-se contornar um pouco este problema, pois é característica desta técnica a possibilidade de se ir para outros pontos do espaço de busca, partindo de outro ponto mais afastado da última solução encontrada (RUSSEL & NORVIG, 1995).

## 3.6 REDES NEURAIIS – MODELOS

### 3.6.1 Perceptron / Adaline

Possui entrada intervalar, aprendizado supervisionado e propagação para frente. Utiliza o *Combinador Adaptativo Linear*, onde a saída de um elemento processador (**ep**) é a combinação linear das entradas. Estas são um vetor que é multiplicado por pesos pré-existentes no **ep**, assumindo a forma.

$$S(t) = \sum_{i=0}^{n-1} p_i(t)e_i(t) \quad (2)$$

onde  $p_i$  é um elemento do vetor de pesos,  $e_i$  é um elemento do vetor de entradas,  $S$  a saída do **ep** e  $t$  o número da amostra de entrada (KOVÁCS, 1996; ZURADA, 1995).

### 3.6.2 Backpropagation

É um modelo com entrada intervalar, aprendizado supervisionado e com propagação para frente.

O *backpropagation* deriva-se do modelo perceptron/*adaline*. Seus neurônios são compostos por três ou mais camadas de *adalines* interconectados. Estes *adalines* têm uma diferença fundamental: eles utilizam uma função do tipo *sigmóide* como função de limiar (KOVÁCS, 1996; ZURADA, 1995; SANTOS, 2001). A função de limiar é do tipo *sigmóide*, tendo em vista ser necessária uma função não-linear, para ampliar o potencial de classificação de um modelo.

Neste modelo, o erro obtido na saída é transferido para as camadas intermediárias. Daí o nome *retropropagação (backpropagation)*. Isso se dá pela necessidade de ajuste dos

neurônios que não têm contato com a saída, necessitando, assim, de algum parâmetro para atualização dos pesos/coeficientes.

### 3.7 COMITÊS DE REDES NEURAIS

Na busca de um melhor resultado de uma RNA, muitas vezes utilizam-se vários esquemas de classificadores. Um desses esquemas pode obter o melhor resultado para um determinado conjunto de padrões, e o mesmo raciocínio vale para todos os outros classificadores, com relação a diferentes conjuntos de padrões. Logo, se existisse uma forma de se utilizar os bons resultados obtidos pelos classificadores, individualmente, em um resultado final, poder-se-ia chegar ao melhor resultado para a solução de um determinado problema de classificação.

Com isso, demonstra-se a importância de se utilizar métodos de combinação de classificadores, onde temos classificadores bem especializados, em diferentes partes do espaço de entrada, mas que apresentam erros em partes distintas, como ilustrado na Figura 15.

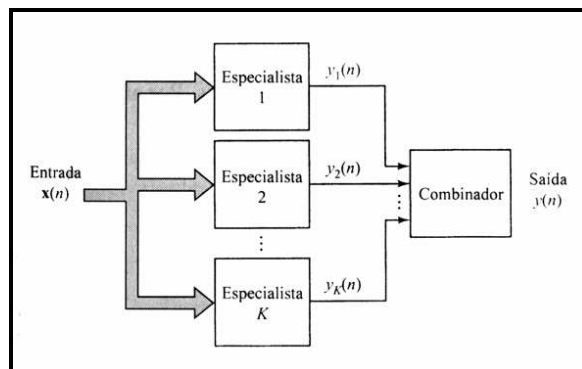


Figura 15 - Máquina de Comitê (HAYKIN, 2001, p. 387).

### 3.7.1 Níveis de classificação

Pode-se separar os esquemas de combinação de classificadores em três níveis: nível abstrato, nível de ranking e nível de medida. No nível abstrato, cada classificador  $e_k$  produz um único rótulo:  $e_k(x) = j_k$ , com  $J_k \in L \cup \{M + 1\}$ , como resposta de análise do padrão  $x$ . No nível de ranking, como resposta da análise de um padrão  $x$ , cada  $e_k$  produz uma lista de rótulos  $L_k \in L$ , onde todos os rótulos de  $L_k$  são ordenados em uma espécie de ranking. No nível de medida, como resposta da análise do padrão  $x$ , cada  $e_k$  produz um vetor real  $M_k(x) = [m_k(1), \dots, m_k(M)]t$ , onde  $m_k(i)$  é um tipo de medida do quão bem  $x$  recai sobre a classe  $i$  (SANTOS, 2001).

### 3.7.2 Métodos de combinação

Existem diversos métodos de combinação de classificadores; todavia, como não é escopo deste trabalho discorrer sobre cada um deles, será abordado o tipo escolhido, o qual será utilizado nos testes posteriores. A escolha do método de combinação se baseou no trabalho de SANTOS (2001), que chegou à conclusão que o *Método de Combinação por Média* (nível de medida) é o que apresenta o melhor resultado para uma determinada faixa de problemas de classificação.

A saída do classificador integrado é dada pela média das saídas dos classificadores envolvidos. Formalmente, pode-se definir assim:  $d_i^n(x)$ , onde  $i$  indica a classe à qual  $d$  se relaciona;  $n$  indica qual o classificador considerado e  $x$  indica um padrão genérico. A função discriminante do comitê fica:  $d_i^{\text{comitê}}(x) = \frac{1}{k} \sum_1^k d_i^j(x)$ ,  $i = \{1, 2, \dots, M\}$ , onde  $M$  é o número de classes e  $K$  o número de classificadores.

### 3.7.3 Formação de comitês

Um sistema com múltiplas redes pode ser *modular* ou *comissionado*. A denominação chamada de comissionada é utilizada quando as redes são redundantes entre si, mas juntas melhoram o resultado final. A denominação modular é utilizada quando as redes não são redundantes e utilizam o conceito “dividir para conquistar” (SANTOS, 2001; HAYKIN, 2001). Ambas não são, necessariamente, mutuamente exclusivas, podendo haver sistemas com múltiplas redes tendo comitês e sub-sistemas modulares (SHARKEY, 1999). Como visto em SANTOS (2001), as redes utilizadas neste trabalho são redundantes, mas foram combinadas na tentativa de se obter um melhor desempenho na classificação.

Dentro das categorias existentes para formação dos membros do comitê, apenas duas serão utilizadas neste trabalho. São elas: variação da arquitetura da rede e variação do conjunto de treinamento. Na categoria de variação do conjunto de treinamento, foi utilizado o método chamado RDP (*Replicação Dirigida de Padrões*), que objetiva criar redes especializadas por classes. Para especializar uma rede na classe  $k$ , basta replicar, no conjunto de treinamento original, os  $n$  padrões que representam a classe em questão por um fator inteiro determinado que seja maior do que 1.

Exemplificando e adaptando a Tabela de treinamento em relação à tipagem sangüínea, temos abaixo a Tabela 15 com o conjunto de treinamento original.

<i>A</i>	<i>B</i>	<i>AB</i>	<i>D</i>	<i>C</i>	<i>RA</i>	<i>RB</i>	<i>CLASSE</i>
0	0	0	3	0	0	3	1
0	0	0	4	0	4	4	1
4	0	4	4	0	0	4	2
4	0	4	4	0	0	4	2
0	4	4	4	0	3	0	3
0	0	0	4	0	4	4	1

**Tabela 15 - Conjunto de treinamento original onde A, B, AB, D, C, RA e RB são atributos e CLASSE representa as três classes existentes.**

Usando-se RDP com fator 2, teremos três novos conjuntos de treinamento que servirão, no futuro, para treinamento de cada membro de um comitê de RNA's (Seção 3.7). Estes novos conjuntos podem ser vistos nas Tabelas 16, 17 e 18.

<i>A</i>	<i>B</i>	<i>AB</i>	<i>D</i>	<i>C</i>	<i>RA</i>	<i>RB</i>	<i>CLASSE</i>
0	0	0	3	0	0	3	1
0	0	0	4	0	4	4	1
0	0	0	3	0	0	3	1
0	0	0	4	0	4	4	1
4	0	4	4	0	0	4	2
4	0	4	4	0	0	4	2
0	4	4	4	0	3	0	3
0	0	0	4	0	4	4	1
0	0	0	4	0	4	4	1

**Tabela 16 - Conjunto de treinamento especializado na Classe 1. *A, B, AB, D, C, RA* e *RB* são atributos.**

<i>A</i>	<i>B</i>	<i>AB</i>	<i>D</i>	<i>C</i>	<i>RA</i>	<i>RB</i>	<i>CLASSE</i>
0	0	0	3	0	0	3	1
0	0	0	4	0	4	4	1
4	0	4	4	0	0	4	2
4	0	4	4	0	0	4	2
4	0	4	4	0	0	4	2
4	0	4	4	0	0	4	2
0	4	4	4	0	3	0	3
0	0	0	4	0	4	4	1

**Tabela 17 - Conjunto de treinamento especializado na Classe 2. *A, B, AB, D, C, RA* e *RB* são atributos.**

<i>A</i>	<i>B</i>	<i>AB</i>	<i>D</i>	<i>C</i>	<i>RA</i>	<i>RB</i>	<i>CLASSE</i>
0	0	0	3	0	0	3	1
0	0	0	4	0	4	4	1
4	0	4	4	0	0	4	2
4	0	4	4	0	0	4	2
0	4	4	4	0	3	0	3
0	4	4	4	0	3	0	3
0	0	0	4	0	4	4	1

**Tabela 18 - Conjunto de treinamento especializado na Classe 3. A, B, AB, D, C, RA e RB são atributos.**

### 3.7.4 Comentários do capítulo

Pôde-se notar, pela teoria demonstrada nas seções deste capítulo, os inúmeros parâmetros e quesitos que devem ser levados em conta para a utilização das RNA's. Este trabalho se aterá à utilização de classificadores supervisionados, que necessitam de uma seleção criteriosa dos dados que formarão o conjunto de treinamento e, também, de um ajuste nos parâmetros que cada rede utilizará. A arquitetura final das MLP's utilizadas resultará de dois aspectos: número de componentes de entrada e disposição dos processadores nas camadas da rede.

Com o intuito de tentar aprimorar resultados e, até mesmo, viabilizar soluções aonde não existiam, é que foi introduzido o conceito de Comitês de Redes Neurais (Seção 3.7). Existem diversos métodos para a formação de comitês, mas não é objetivo deste trabalho se ater a isso. A técnica utilizada para a formação dos membros do comitê será a variação do conjunto de treinamento (RDP). Será feito um experimento, a fim de se avaliar a utilização desta técnica.

Pode-se discorrer sobre inúmeras possibilidades, que se abrem com a utilização das Redes Neurais Artificiais; como exemplo, pode-se citar aplicações na área de visão computacional, tais como as mencionadas em Santos (2001). Na área de reconhecimento da fala, existem aplicações que permitem realizar o controle biométrico, para autenticação de



acesso a sistemas computacionais, ou, até mesmo, em espaços físicos; verificação de autoria de gravação, para elucidação de crimes; entrada de texto, por intermédio da voz, e tantas outras. Diversos trabalhos foram realizados nesta área, dentre eles, os realizados por CARICATTI & WEIGANG (2001), e TAFNER (1996).

Diversas outras aplicações podem ser mencionadas, tais como: avaliação de riscos de investimento, crédito financeiro, diagnósticos os mais diversos na área médica, previsão de crises econômicas e energéticas.

## 4 DESCRIÇÃO DOS EXPERIMENTOS

### 4.1 INTRODUÇÃO

Nesta dissertação é tratada uma aplicação que tem por meta testar a capacidade de uma RNA em prever a etnia simplificada de um paciente, fornecendo-se para a mesma as características do sangue colhido, utilizando-se dos conceitos definidos no Capítulo 3.

Utilizou-se uma plataforma computacional, a saber: Microcomputador com processador Pentium III de 700 Mhz (Intel), 256 Mb de RAM e sistema operacional Windows XP (Microsoft). Além disso, vale ressaltar que todas as rotinas que foram desenvolvidas para este trabalho foram escritas no MATLAB 6.0, da *Mathworks* (MATLAB, 1997).

A análise baseia-se em uma amostra aleatória de dados colhidos no Banco de Sangue do Hospital Universitário Antonio Pedro – Universidade Federal Fluminense – Niterói, Estado do Rio de Janeiro, nos anos de 2000 e 2001. O método de *randomização* utilizado foi a utilização de dados de pacientes cujo último sobrenome começasse pela letra “A”. Foram colhidos 950 registros de pacientes que doaram sangue nos anos de 2000 a 2001. Apenas uma amostra aleatória de cada indivíduo foi selecionada.

A classificação utilizada para os tipos sangüíneos, segundo os grupos *ABO* e *Rh*, foi a classificação já descrita na literatura e universalmente aceita, cujas características estão discriminadas na Tabela 19.

REAGENTE GRUPO SANGUINEO	A	B	AB	D	C	RA	RB
A	+	-	+	+ ou -	+ ou -	+ ou -	+ ou -
B	-	+	+	+ ou -	+ ou -	+ ou -	+ ou -
AB	+	+	+	+ ou -	+ ou -	+ ou -	+ ou -
O	-	-	-	+ ou -	+ ou -	+ ou -	+ ou -

Tabela 19 - Reatividade dos grupos sangüíneos, segundo reagentes.

Na Tabela 20 observa-se o significado da notação utilizada na Tabela 19, acima.

- (negativo) significa ausência de reatividade para o reagente; + (positivo) significa presença de reatividade para o reagente, podendo ser expresso em número de 1 a 4, com os valores atribuídos aos números, segundo o maior, ou menor, grau de reatividade.	0 ou negativo – não reator 1 – pouco reator 2 – moderado reator 3 – forte reator 4 – muito forte reator
--	---

Tabela 20 - Notação.

O sangue dos doadores foi classificado através de prova direta com os reagentes *anti-A*, *anti-B*, *anti-AB*, *anti-D* (fator *Rh*), e através de prova reversa com os reagentes *RA* e *RB* (*reverso de A* e *reverso de B*). Os grupos sangüíneos podem ser classificados segundo tipo *ABO* (seção 2.1.2) ou *fator Rh*. Serão utilizados, para os tipos *ABO*, os subtipos *A*, *B*, *AB* e *O* e para o fator *Rh* os subtipos positivo e negativo. Os reagentes *A*, *B* e *AB* são suficientes para uma eficaz classificação dos tipos *ABO* e o reagente *D* para o fator *Rh*. O ***Rh*** + reage com *D* e o ***Rh*** – não reage (valor 0). Em relação a **etnia**, será utilizada a classificação *fenotípica* simplificada, segundo o senso comum pois, tratando-se de um trabalho retrospectivo, a etnia foi definida pelo técnico do laboratório, no momento da coleta do sangue. A etnia simplificada foi classificada como: **branca, parda e negra**.

Será pesquisada a possibilidade de haver influência da reatividade com os reagentes *A*, *B*, *AB*, *C*, *D*, *RA* e *RB* na determinação da etnia de um indivíduo. O estudo se dará através da

utilização de Redes Neurais Artificiais (RNAs). Para tal, os dados serão testados nas seguintes etapas: etapa de treinamento da rede; etapa de aferição da rede e etapa de correlação de dados com a etnia simplificada.

Para a classificação dos dados analisados utilizaram-se técnicas de redes neurais do tipo “Multilayer Perceptron” (MLP), contendo uma camada escondida de neurônios (HAYKIN, 2001); o algoritmo de aprendizado utilizado foi o de “backpropagation error”. O número de neurônios na camada de entrada variou com o tipo de experimento que foi realizado. Em geral, foi equivalente ao número de atributos relevantes no domínio do problema. Na camada de saída, teve-se, em geral, um número igual ao número de classes que precisavam ser classificadas.

Para a determinação do número de processadores na camada escondida, foi utilizada a métrica de Hetch-Nielsen (1990) que diz que qualquer função pode ser representada por  $2n + 1$  funções de uma variável. Então, se tivéssemos 5 entradas ter-se-ia:  $N_{hidden} \leq 2N_{in} + 1$ , onde  $N_{hidden}$  é o número de processadores na camada escondida e  $N_{in}$  é o número de entradas (HAYKIN, 2001).

Para a determinação do número de épocas a ser utilizada pela rede, foi utilizada a técnica de *validação cruzada múltipla* (HAYKIN, 2001), que consiste em dividir o conjunto de dados de  $N$  exemplos em  $K$  subconjuntos, sendo  $K > 1$ . O modelo é treinado com todos os subconjuntos, exceto um, e o erro de validação é medido testando com este subconjunto, deixado de lado no treinamento. Este procedimento foi realizado para um total de  $K$  tentativas, cada uma delas usando um subconjunto diferente para a validação. O desempenho do modelo é avaliado pela média do erro quadrado obtido na validação sobre todas as tentativas do experimento. A técnica utilizada aqui é conhecida como *validação cruzada múltipla do método deixe um de fora*. Além disso, foi utilizada uma taxa de aprendizado adaptativa, iniciando em 0,2 para todos os experimentos.

Na modelagem da rede, inicialmente foi realizado um pré-processamento no conjunto de dados, a fim de eliminar valores incorretos, dados ausentes ou inconsistentes. Depois, continuando com o pré-processamento, os dados foram todos normalizados entre 0 e 1.

Com o objetivo de se ter um melhor resultado na classificação dos padrões utilizados, em alguns experimentos foi utilizada uma combinação de classificadores (comitê de redes) no nível abstrato que é a de votação por média (seções 3.7.2 e 3.7.3). Neste tipo de combinação, a saída do classificador integrado é dada pela média das saídas equivalentes em seus membros. É importante ressaltar que, para se utilizar comitês de redes neurais, é necessário que os classificadores utilizados sejam bem acurados individualmente (SANTOS, 2001).

No caso dos experimentos que utilizaram comitês, as redes utilizadas foram redundantes entre si, mas, quando combinadas, deveriam apresentar uma melhora no desempenho global do sistema. Cada rede foi conseguida através da variação do conjunto de treinamento, utilizando-se a técnica de *Replicação Dirigida de Padrões* (RDP), como visto na seção 3.7.3, visando a criação de redes especializadas por classe.

Para medir os desempenhos dos experimentos, foi utilizada a técnica de construção de matrizes de confusão. Ela é definida como: sendo a matriz  $\phi_k = (n_{ij})_k$  a matriz de confusão para o classificador  $e_k$ , quadrada e com dimensões número de classes  $\times$  número de classes. Nesta matriz, se  $i = j$ ,  $(n_{ij})_k$  indica o número de padrões da classe  $i$  que foram corretamente classificados por  $e_k$ . Se, por outro lado,  $i \neq j$ ,  $(n_{ij})_k$  indica o número de padrões da classe  $i$  que foram erroneamente classificados como sendo da classe  $j$ . Como exemplo, na Tabela 21, pode-se observar 70 % dos casos de classe 1 (linha 2 da Tabela) classificados pela RNA como sendo realmente da classe 1 e 30 % dos que deveriam ser da classe 1 sendo classificados erroneamente como sendo da classe 2. Em relação à classe 2 (linha 3 da Tabela), temos 10 % classificados pela RNA, erroneamente, como sendo da classe 1 e 90 % classificados, corretamente, como sendo da classe 2.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	70.0000	30.0000
<i>Classe 2</i>	10.0000	90.0000

**Tabela 21 - Exemplo de Matriz de Confusão Percentual.**

Além da matriz de confusão percentual, utilizou-se, como parâmetro para se verificar a eficácia da RNA, o percentual de reconhecimento por conjunto e, também, o reconhecimento médio por classe. Ambos os percentuais foram verificados nos conjuntos de treinamento e

teste. Também será documentado, para cada experimento, o vetor de competência, onde se pode observar a especialização de cada classe (quanto mais próxima de 1 em relação a todas as classes, melhor será a generalização da RNA) e também será mostrada a curva conceitual do aprendizado no período de treinamento. Ter-se-ia um bom resultado se a curva resultante não fosse muito oscilante, nem erros finais considerados altos para a aplicação em questão. O ideal é se ter uma convergência para um mínimo local no menor número de épocas possível (SANTOS, 2001).

A fase de teste da rede teve início realizando-se o treinamento da rede com um conjunto de dados não estratificados, cinco entradas, três neurônios na camada oculta e quatro na de saída. Este passo visou o teste de consistência da base de dados, onde se verificou a correta classificação da tipagem sanguínea.

## **4.2 EXPERIMENTOS**

Nesta seção encontram-se os diversos experimentos realizados. Em geral, utilizou-se uma taxa de 1000 épocas de treinamento, taxa de aprendizado de 0,2 e um total de 10 interações para todos os experimentos. Nos diversos experimentos foi variada a arquitetura da rede, em relação ao número de neurônios na camada de entrada, oculta e de saída. Também se trabalhou com variações no conjunto de dados utilizados para treinamento e teste.

### **4.2.1 Experimento A: Validação da Rede pela Tipagem Sanguínea**

Este grupo de experimentos foi utilizado para aferir a estabilidade e robustez da base de dados. Verifica a eficácia na classificação da tipagem sanguínea.

#### 4.2.1.1 Experimento A.1

No Experimento *A.1* foi utilizada a base de dados referente a tipagem sanguínea classificada nos tipos A, B, AB e O. A arquitetura da RNA utilizada neste experimento foi composta por dados de entrada não estratificados com 246 padrões do tipo sanguíneo A, 61 do tipo B, 136 do tipo AB e 241 do tipo O, dando um total de 684 padrões para treinamento. Os dados de saída também não eram estratificados com um total de 215 padrões sendo representados por 4 bits. Dos 7 reagentes originais, 2 (reagente do *fator Rh* e reagente C) foram excluídos no processo de pré-processamento da base de dados, pois não influenciavam na determinação da tipagem sanguínea. Com isso, 5 neurônios foram utilizados na camada de entrada. Na camada oculta, utilizando-se conceitos vistos na Seção 4.1 e, com o objetivo de se variar a arquitetura da rede, para se observar a performance da mesma, tomou-se como ponto de início utilizar 3 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída foram utilizados 4 neurônios, também com função *logsig*.

Na Tabela 22 pode-se observar a matriz de confusão percentual (Seção 4.1) do conjunto de treinamento. Na Tabela 23, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 24, observa-se o percentual de reconhecimento por conjunto de treinamento e teste e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 25 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 16 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	70.0000	9.5122	10.4065	10.0813
<i>Classe 2</i>	1.1475	78.8525	19.8361	0.1639
<i>Classe 3</i>	0	0	99.9265	0.0735
<i>Classe 4</i>	0.4149	0.0415	9.6266	89.9170

Tabela 22 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.1.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	69.9115	9.6460	10.4425	10.0000
<i>Classe 2</i>	0	80.0000	20.0000	0
<i>Classe 3</i>	0	0	100.0000	0
<i>Classe 4</i>	0.4444	0	9.5556	90.0000

Tabela 23 - Matriz de confusão percentual do conjunto de teste do Experimento A.1.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	83.7573
<i>Por conjunto de teste</i>	81.9070
<i>Médio por classe no conjunto de treinamento</i>	84.6740
<i>Médio por classe no conjunto de teste</i>	84.9779

Tabela 24 - Percentuais de reconhecimento do Experimento A.1.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
0.6972	0.7093	0.8798	0.8477

Tabela 25 - Vetor de competência do Experimento A.1.

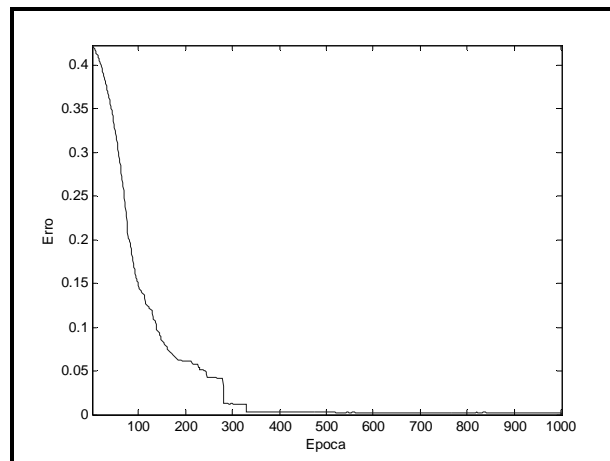


Figura 16 - Curva conceitual de aprendizado do Experimento A.1.



#### 4.2.1.2 Experimento A.2

O Experimento A.2 apresenta a arquitetura da RNA análoga ao experimento anterior, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 5. Isto foi feito para se verificar o efeito na performance da rede. Na Tabela 26 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 27, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 28, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 29 demonstra o nível de especialização, através do vetor de competência, no conjunto de treinamento e, finalmente na Figura 17 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	100.0000	0	0	0
<i>Classe 2</i>	1.6393	88.5246	9.6721	0.1639
<i>Classe 3</i>	0	9.9265	90.0000	0.0735
<i>Classe 4</i>	0	0.0415	0	99.9585

Tabela 26 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.2.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	100	0	0	0
<i>Classe 2</i>	0	90	10	0
<i>Classe 3</i>	0	10	90	0
<i>Classe 4</i>	0	0	0	100

Tabela 27 - Matriz de confusão percentual do conjunto de teste do Experimento A.2.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	96.9737
<i>Por conjunto de teste</i>	97.3488
<i>Médio por classe no conjunto de treinamento</i>	94.6208
<i>Médio por classe no conjunto de teste</i>	95

Tabela 28 - Percentuais de reconhecimento do Experimento A.2.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
0.9960	0.8159	0.8697	0.9988

Tabela 29 - Vetor de competência

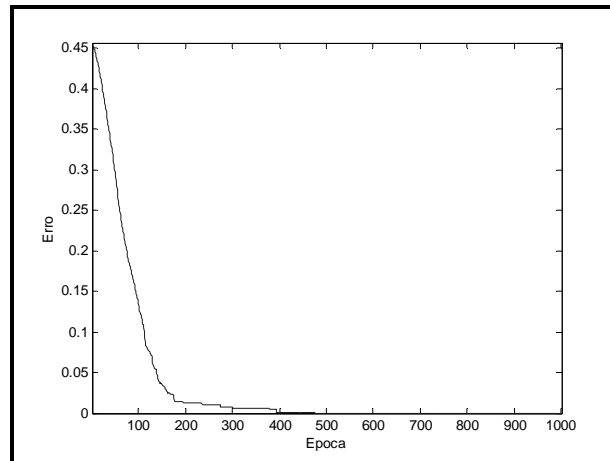


Figura 17 - Curva conceitual de aprendizado do Experimento A.2.

#### 4.2.1.3 Experimento A.3

O Experimento A.3 apresenta a arquitetura da RNA análoga ao Experimento A.1, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 7. Na Tabela 30 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 31, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 32 observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 33 demonstra o nível de especialização, através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 18 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	100.0000	0	0	0
<i>Classe 2</i>	1.8033	88.5246	9.5082	0.1639
<i>Classe 3</i>	0	0	100.0000	0
<i>Classe 4</i>	0	0	0	100.0000

Tabela 30 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.3.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	100	0	0	0
<i>Classe 2</i>	0	90	10	0
<i>Classe 3</i>	0	0	100	0
<i>Classe 4</i>	0	0	0	100

Tabela 31 - Matriz de confusão percentual do conjunto de teste do Experimento A.3.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	98.9766
<i>Por conjunto de teste</i>	99.9070
<i>Médio por classe no conjunto de treinamento</i>	97.1311
<i>Médio por classe no conjunto de teste</i>	97.5000

Tabela 32 - Percentuais de reconhecimento do Experimento A.3.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
0.9955	0.8852	0.9701	0.9996

Tabela 33 - Vetor de competência do Experimento A.3.

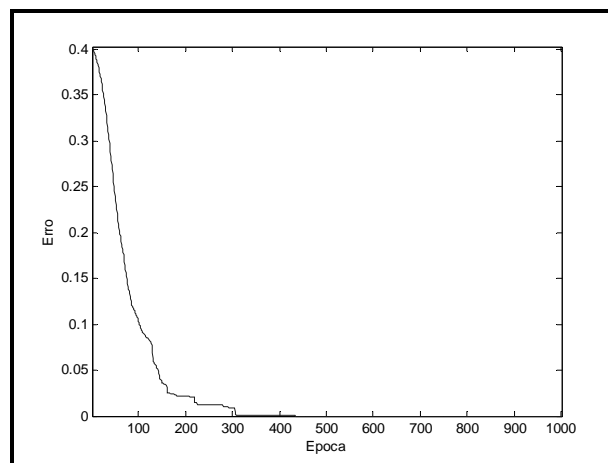


Figura 18 - Curva conceitual de aprendizado do Experimento A.3.

#### 4.2.1.4 Experimento A.4

O Experimento A.4 apresenta a arquitetura da RNA análoga ao Experimento A.1, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 10. Na Tabela 34 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 35, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 36, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 37 demonstra o nível de especialização, através do vetor de competência no conjunto de treinamento e, finalmente, na Figura 19 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	100.0000	0	0	0
<i>Classe 2</i>	1.6393	88.5246	1.3115	8.5246
<i>Classe 3</i>	0	0	100.0000	0
<i>Classe 4</i>	0	0	0	100.0000

Tabela 34 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.4.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	99.9115	0	0.0885	0
<i>Classe 2</i>	0	90.0000	0	10.0000
<i>Classe 3</i>	0	0	100.0000	0
<i>Classe 4</i>	0	0	0	100.0000

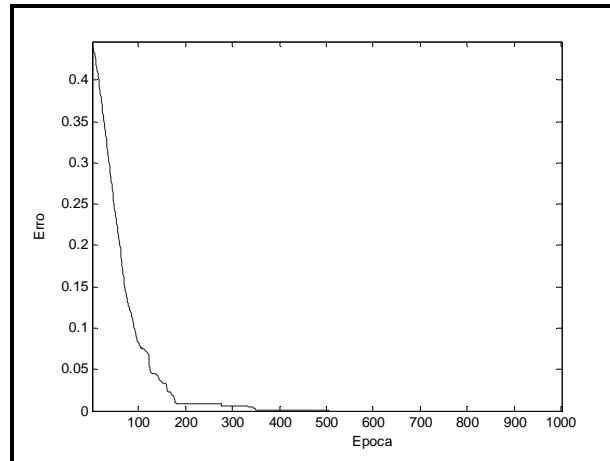
Tabela 35 - Matriz de confusão percentual do conjunto de teste do Experimento A.4.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	98.9766
<i>Por conjunto de teste</i>	99.8605
<i>Médio por classe no conjunto de treinamento</i>	97.1311
<i>Médio por classe no conjunto de teste</i>	97.4779

Tabela 36 - Percentuais de reconhecimento do Experimento A.4.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
0.9960	0.8852	0.9944	0.9823

**Tabela 37 - Vetor de competência do Experimento A.4.**



**Figura 19 - Curva conceitual de aprendizado do Experimento A.4.**

#### **4.2.1.5 Experimento A.5**

O Experimento A.5 apresenta a arquitetura da RNA análoga ao Experimento A.2, exceto por apresentar a função de transferência *tansig* na camada oculta. Isto também foi feito para se verificar o efeito na performance da rede. Na Tabela 38, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 39, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 40, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 41 demonstra o nível de especialização, através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 20 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	100.0000	0	0	0
<i>Classe 2</i>	1.6393	88.5246	9.6721	0.1639
<i>Classe 3</i>	0.0735	0.0735	99.8529	0
<i>Classe 4</i>	0	0	0	100.0000

Tabela 38 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.5.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	99.9115	0	0.0885	0
<i>Classe 2</i>	0	90.0000	10.0000	0
<i>Classe 3</i>	0	0	100.0000	0
<i>Classe 4</i>	0	0.2222	0	99.7778

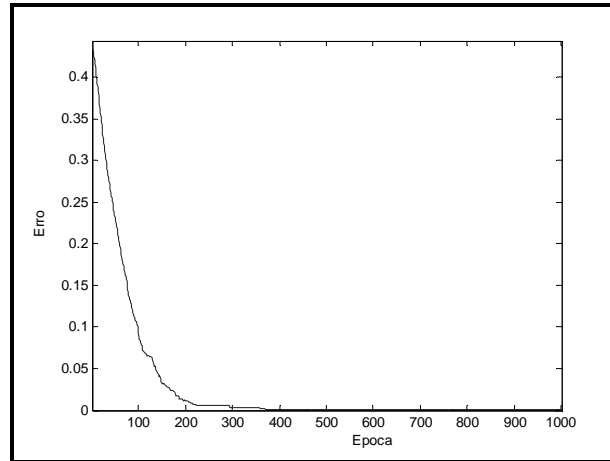
Tabela 39 - Matriz de confusão percentual do conjunto de teste do Experimento A.5.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	98.9474
<i>Por conjunto de teste</i>	99.8140
<i>Médio por classe no conjunto de treinamento</i>	97.0944
<i>Médio por classe no conjunto de teste</i>	97.4223

Tabela 40 - Percentuais de reconhecimento do Experimento A.5.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
0.9955	0.8837	0.9683	0.9996

Tabela 41 - Vetor de competência do Experimento A.5.



**Figura 20 - Curva conceitual de aprendizado do Experimento A.5.**

#### 4.2.1.6 Experimento A.6

O Experimento A.6 apresenta a arquitetura da RNA análoga ao Experimento A.2, exceto por apresentar duas camadas ocultas, com cinco neurônios cada. Isto também foi feito para se verificar o efeito na performance da rede. Na Tabela 42, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 43, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 44, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 45 demonstra o nível de especialização, através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 21 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	100.0000	0	0	0
<i>Classe 2</i>	1.6393	98.3607	0	0
<i>Classe 3</i>	5.8088	4.1912	90.0000	0
<i>Classe 4</i>	0	0	0	100.0000

**Tabela 42 - Matriz de confusão percentual do conjunto de treinamento do Experimento A.6.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Classe 1</i>	99.9115	0	0.0885	0
<i>Classe 2</i>	0	100.0000	0	0
<i>Classe 3</i>	5.2727	4.7273	90.0000	0
<i>Classe 4</i>	0	0	0	100.0000

Tabela 43 - Matriz de confusão percentual do conjunto de teste do Experimento A.6.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	97.8655
<i>Por conjunto de teste</i>	97.3953
<i>Médio por classe no conjunto de treinamento</i>	97.0902
<i>Médio por classe no conjunto de teste</i>	97.4779

Tabela 44 - Percentuais de reconhecimento do Experimento A.6.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
0.9718	0.9361	0.9000	1.0000

Tabela 45 - Vetor de competência do Experimento A.6.

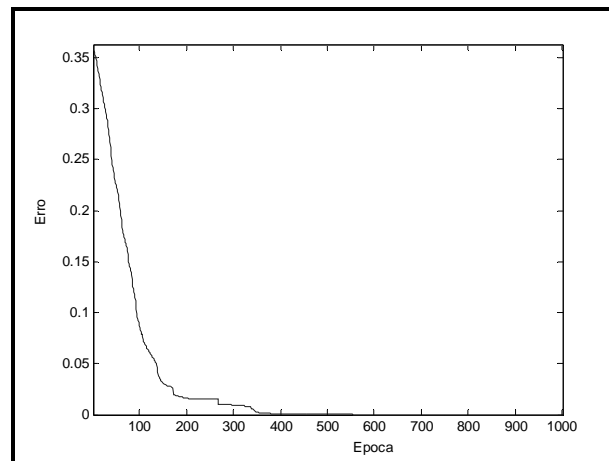


Figura 21 - Curva conceitual de aprendizado do Experimento A.6.



## 4.2.2 Experimento B: Determinação da Etnia simplificada

Este grupo de experimentos foi utilizado para aferir a eficácia do classificador neural em relação à etnia simplificada.

### 4.2.2.1 Experimento B.1

No Experimento *B.1* foi utilizada a base de dados referente a Etnia Simplificada classificada em etnia branca, parda e negra. A arquitetura da RNA utilizada neste experimento foi composta por dados de entrada não estratificados com 97 padrões da etnia negra, 459 da etnia branca, 128 da etnia parda, dando um total de 684 padrões para treinamento. Os dados de saída também não eram estratificados com um total de 215 padrões sendo representados por 3 bits. Foram utilizados 5 neurônios na camada de entrada. Neste experimento, assim como nos que se seguirão foi feita uma variação no número de neurônios na camada oculta, utilizando-se conceitos vistos na Seção 4.1 e, com o objetivo de se variar a arquitetura da rede para se observar a performance da mesma, tomou-se como ponto de início utilizar 3 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída foram utilizados 3 neurônios, também com função *logsig*.

Na Tabela 46 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 47, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 48, observa-se o percentual de reconhecimento por conjunto de treinamento e teste e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 49 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 22 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	0	100.0000	0
<i>Classe 3</i>	0	99.1406	0.8594

Tabela 46 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.1.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	0	99.6552	0.3448
<i>Classe 3</i>	0	100.0000	0

Tabela 47 - Matriz de confusão percentual do conjunto de teste do Experimento B.1.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.2661
<i>Por conjunto de teste</i>	53.7674
<i>Médio por classe no conjunto de treinamento</i>	33.6198
<i>Médio por classe no conjunto de teste</i>	33.2184

Tabela 48 - Percentuais de reconhecimento do Experimento B.1.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6721	0.0086

Tabela 49 - Vetor de competência do Experimento B.1.

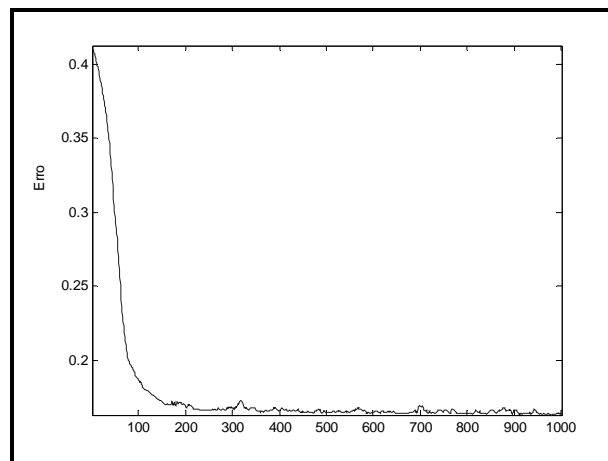


Figura 22 - Curva conceitual de aprendizado do Experimento B.1.

#### 4.2.2.2 Experimento B.2

O Experimento *B.2* apresenta a arquitetura da RNA análoga ao experimento anterior exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 5. Na Tabela 50, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 51, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 52, observa-se o percentual de reconhecimento por conjunto de treinamento e teste e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 53 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 23 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	0	100.0000	0
<i>Classe 3</i>	0	99.3750	0.6250

Tabela 50 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.2.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100	0
<i>Classe 2</i>	0	100	0
<i>Classe 3</i>	0	100	0

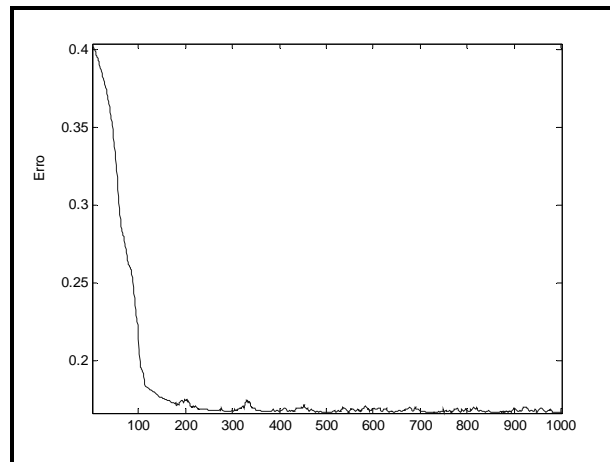
Tabela 51 - Matriz de confusão percentual do conjunto de teste do Experimento B.2.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.2222
<i>Por conjunto de teste</i>	53.9535
<i>Médio por classe no conjunto de treinamento</i>	33.5417
<i>Médio por classe no conjunto de teste</i>	33.3333

Tabela 52 - Percentuais de reconhecimento do Experimento B.2.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6718	0.0063

Tabela 53 - Vetor de competência do Experimento B.2.



**Figura 23 - Curva conceitual de aprendizado do Experimento B.2.**

#### 4.2.2.3 Experimento B.3

O Experimento *B.3* apresenta a arquitetura da RNA análoga ao Experimento *B.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 7. Na Tabela 54, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 55, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 56, observa-se o percentual de reconhecimento por conjunto de treinamento e teste e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 57 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 24 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8969	0.1031
<i>Classe 2</i>	0	99.9346	0.0654
<i>Classe 3</i>	0	98.0469	1.9531

**Tabela 54 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.3.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	0	99.6552	0.3448
<i>Classe 3</i>	0	99.7872	0.2128

Tabela 55 - Matriz de confusão percentual do conjunto de teste do Experimento B.3.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.4269
<i>Por conjunto de teste</i>	53.8140
<i>Médio por classe no conjunto de treinamento</i>	33.9626
<i>Médio por classe no conjunto de teste</i>	33.2893

Tabela 56 - Percentuais de reconhecimento do Experimento B.3.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6732	0.0194

Tabela 57 - Vetor de competência do Experimento B.3.

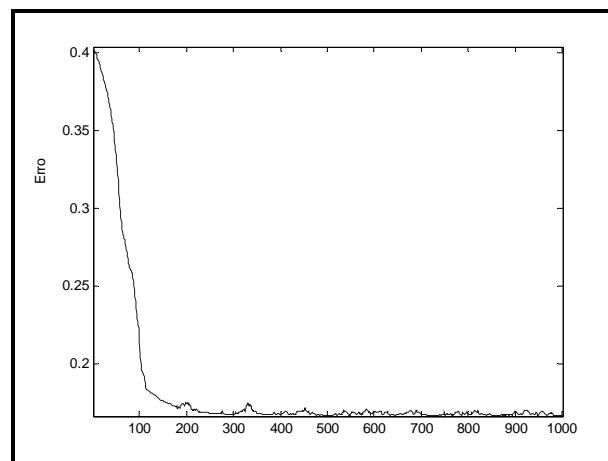


Figura 24 - Curva conceitual de aprendizado do Experimento B.3.

#### 4.2.2.4 Experimento B.4

O Experimento *B.4* apresenta a arquitetura da RNA análoga ao Experimento *B.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 10. Na Tabela 58, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 59, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 60, observa-se o percentual de reconhecimento por conjunto de treinamento e teste e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 61 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 25 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8969	0.1031
<i>Classe 2</i>	0	99.9564	0.0436
<i>Classe 3</i>	0	97.0313	2.9688

Tabela 58 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.4.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.4231	0.5769
<i>Classe 2</i>	0	99.4828	0.5172
<i>Classe 3</i>	0	100.0000	0

Tabela 59 - Matriz de confusão percentual do conjunto de teste do Experimento B.4.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.6316
<i>Por conjunto de teste</i>	53.6744
<i>Médio por classe no conjunto de treinamento</i>	34.3084
<i>Médio por classe no conjunto de teste</i>	33.1609

Tabela 60 - Percentuais de reconhecimento do Experimento B.4.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6746	0.0296

Tabela 61 - Vetor de competência do Experimento B.4.

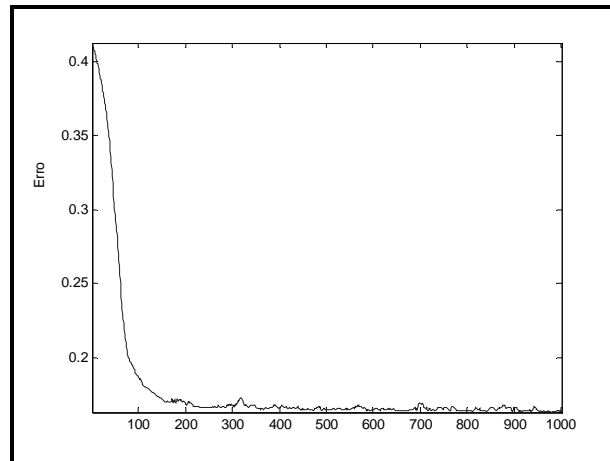


Figura 25- Curva conceitual de aprendizado do Experimento B.4.

#### 4.2.2.5 Experimento B.5

O Experimento *B.5* apresenta a arquitetura da RNA análoga ao Experimento *B.2*, exceto por apresentar a função de transferência *tansig* na camada oculta. Isto também foi feito para se verificar o efeito na performance da rede. Na Tabela 62, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 63, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 64, observa-se o percentual de reconhecimento por conjunto de treinamento e teste e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 65 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 26 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.1031	99.8969	0
<i>Classe 2</i>	0	99.9782	0.0218
<i>Classe 3</i>	0.0781	98.3594	1.5625

Tabela 62 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.5.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8077	0.1923
<i>Classe 2</i>	0	99.5690	0.4310
<i>Classe 3</i>	0	100.0000	0

Tabela 63 - Matriz de confusão percentual do conjunto de teste do Experimento B.5.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.3977
<i>Por conjunto de teste</i>	53.7209
<i>Médio por classe no conjunto de treinamento</i>	33.8813
<i>Médio por classe no conjunto de teste</i>	33.1897

Tabela 64 - Percentuais de reconhecimento do Experimento B.5.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.0010	0.6731	0.0156

Tabela 65 - Vetor de competência do Experimento B.5.

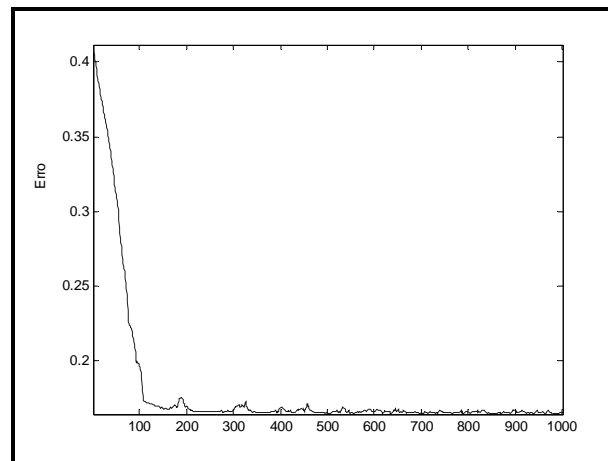


Figura 26 - Curva conceitual de aprendizado do Experimento B.5.



#### 4.2.2.6 Experimento B.6

O Experimento *B.6* apresenta a arquitetura da RNA análoga ao Experimento *B.2*, exceto por apresentar duas camadas ocultas, com cinco neurônios cada. Na Tabela 66, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 67, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 68, observa-se o percentual de reconhecimento por conjunto de treinamento e teste e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 69 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 27 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.1031	99.7938	0.1031
<i>Classe 2</i>	0.0218	99.9564	0.0218
<i>Classe 3</i>	0	98.0469	1.9531

Tabela 66 - Matriz de confusão percentual do conjunto de treinamento do Experimento B.6.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.6154	0.3846
<i>Classe 2</i>	0.1724	99.2241	0.6034
<i>Classe 3</i>	0	99.7872	0.2128

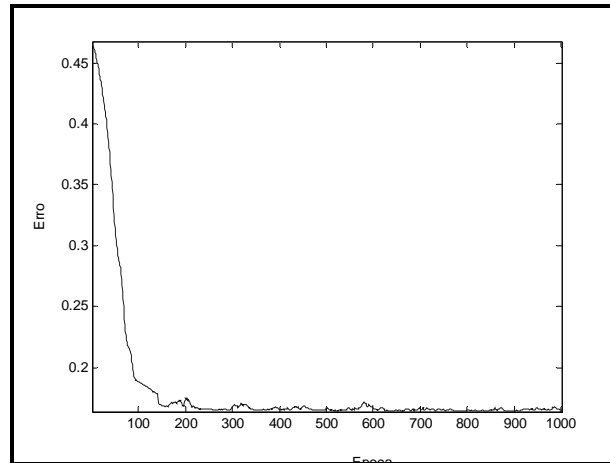
Tabela 67 - Matriz de confusão percentual do conjunto de teste do Experimento B.6.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.4561
<i>Por conjunto de teste</i>	53.5814
<i>Médio por classe no conjunto de treinamento</i>	34.0042
<i>Médio por classe no conjunto de teste</i>	33.1456

Tabela 68 - Percentuais de reconhecimento do Experimento B.6.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.0010	0.6734	0.0195

**Tabela 69 - Vetor de competência do Experimento B.6.**



**Figura 27 - Curva conceitual de aprendizado do Experimento B.6.**

### **4.2.3 Experimento C: Estratificação do conjunto de dados**

Este grupo de experimentos foi utilizado para aferir a eficácia do classificador neural, em relação à etnia simplificada, utilizando-se de técnicas para estratificação da base de dados. Este experimento utilizou o conjunto de treinamento estratificado.

#### **4.2.3.1 Experimento C.1**

No Experimento *C.1* foi utilizada a base de dados referente a Etnia Simplificada, classificada em etnia branca, parda e negra. A arquitetura da RNA utilizada neste experimento foi composta por dados de entrada estratificados com 485 padrões da etnia negra, 459 da etnia branca, 512 da etnia parda, formando o conjunto de treinamento. Os dados de saída não eram estratificados, com um total de 215 padrões sendo representados por 3 bits. Foram utilizados 5 neurônios na camada de entrada. Neste experimento, assim como nos que se seguirão, foi feita uma variação no número de neurônios na camada oculta, utilizando-se conceitos vistos na

Seção 4.1 e, com o objetivo de se variar a arquitetura da rede, para se observar a performance da mesma, tomou-se como ponto de início utilizar 3 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída foram utilizados 3 neurônios, também com função *logsig*.

Na Tabela 70 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 71, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 72, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 73 demonstra o nível de especialização, através do vetor de competência no conjunto de treinamento, e, finalmente, na Figura 28 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	18.8660	14.2268	66.9072
<i>Classe 2</i>	17.9956	15.8606	66.1438
<i>Classe 3</i>	17.8125	13.5156	68.6719

Tabela 70 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.1.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	17.8846	10.0000	72.1154
<i>Classe 2</i>	16.8966	10.6034	72.5000
<i>Classe 3</i>	18.2979	12.7660	68.9362

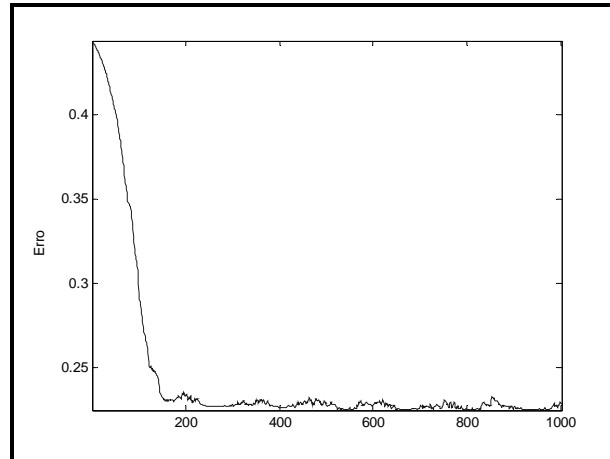
Tabela 71 - Matriz de confusão percentual do conjunto de teste do Experimento C.1.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	35.4327
<i>Por conjunto de teste</i>	25.1163
<i>Médio por classe no conjunto de treinamento</i>	34.4661
<i>Médio por classe no conjunto de teste</i>	32.4747

Tabela 72 - Percentuais de reconhecimento do Experimento C.1.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.0853	0.0990	0.2823

**Tabela 73 - Vetor de competência do Experimento C.1.**



**Figura 28 - Curva conceitual de aprendizado do Experimento C.1.**

#### 4.2.3.2 Experimento C.2

O Experimento C.2 apresenta a arquitetura da RNA análoga ao experimento anterior exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 5. Na Tabela 74, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 75, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 76, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 77 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 29 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	48.4536	22.2680	29.2784
<i>Classe 2</i>	38.9978	27.1895	33.8126
<i>Classe 3</i>	39.8438	20.3906	39.7656

**Tabela 74 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.2.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	51.1538	14.8077	34.0385
<i>Classe 2</i>	50.0862	15.1724	34.7414
<i>Classe 3</i>	46.3830	20.0000	33.6170

Tabela 75 - Matriz de confusão percentual do conjunto de teste do Experimento C.2.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	38.6951
<i>Por conjunto de teste</i>	27.9070
<i>Médio por classe no conjunto de treinamento</i>	38.4696
<i>Médio por classe no conjunto de teste</i>	33.3144

Tabela 76 - Percentuais de reconhecimento do Experimento C.2.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.2537	0.1796	0.2415

Tabela 77 - Vetor de competência do Experimento C.2.

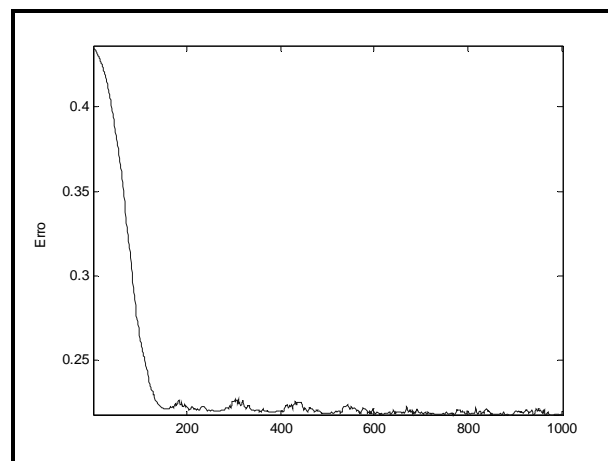


Figura 29 - Curva conceitual de aprendizado do Experimento C.2.

### 4.2.3.3 Experimento C.3

O Experimento C.3 apresenta a arquitetura da RNA análoga ao Experimento C.1, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 7. Na Tabela 78, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 79, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 80, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 81 demonstra o nível de especialização, através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 30 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	48.7629	21.4433	29.7938
<i>Classe 2</i>	38.9542	28.5621	32.4837
<i>Classe 3</i>	40.0000	21.0156	38.9844

Tabela 78 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.3.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	49.2308	13.6538	37.1154
<i>Classe 2</i>	46.1207	15.0000	38.8793
<i>Classe 3</i>	44.4681	20.0000	35.5319

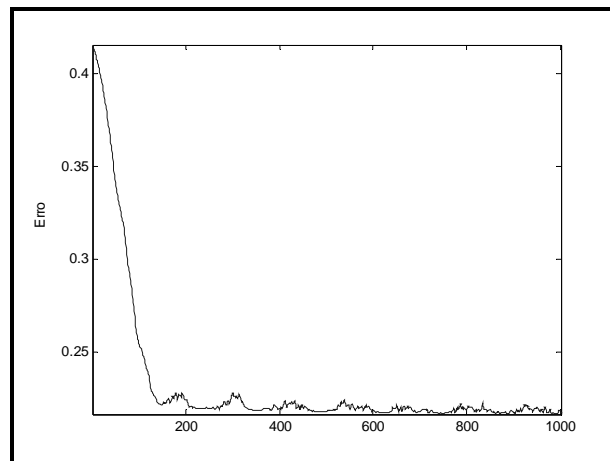
Tabela 79 - Matriz de confusão percentual do conjunto de teste do Experimento C.3.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	38.9560
<i>Por conjunto de teste</i>	27.7674
<i>Médio por classe no conjunto de treinamento</i>	38.7698
<i>Médio por classe no conjunto de teste</i>	33.2542

Tabela 80 - Percentuais de reconhecimento do Experimento C.3.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.2562	0.1824	0.2353

**Tabela 81 - Vetor de competência do Experimento C.3.**



**Figura 30 - Curva conceitual de aprendizado do Experimento C.3.**

#### 4.2.3.4 Experimento C.4

O Experimento *C.4* apresenta a arquitetura da RNA análoga ao Experimento *C.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 10. Na Tabela 82, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 83, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 84, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 85 demonstra o nível de especialização, através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 31 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	57.9381	23.2990	18.7629
<i>Classe 2</i>	44.2702	33.0283	22.7015
<i>Classe 3</i>	45.3125	24.2188	30.4688

Tabela 82 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.4.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	59.0385	15.1923	25.7692
<i>Classe 2</i>	57.4138	16.9828	25.6034
<i>Classe 3</i>	54.6809	21.9149	23.4043

Tabela 83 - Matriz de confusão percentual do conjunto de teste do Experimento C.4.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	40.4258
<i>Por conjunto de teste</i>	28.5581
<i>Médio por classe no conjunto de treinamento</i>	40.4784
<i>Médio por classe no conjunto de teste</i>	33.1418

Tabela 84 - Percentuais de reconhecimento do Experimento C.4.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.3010	0.2116	0.2141

Tabela 85 - Vetor de competência do Experimento C.4.

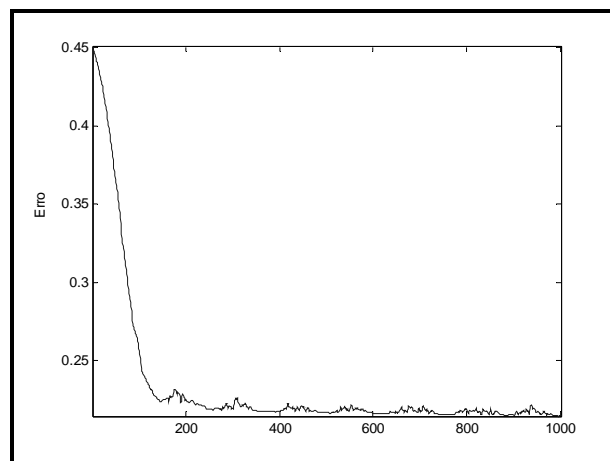


Figura 31 - Curva conceitual de aprendizado do Experimento C.4.



#### 4.2.3.5 Experimento C.5

O Experimento C.5 apresenta a arquitetura da RNA análoga ao Experimento C.2, exceto por apresentar a função de transferência *tansig* na camada oculta. Na Tabela 86, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 87, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 88, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 89 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 32 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	48.3505	27.6289	24.0206
<i>Classe 2</i>	38.9107	33.3333	27.7560
<i>Classe 3</i>	40.6250	27.2656	32.1094

Tabela 86 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.5.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	52.6923	16.3462	30.9615
<i>Classe 2</i>	52.3276	17.0690	30.6034
<i>Classe 3</i>	46.5957	24.6809	28.7234

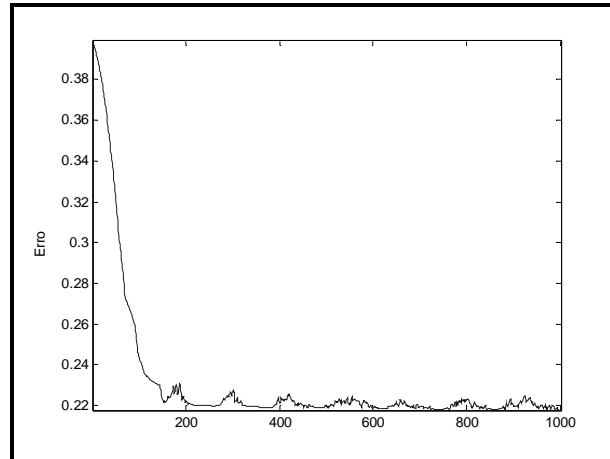
Tabela 87 - Matriz de confusão percentual do conjunto de teste do Experimento C.5.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	37.9052
<i>Por conjunto de teste</i>	28.2326
<i>Médio por classe no conjunto de treinamento</i>	37.9311
<i>Médio por classe no conjunto de teste</i>	32.8282

Tabela 88 - Percentuais de reconhecimento do Experimento C.5.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.2544	0.1926	0.1909

**Tabela 89 - Vetor de competência do Experimento C.5.**



**Figura 32 - Curva conceitual de aprendizado do Experimento C.5.**

#### **4.2.3.6 Experimento C.6**

O Experimento C.6 apresenta a arquitetura da RNA análoga ao Experimento C.2, exceto por apresentar duas camadas ocultas, com cinco neurônios cada. Isto também foi feito para se verificar o efeito na performance da rede. Na Tabela 90, pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 91, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 92, observa-se o percentual de reconhecimento, por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 93 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 33 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	48.8660	17.2165	33.9175
<i>Classe 2</i>	38.6928	23.8126	37.4946
<i>Classe 3</i>	39.0625	17.9688	42.9688

Tabela 90 - Matriz de confusão percentual do conjunto de treinamento do Experimento C.6.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	49.8077	10.3846	39.8077
<i>Classe 2</i>	46.2069	12.1552	41.6379
<i>Classe 3</i>	44.6809	14.6809	40.6383

Tabela 91 - Matriz de confusão percentual do conjunto de teste do Experimento C.6.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	38.8942
<i>Por conjunto de teste</i>	27.4884
<i>Médio por classe no conjunto de treinamento</i>	38.5491
<i>Médio por classe no conjunto de teste</i>	34.2004

Tabela 92 - Percentuais de reconhecimento do Experimento C.6.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.2521	0.1515	0.2438

Tabela 93 - Vetor de competência do Experimento C.6.

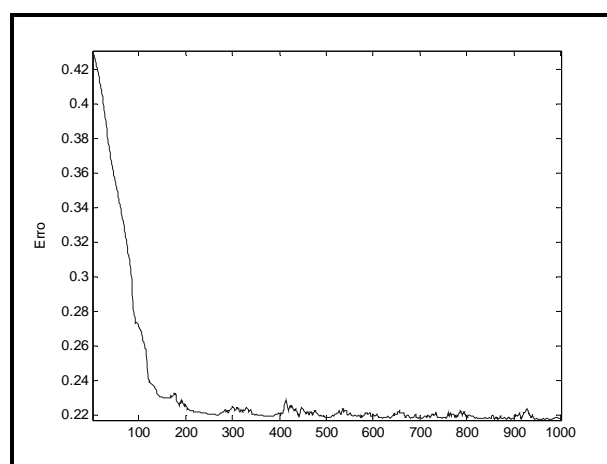


Figura 33 - Curva conceitual de aprendizado do Experimento C.6.

#### 4.2.4 Experimento D: Estratificação do conjunto de dados

Este grupo de experimentos foi utilizado para aferir a eficácia do classificador neural em relação à etnia simplificada, utilizando-se de técnicas para estratificação da base de dados. Este experimento utilizou o conjunto de teste estratificado.

##### 4.2.4.1 Experimento D.1

No Experimento *D.1* foi utilizada a base de dados referente a Etnia Simplificada classificada em etnia branca, parda e negra. A arquitetura da RNA utilizada neste experimento, foi composta por dados de entrada não estratificados, com 684 padrões formando o conjunto de treinamento. Os dados de saída eram estratificados, com 104 padrões da etnia negra, 116 padrões da etnia branca e 94 padrões da etnia parda dando um total de 314 padrões representados por 3 bits. Foram utilizados 5 neurônios na camada de entrada. Neste experimento, assim como nos que se seguirão, foi feita uma variação no número de neurônios na camada oculta, utilizando-se conceitos vistos na Seção 4.1 e, com o objetivo de se variar a arquitetura da rede, para se observar a performance da mesma, tomou-se como ponto de início utilizar 3 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída foram utilizados 3 neurônios, também com função *logsig*.

Na Tabela 94 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 95, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 96, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 97 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento e, finalmente, na Figura 34 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.7938	0.2062
<i>Classe 2</i>	0	99.9564	0.0436
<i>Classe 3</i>	0	98.3594	1.6406

Tabela 94 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.1.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8077	0.1923
<i>Classe 2</i>	0	99.4828	0.5172
<i>Classe 3</i>	0	99.7872	0.2128

Tabela 95 - Matriz de confusão percentual do conjunto de teste do Experimento D.1.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.3830
<i>Por conjunto de teste</i>	36.8153
<i>Médio por classe no conjunto de treinamento</i>	33.8657
<i>Médio por classe no conjunto de teste</i>	33.2318

Tabela 96 - Percentuais de reconhecimento do Experimento D.1.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6730	0.0163

Tabela 97 - Vetor de competência do Experimento D.1.

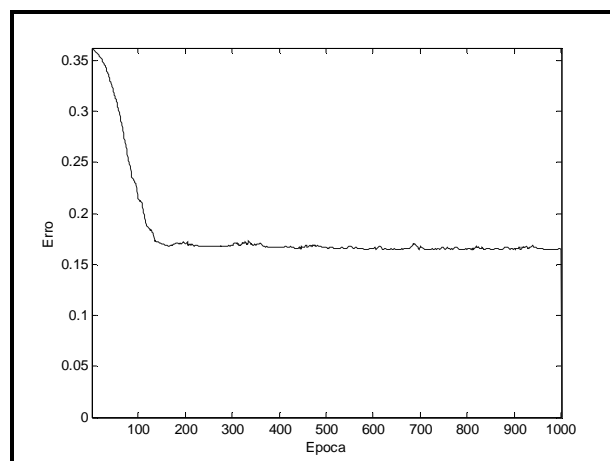


Figura 34 - Curva conceitual de aprendizado do Experimento D.1.

#### 4.2.4.2 Experimento D.2

O Experimento *D.2* apresenta a arquitetura da RNA análoga ao experimento anterior, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 5. Na Tabela 98 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 99, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 100, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 101 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 35 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8969	0.1031
<i>Classe 2</i>	0	99.8693	0.1307
<i>Classe 3</i>	0	97.9688	2.0313

Tabela 98 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.2.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8077	0.1923
<i>Classe 2</i>	0	99.6552	0.3448
<i>Classe 3</i>	0	99.7872	0.2128

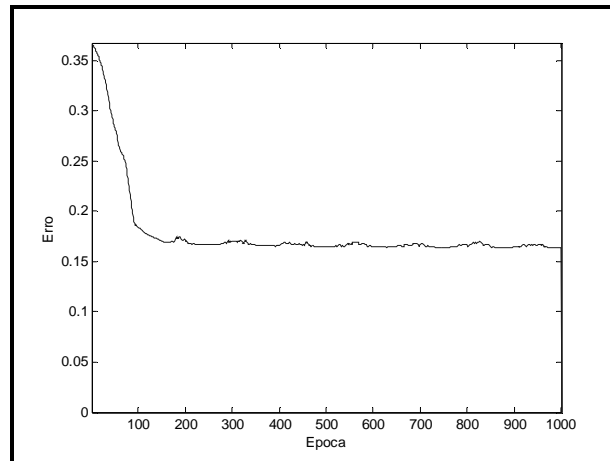
Tabela 99 - Matriz de confusão percentual do conjunto de teste do Experimento D.2.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.3977
<i>Por conjunto de teste</i>	36.8790
<i>Médio por classe no conjunto de treinamento</i>	33.9668
<i>Médio por classe no conjunto de teste</i>	33.2893

Tabela 100 - Percentuais de reconhecimento do Experimento D.2.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6728	0.0201

**Tabela 101 - Vetor de competência do Experimento D.2.**



**Figura 35 - Curva conceitual de aprendizado do Experimento D.2.**

#### 4.2.4.3 Experimento D.3

O Experimento *D.3* apresenta a arquitetura da RNA análoga ao Experimento *D.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 7. Na Tabela 102 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 103, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 104, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 105 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 36 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	0	99.9564	0.0436
<i>Classe 3</i>	0	98.0469	1.9531

**Tabela 102 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.3.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8077	0.1923
<i>Classe 2</i>	0	99.3103	0.6897
<i>Classe 3</i>	0	100.0000	0

Tabela 103 - Matriz de confusão percentual do conjunto de teste do Experimento D.3.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.4415
<i>Por conjunto de teste</i>	36.6879
<i>Médio por classe no conjunto de treinamento</i>	33.9699
<i>Médio por classe no conjunto de teste</i>	33.1034

Tabela 104 - Percentuais de reconhecimento do Experimento D.3.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6732	0.0195

Tabela 105 - Vetor de competência do Experimento D.3.

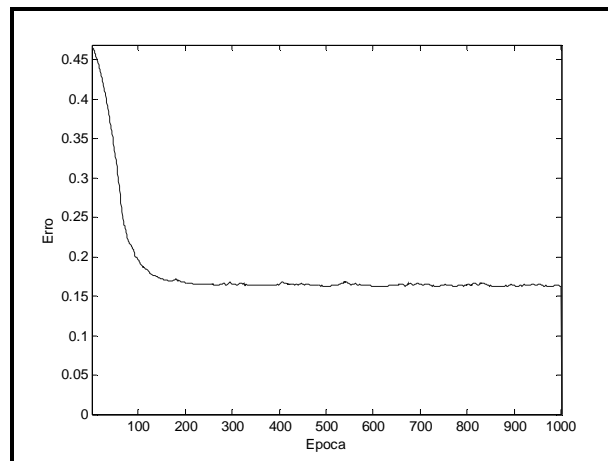


Figura 36 - Curva conceitual de aprendizado do Experimento D.3.



#### 4.2.4.4 Experimento D.4

O Experimento *D.4* apresenta a arquitetura da RNA análoga ao Experimento *D.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 10. Na Tabela 106 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 107, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 108, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 109 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 37 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.1031	99.6907	0.2062
<i>Classe 2</i>	0	99.9129	0.0871
<i>Classe 3</i>	0.2344	96.9531	2.8125

Tabela 106 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.4.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.2308	0.7692
<i>Classe 2</i>	0	99.6552	0.3448
<i>Classe 3</i>	0	99.7872	0.2128

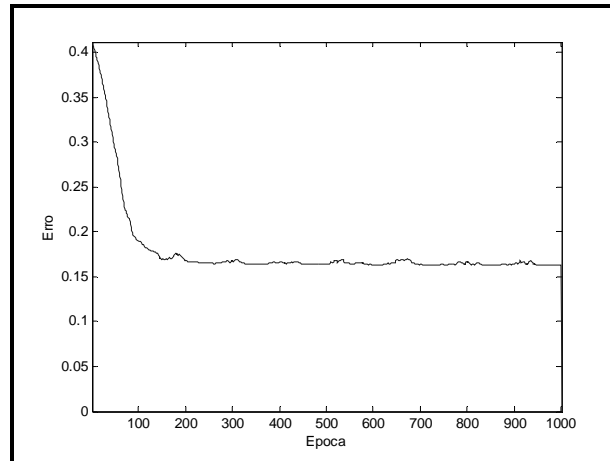
Tabela 107 - Matriz de confusão percentual do conjunto de teste do Experimento D.4.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.5877
<i>Por conjunto de teste</i>	36.8790
<i>Médio por classe no conjunto de treinamento</i>	34.2761
<i>Médio por classe no conjunto de teste</i>	33.2893

Tabela 108 - Percentuais de reconhecimento do Experimento D.4.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.0010	0.6746	0.0280

**Tabela 109 - Vetor de competência do Experimento D.4.**



**Figura 37 - Curva conceitual de aprendizado do Experimento D.4.**

#### 4.2.4.5 Experimento D.5

O Experimento *D.5* apresenta a arquitetura da RNA análoga ao Experimento *D.2*, exceto por apresentar a função de transferência *tansig* na camada oculta. Na Tabela 110 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 111, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 112, observe o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 113 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 38 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8969	0.1031
<i>Classe 2</i>	0	100.0000	0
<i>Classe 3</i>	0	97.7344	2.2656

**Tabela 110 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.5.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.6154	0.3846
<i>Classe 2</i>	0	99.4828	0.5172
<i>Classe 3</i>	0	99.7872	0.2128

Tabela 111 - Matriz de confusão percentual do conjunto de teste do Experimento D.5.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.5292
<i>Por conjunto de teste</i>	36.8153
<i>Médio por classe no conjunto de treinamento</i>	34.0885
<i>Médio por classe no conjunto de teste</i>	33.2318

Tabela 112 - Percentuais de reconhecimento do Experimento D.5.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0	0.6740	0.0226

Tabela 113 - Vetor de competência do Experimento D.5.

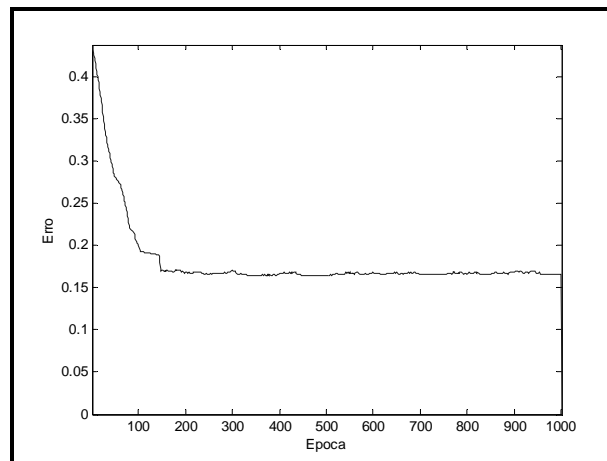


Figura 38 - Curva conceitual de aprendizado do Experimento D.5.

#### 4.2.4.6 Experimento D.6

O Experimento *D.6* apresenta a arquitetura da RNA análoga ao Experimento *D.2*, exceto por apresentar duas camadas ocultas, com cinco neurônios cada. Na Tabela 114 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 115, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 116, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 117 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 39 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.2062	99.6907	0.1031
<i>Classe 2</i>	0	100.0000	0
<i>Classe 3</i>	0.1563	97.9688	1.8750

Tabela 114 - Matriz de confusão percentual do conjunto de treinamento do Experimento D.6.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8077	0.1923
<i>Classe 2</i>	0	99.6552	0.3448
<i>Classe 3</i>	0	100.0000	0

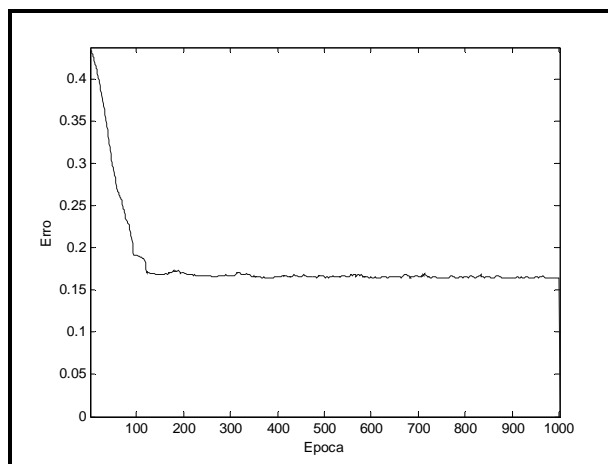
Tabela 115 - Matriz de confusão percentual do conjunto de teste do Experimento D.6.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.4854
<i>Por conjunto de teste</i>	36.8153
<i>Médio por classe no conjunto de treinamento</i>	34.0271
<i>Médio por classe no conjunto de teste</i>	33.2184

Tabela 116 - Percentuais de reconhecimento do Experimento D.6.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.0020	0.6739	0.0187

**Tabela 117 - Vetor de competência do Experimento D.6.**



**Figura 39 - Curva conceitual de aprendizado do Experimento D.6.**

#### **4.2.5 Experimento E: Utilização da técnica de RDP e Comitês de RNA's**

Este grupo de experimentos foi utilizado para aferir a eficácia do classificador neural em relação à etnia simplificada, utilizando-se de técnicas de RDP e comitês de Redes Neurais Artificiais.

##### **4.2.5.1 Experimento E.1**

No Experimento *E.1* foi utilizada a base de dados referente a Etnia Simplificada classificada em etnia branca, parda e negra. A arquitetura da RNA utilizada neste experimento foi composta por dados de entrada estratificados, com 1456 padrões inicialmente. Os dados de saída não eram estratificados, com 215 padrões no total representados por 3 bits. Foram utilizados 5 neurônios na camada de entrada. Neste experimento, assim como nos que se seguirão será aplicada a técnica de RDP iniciando-se com um fator 5 e variando-se até 25, utilizando-se conceitos vistos na Seção 3.7.3, isto com o objetivo de se variar o conjunto de treinamento e especializar cada classe. Esta técnica foi utilizada no comitê de redes neurais

cujo resultado era obtido por média, (Seção 3.7) a fim de se tentar resultados satisfatórios, ou, até mais eficientes que os anteriores. Neste experimento, assim como nos demais, foram 5 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída foram utilizados 3 neurônios, também com função *logsig*. Isto para cada rede componente do Comitê (Seção 3.7).

Na Tabela 118 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 119, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 120, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 121 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 40 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	50.5155	20.9278	28.5567
<i>Classe 2</i>	41.5904	26.9717	31.4379
<i>Classe 3</i>	41.9531	20.8594	37.1875

**Tabela 118 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.1.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	50.9615	12.5000	36.5385
<i>Classe 2</i>	50.3448	13.6207	36.0345
<i>Classe 3</i>	46.1702	18.7234	35.1064

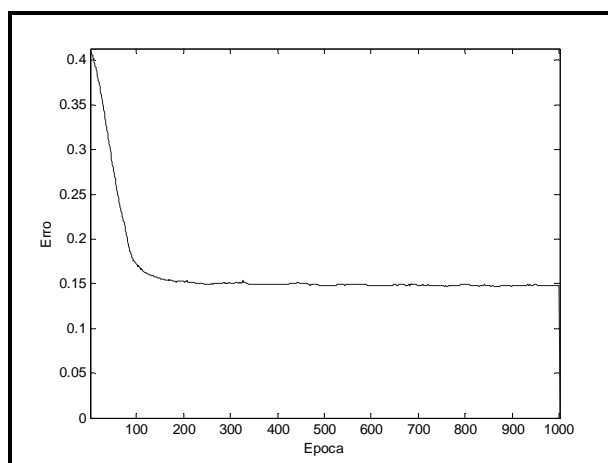
**Tabela 119 - Matriz de confusão percentual do conjunto de teste do Experimento E.1.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	38.4066
<i>Por conjunto de teste</i>	27.3488
<i>Médio por classe no conjunto de treinamento</i>	38.2249
<i>Médio por classe no conjunto de teste</i>	33.2295

**Tabela 120 - Percentuais de reconhecimento do Experimento E.1.**

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.2589	0.1650	0.2202

**Tabela 121 - Vetor de competência do Experimento E.1.**



**Figura 40 - Curva conceitual de aprendizado do Experimento E.1.**

#### 4.2.5.2 Experimento E.2

O Experimento *E.2* apresenta a arquitetura análoga ao experimento anterior, exceto por utilizar RDP com índice 10. Na Tabela 122 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 123, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 124, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 125 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 41 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	30.3093	41.0309	28.6598
<i>Classe 2</i>	30.5011	39.8257	29.6732
<i>Classe 3</i>	30.6250	39.6875	29.6875

**Tabela 122 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.2.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	32.3077	35.7692	31.9231
<i>Classe 2</i>	33.9655	34.6552	31.3793
<i>Classe 3</i>	34.4681	34.8936	30.6383

Tabela 123 - Matriz de confusão percentual do conjunto de teste do Experimento E.2.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	33.0907
<i>Por conjunto de teste</i>	33.2093
<i>Médio por classe no conjunto de treinamento</i>	33.2742
<i>Médio por classe no conjunto de teste</i>	32.5337

Tabela 124 - Percentuais de reconhecimento do Experimento E.2.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.1353	0.1839	0.1245

Tabela 125 - Vetor de competência do Experimento E.2.

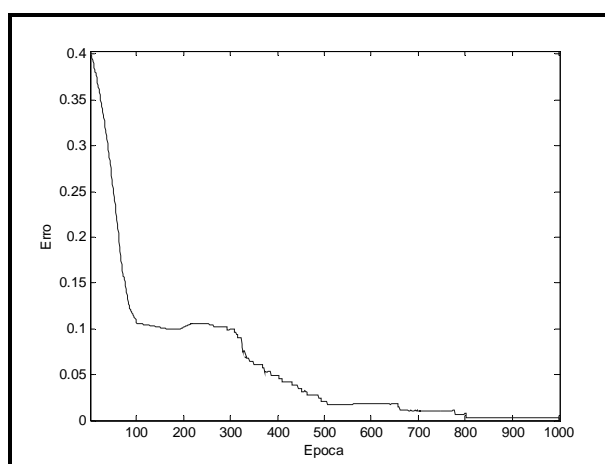


Figura 41 - Curva conceitual de aprendizado do Experimento E.2.



### 4.2.5.3 Experimento E.3

O Experimento *E.3* apresenta a arquitetura análoga ao Experimento *E.1*, exceto por utilizar RDP com índice 15. Na Tabela 126 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 127, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 128, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 129 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 42 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	47.5258	38.5567	13.9175
<i>Classe 2</i>	46.4488	39.1503	14.4009
<i>Classe 3</i>	45.9375	38.9063	15.1563

Tabela 126 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.3.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	45.7692	38.4615	15.7692
<i>Classe 2</i>	43.6207	38.7069	17.6724
<i>Classe 3</i>	44.0426	39.3617	16.5957

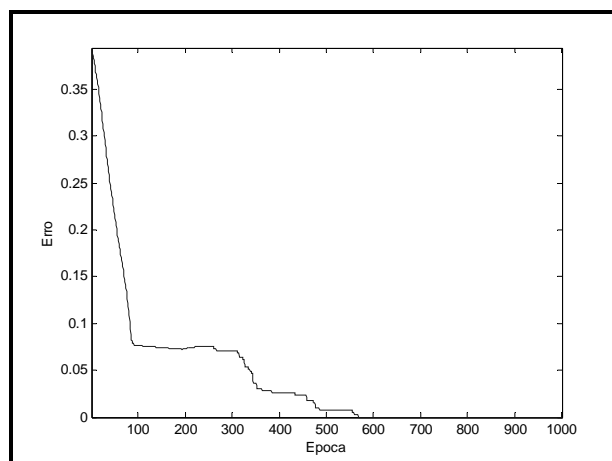
Tabela 127 - Matriz de confusão percentual do conjunto de teste do Experimento E.3.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	33.5027
<i>Por conjunto de teste</i>	35.5814
<i>Médio por classe no conjunto de treinamento</i>	33.9441
<i>Médio por classe no conjunto de teste</i>	33.6906

Tabela 128 - Percentuais de reconhecimento do Experimento E.3.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.1932	0.1370	0.0790

**Tabela 129 - Vetor de competência do Experimento E.3.**



**Figura 42 - Curva conceitual de aprendizado do Experimento E.3.**

#### 4.2.5.4 Experimento E.4

O Experimento *E.4* apresenta a arquitetura ao Experimento *E.1*, exceto por utilizar RDP com índice 20. Na Tabela 130 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 131, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 132, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 133 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 43 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	49.4845	36.8041	13.7113
<i>Classe 2</i>	48.8453	37.9739	13.1808
<i>Classe 3</i>	48.9063	36.8750	14.2188

**Tabela 130 - Matriz de confusão percentual do conjunto de treinamento do Experimento E.4.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	47.8846	39.8077	12.3077
<i>Classe 2</i>	47.1552	40.9483	11.8966
<i>Classe 3</i>	47.0213	41.4894	11.4894

Tabela 131 - Matriz de confusão percentual do conjunto de teste do Experimento E.4.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	33.4547
<i>Por conjunto de teste</i>	36.1860
<i>Médio por classe no conjunto de treinamento</i>	33.8924
<i>Médio por classe no conjunto de teste</i>	33.4408

Tabela 132 - Percentuais de reconhecimento do Experimento E.4.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.1848	0.1542	0.0780

Tabela 133 - Vetor de competência do Experimento E.4.

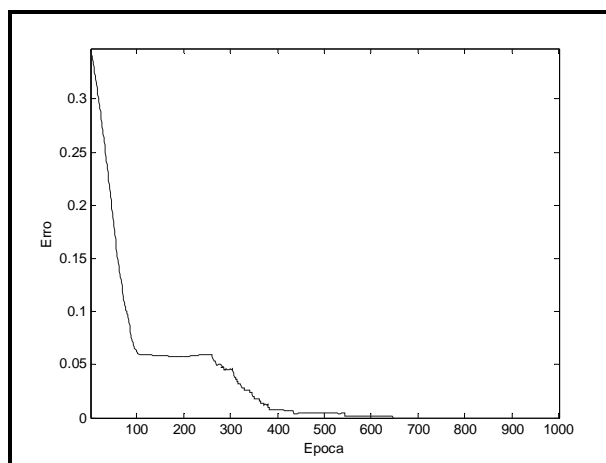


Figura 43 - Curva conceitual de aprendizado do Experimento E.4.

#### 4.2.6 Experimento F: Classificação sem a Classe Dominante

Este grupo de experimentos foi utilizado para aferir a eficácia do classificador neural em relação à etnia simplificada, removendo-se da base de dados a classe dominante (branca). Isto foi feito, intuitivamente, para se verificar o efeito na classificação das outras duas etnias.

##### 4.2.6.1 Experimento F.1

No Experimento *F.1* foi utilizada a base de dados referente a Etnia Simplificada classificada em etnia parda e negra sem a etnia branca. A arquitetura da RNA utilizada neste experimento foi composta por dados de entrada estratificados, com 106 padrões da etnia negra e 137 padrões da etnia parda formando o conjunto de treinamento. Os dados de saída não eram estratificados e foram representados por 2 bits. Foram utilizados 5 neurônios na camada de entrada. Neste experimento, assim como nos que se seguirão, foi feita uma variação na arquitetura da rede, variando-se o número de neurônios na camada oculta. Neste experimento tomou-se como ponto de início utilizar 3 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída foram utilizados 2 neurônios, também com função *logsig*. O objetivo de se retirar a etnia branca era verificar o efeito nos resultados já que na base de dados original a etnia branca era majoritária.

Na Tabela 134 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 135, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 136, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 137 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 44 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	2.3585	97.6415
<i>Classe 2</i>	2.4818	97.5182

**Tabela 134- - Matriz de confusão percentual do conjunto de treinamento do Experimento F.1.**

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	3.2558	96.7442
<i>Classe 2</i>	3.4211	96.5789

Tabela 135 - Matriz de confusão percentual do conjunto de teste do Experimento F.1.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	56.0082
<i>Por conjunto de teste</i>	47.0370
<i>Médio por classe no conjunto de treinamento</i>	49.9384
<i>Médio por classe no conjunto de teste</i>	49.9174

Tabela 136 - Percentuais de reconhecimento do Experimento F.1.

<i>Classe 1</i>	<i>Classe 2</i>
0.0208	0.0407

Tabela 137 - Vetor de competência do Experimento F.1.

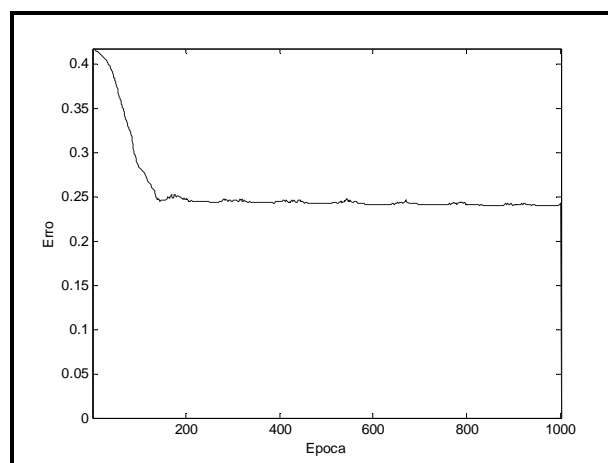


Figura 44 - Curva conceitual de aprendizado do Experimento F.1.

#### 4.2.6.2 Experimento F.2

O Experimento *F.2* apresenta a arquitetura da RNA análoga ao experimento anterior, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 5. Na Tabela 138 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 139, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 140, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 141 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 45 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	16.2264	83.7736
<i>Classe 2</i>	13.2117	86.7883

Tabela 138 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.2.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	16.2791	83.7209
<i>Classe 2</i>	15.2632	84.7368

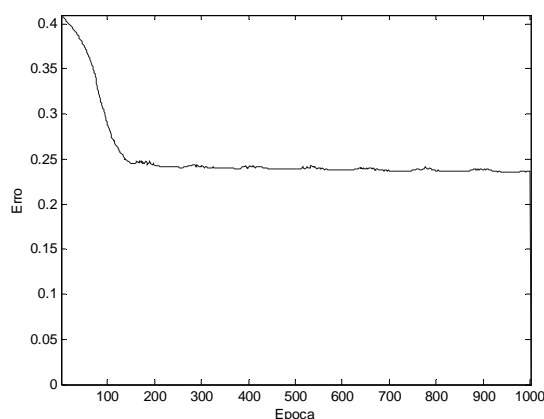
Tabela 139 - Matriz de confusão percentual do conjunto de teste do Experimento F.2.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	56.0082
<i>Por conjunto de teste</i>	48.3951
<i>Médio por classe no conjunto de treinamento</i>	51.5074
<i>Médio por classe no conjunto de teste</i>	50.5080

Tabela 140 - Percentuais de reconhecimento do Experimento F.2.

<i>Classe 1</i>	<i>Classe 2</i>
0.1103	0.5153

Tabela 141 - Vetor de competência do Experimento F.2.



**Figura 45 - Curva conceitual de aprendizado do Experimento F.2.**

#### 4.2.6.3 Experimento F.3

O Experimento *F.3* apresenta a arquitetura da RNA análoga ao Experimento *F.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 7. Na Tabela 142 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 143, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 144, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 145 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 46 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	21.6981	78.3019
<i>Classe 2</i>	17.8832	82.1168

**Tabela 142 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.3.**

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	21.3953	78.6047
<i>Classe 2</i>	19.4737	80.5263

**Tabela 143 - Matriz de confusão percentual do conjunto de teste do Experimento F.3.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	55.7613
<i>Por conjunto de teste</i>	49.1358
<i>Médio por classe no conjunto de treinamento</i>	51.9075
<i>Médio por classe no conjunto de teste</i>	50.9608

Tabela 144 - Percentuais de reconhecimento do Experimento F.3.

<i>Classe 1</i>	<i>Classe 2</i>
0.1396	0.4958

Tabela 145 - Vetor de competência do Experimento F.3.

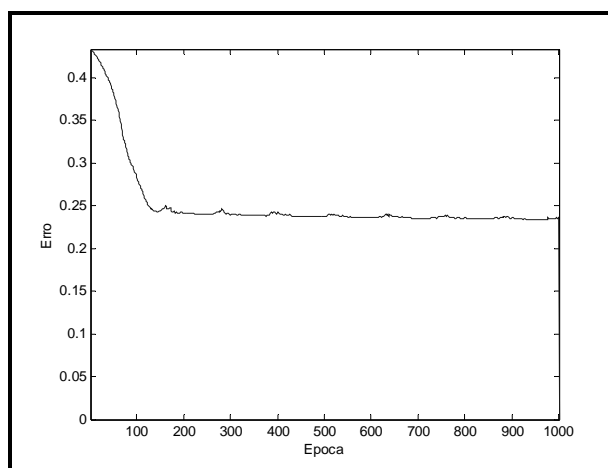


Figura 46 - Curva conceitual de aprendizado do Experimento F.3.

#### 4.2.6.4 Experimento F.4

O Experimento *F.4* apresenta a arquitetura da RNA análoga ao Experimento *F.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 10. Na Tabela 146 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 147, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 148, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 149 demonstra o nível de especialização através do vetor de



competência no conjunto de treinamento, e, finalmente, na Figura 47 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	23.6792	76.3208
<i>Classe 2</i>	17.2263	82.7737

Tabela 146 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.4.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	20.6977	79.3023
<i>Classe 2</i>	18.1579	81.8421

Tabela 147 - Matriz de confusão percentual do conjunto de teste do Experimento F.4.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	56.9959
<i>Por conjunto de teste</i>	49.3827
<i>Médio por classe no conjunto de treinamento</i>	53.2265
<i>Médio por classe no conjunto de teste</i>	51.2699

Tabela 148 - Percentuais de reconhecimento do Experimento F.4.

<i>Classe 1</i>	<i>Classe 2</i>
0.1499	0.5032

Tabela 149 - Vetor de competência do Experimento F.4.

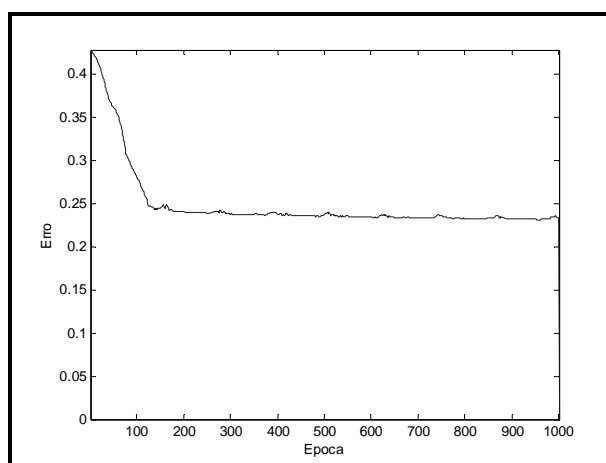


Figura 47 - Curva conceitual de aprendizado do Experimento F.4.

#### 4.2.6.5 Experimento F.5

O Experimento *F.5* apresenta a arquitetura da RNA análoga ao Experimento *F.4*, exceto por apresentar a função de transferência *tansig* na camada oculta. Na Tabela 150 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 151, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 152, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 153 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 48 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	18.1132	81.8868
<i>Classe 2</i>	11.6788	88.3212

Tabela 150 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.5.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	15.5814	84.4186
<i>Classe 2</i>	12.1053	87.8947

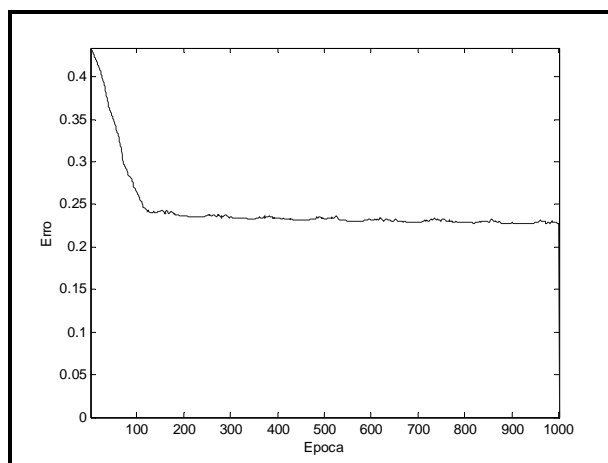
Tabela 151 - Matriz de confusão percentual do conjunto de teste do Experimento F.5.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	57.6955
<i>Por conjunto de teste</i>	49.5062
<i>Médio por classe no conjunto de treinamento</i>	53.2172
<i>Médio por classe no conjunto de teste</i>	51.7381

Tabela 152 - Percentuais de reconhecimento do Experimento F.5.

<i>Classe 1</i>	<i>Classe 2</i>
0.1437	0.5376

Tabela 153 - Vetor de competência do Experimento F.5.



**Figura 48 - Curva conceitual de aprendizado do Experimento F.5.**

#### 4.2.6.6 Experimento F.6

O Experimento *F.6* apresenta a arquitetura da RNA análoga ao Experimento *F.4*, exceto por apresentar duas camadas ocultas, com 10 neurônios cada. Na Tabela 154 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 155, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 156, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 157 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 49 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	41.5094	58.4906
<i>Classe 2</i>	30.6569	69.3431

**Tabela 154 - Matriz de confusão percentual do conjunto de treinamento do Experimento F.6.**

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	39.0698	60.9302
<i>Classe 2</i>	33.4211	66.5789

**Tabela 155 - Matriz de confusão percentual do conjunto de teste do Experimento F.6.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	57.2016
<i>Por conjunto de teste</i>	51.9753
<i>Médio por classe no conjunto de treinamento</i>	55.4262
<i>Médio por classe no conjunto de teste</i>	52.8244

Tabela 156 - Percentuais de reconhecimento do Experimento F.6.

<i>Classe 1</i>	<i>Classe 2</i>
0.2634	0.4590

Tabela 157 - Vetor de competência do Experimento F.6.

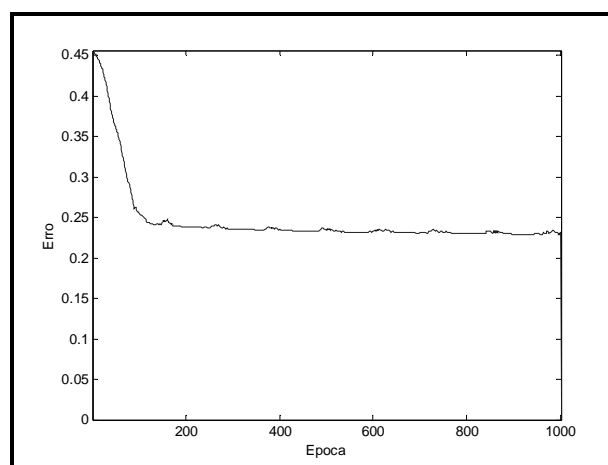


Figura 49 - Curva conceitual de aprendizado do Experimento F.6.

#### 4.2.7 Experimento G: Separação dos Atributos de Entrada

Este grupo de experimentos foi utilizado para aferir a eficácia do classificador neural, em relação à etnia simplificada, utilizando-se dos atributos de entrada binarizados. Isto foi feito a fim de verificar o efeito de uma melhor separação dos dados de entrada.

#### 4.2.7.1 Experimento G.1

No Experimento *G.1* foi utilizada a base de dados referente a Etnia Simplificada, classificada em etnia branca, parda e negra. A arquitetura da RNA utilizada neste experimento foi composta por dados de entrada não estratificados, com 97 padrões da etnia negra, 459 da etnia branca, 128 da etnia parda, dando um total de 684 padrões para treinamento. Os dados de saída também não eram estratificados, com um total de 215 padrões sendo representados por 3 bits. Foram utilizados 15 neurônios na camada de entrada, pois cada reagente passou a ser representado por 3 dígitos binários. Isto foi feito, tentando-se separar melhor os dados de entrada, a fim de se tentar atingir resultados mais eficientes. Além disso, neste experimento, assim como nos que se seguirão, foi feita uma variação no número de neurônios na camada oculta, utilizando-se conceitos vistos na Seção 4.1 e, com o objetivo de se variar a arquitetura da rede, para se observar a performance da mesma, tomou-se como ponto de início utilizar 5 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída foram utilizados 3 neurônios, também com função *logsig*.

Na Tabela 158 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 159, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 160, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 161 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 50 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.6186	98.2474	1.1340
<i>Classe 2</i>	0	99.8667	0.1333
<i>Classe 3</i>	0.2362	95.1969	4.5669

**Tabela 158 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.1.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8077	0.1923
<i>Classe 2</i>	0	98.9600	1.0400
<i>Classe 3</i>	0	99.3750	0.6250

Tabela 159 - Matriz de confusão percentual do conjunto de teste do Experimento G.1.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.6261
<i>Por conjunto de teste</i>	55.1111
<i>Médio por classe no conjunto de treinamento</i>	35.0174
<i>Médio por classe no conjunto de teste</i>	33.1950

Tabela 160 - Percentuais de reconhecimento do Experimento G.1.

<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
0.0061	0.6746	0.0449

Tabela 161 - Vetor de competência do Experimento G.1.

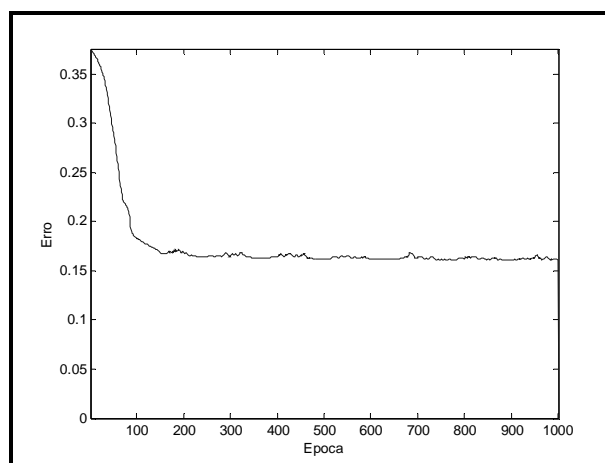


Figura 50 - Curva conceitual de aprendizado do Experimento G.1.

#### 4.2.7.2 Experimento G.2

O Experimento G.2 apresenta a arquitetura da RNA análoga ao experimento anterior, exceto com uma variação no número de neurônios na camada oculta, que de 5 passou para 10.

Na Tabela 162 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 163, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 164, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 165 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 51 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.9278	97.4227	1.6495
<i>Classe 2</i>	0.0667	99.5778	0.3556
<i>Classe 3</i>	0.1575	92.8346	7.0079

**Tabela 162 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.2.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	0.4800	98.0000	1.5200
<i>Classe 3</i>	0	98.9583	1.0417

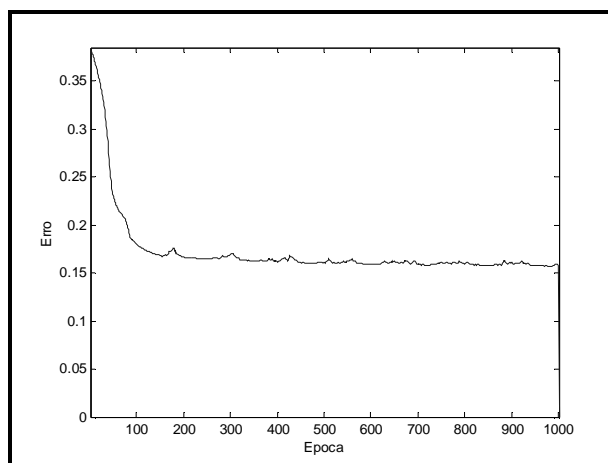
**Tabela 163 - Matriz de confusão percentual do conjunto de teste do Experimento G.2.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	67.9377
<i>Por conjunto de teste</i>	54.6667
<i>Médio por classe no conjunto de treinamento</i>	35.8378
<i>Médio por classe no conjunto de teste</i>	33.0139

**Tabela 164 - Percentuais de reconhecimento do Experimento G.2.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.0092	0.6765	0.0678

**Tabela 165 - Vetor de competência do Experimento G.2.**



**Figura 51 - Curva conceitual de aprendizado do Experimento G.2.**

#### 4.2.7.3 Experimento G.3

O Experimento *G.3* apresenta a arquitetura da RNA análoga ao Experimento *G.1*, exceto por apresentar 15 neurônios na camada oculta. Na Tabela 166 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 167, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 168, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 169 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 52 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	1.9588	96.7010	1.3402
<i>Classe 2</i>	0.1556	99.6000	0.2444
<i>Classe 3</i>	0.2362	93.0709	6.6929

**Tabela 166 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.3.**



	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	1.1200	97.9200	0.9600
<i>Classe 3</i>	0	99.1667	0.8333

Tabela 167 - Matriz de confusão percentual do conjunto de teste do Experimento G.3.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	68.0415
<i>Por conjunto de teste</i>	54.5778
<i>Médio por classe no conjunto de treinamento</i>	36.0839
<i>Médio por classe no conjunto de teste</i>	32.9178

Tabela 168 - Percentuais de reconhecimento do Experimento G.3.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.0193	0.6770	0.0654

Tabela 169 - Vetor de competência do Experimento G.3.

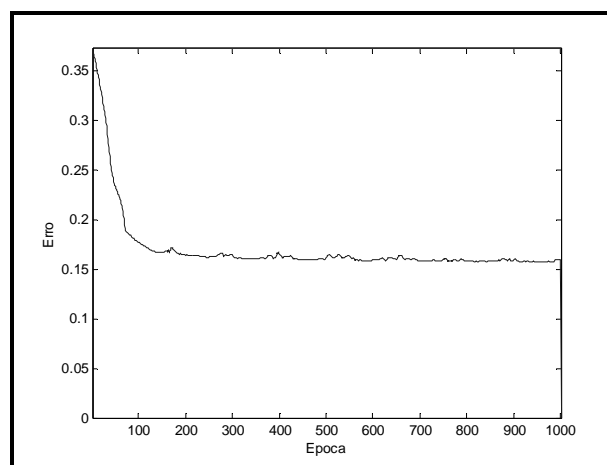


Figura 52 - Curva conceitual de aprendizado do Experimento G.3.

#### 4.2.7.4 Experimento G.4

O Experimento *G.4* apresenta a arquitetura da RNA análoga ao Experimento *G.1*, exceto por apresentar 20 neurônios na camada oculta. Na Tabela 170 pode-se observar a

matriz de confusão percentual do conjunto de treinamento. Na Tabela 171, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 172, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 173 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 53 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	1.7526	97.2165	1.0309
<i>Classe 2</i>	0.0889	99.7778	0.1333
<i>Classe 3</i>	0.2362	93.0709	6.6929

**Tabela 170 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.4.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	100.0000	0
<i>Classe 2</i>	0.6400	98.0000	1.3600
<i>Classe 3</i>	0	99.3750	0.6250

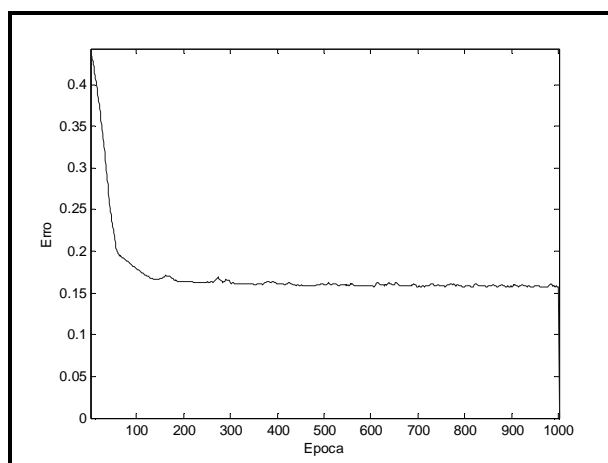
**Tabela 171 - Matriz de confusão percentual do conjunto de teste do Experimento G.4.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	68.1306
<i>Por conjunto de teste</i>	54.5778
<i>Médio por classe no conjunto de treinamento</i>	36.0744
<i>Médio por classe no conjunto de teste</i>	32.8750

**Tabela 172 - Percentuais de reconhecimento do Experimento G.4.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.0173	0.6777	0.0660

**Tabela 173 - Vetor de competência do Experimento G.4.**



**Figura 53 - Curva conceitual de aprendizado do Experimento G.4.**

#### 4.2.7.5 Experimento G.5

O Experimento *G.5* apresenta a arquitetura da RNA análoga ao Experimento *G.1*, exceto por apresentar 25 neurônios na camada oculta. Na Tabela 174 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 175, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 176, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 177 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 54 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	1.8557	97.5258	0.6186
<i>Classe 2</i>	0.1111	99.8667	0.0222
<i>Classe 3</i>	0.3150	93.7008	5.9843

**Tabela 174 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.5.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.1923	99.8077	0
<i>Classe 2</i>	0.8000	98.5600	0.6400
<i>Classe 3</i>	0	100.0000	0

Tabela 175 - Matriz de confusão percentual do conjunto de teste do Experimento G.5.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	68.0712
<i>Por conjunto de teste</i>	54.8000
<i>Médio por classe no conjunto de treinamento</i>	35.9022
<i>Médio por classe no conjunto de teste</i>	32.9174

Tabela 176 - Percentuais de reconhecimento do Experimento G.5.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.0183	0.6772	0.0595

Tabela 177 - Vetor de competência do Experimento G.5.

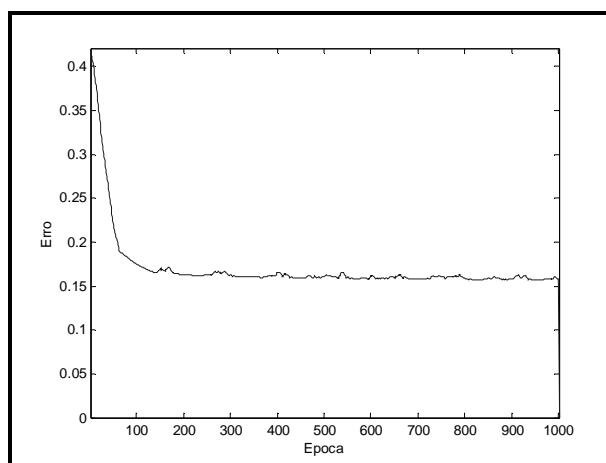


Figura 54 - Curva conceitual de aprendizado do Experimento G.5.

#### 4.2.7.6 Experimento G.6

O Experimento G.6 apresenta a arquitetura da RNA análoga ao Experimento G.1, exceto por apresentar 30 neurônios na camada oculta. Na Tabela 178 pode-se observar a

matriz de confusão percentual do conjunto de treinamento. Na Tabela 179, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 180, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 181 demonstra o nível de especialização através do vetor de competência, no conjunto de treinamento, e, finalmente, na Figura 55 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	1.6495	96.8041	1.5464
<i>Classe 2</i>	0.1111	99.6667	0.2222
<i>Classe 3</i>	0.0787	93.3071	6.6142

**Tabela 178 - Matriz de confusão percentual do conjunto de treinamento do Experimento G.6.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0	99.8077	0.1923
<i>Classe 2</i>	0.6400	98.0000	1.3600
<i>Classe 3</i>	0	99.3750	0.6250

**Tabela 179 - Matriz de confusão percentual do conjunto de teste do Experimento G.6.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	68.0267
<i>Por conjunto de teste</i>	54.5778
<i>Médio por classe no conjunto de treinamento</i>	35.9768
<i>Médio por classe no conjunto de teste</i>	32.8750

**Tabela 180 - Percentuais de reconhecimento do Experimento G.6.**

	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Classe 1</i>	0.0163	0.6771	0.0646

**Tabela 181 - Vetor de competência do Experimento G.6.**

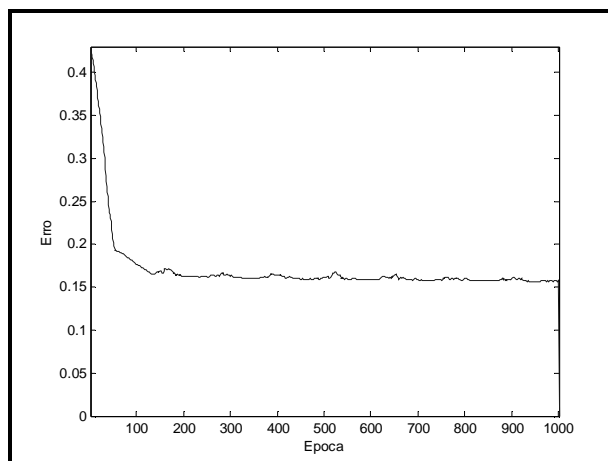


Figura 55 - Curva conceitual de aprendizado do Experimento G.6.

#### 4.2.8 Experimento H: Utilização de Conjuntos Disjuntos de Dados

Este grupo de experimentos foi utilizado para aferir a eficácia do classificador neural em relação à etnia simplificada, sem a classe dominante (branca) e utilizando-se conjuntos disjuntos de dados.

##### 4.2.8.1 Experimento H.1

No Experimento *H.1* foi utilizada a base de dados referente a Etnia Simplificada, classificada em etnia parda e negra sem a etnia branca. A arquitetura da RNA utilizada neste experimento utilizou os dados de entrada de forma que só conjuntos disjuntos tivessem as saídas agrupadas, As classes foram agrupadas em pardos e negros. Com isto, tivemos um total de 22 padrões para toda a base de dados. O conjunto de treinamento foi formado por 50 % dos padrões e o restante formou o conjunto de testes. Os dados de saída não eram estratificados e foram representados por 2 dígitos binários. Foram utilizados 5 neurônios na camada de entrada. Neste experimento, assim como nos que se seguirão, foi feita uma variação na arquitetura da rede, variando-se o número de neurônios na camada oculta. Tomou-se, como ponto de início, utilizar 3 neurônios na camada oculta, com função de transferência *logsig*. Na camada de saída, foram utilizados 2 neurônios, também com função *logsig*.

Na Tabela 182 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 183, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 184, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 185 demonstra o nível de especialização através do vetor de competência no conjunto de treinamento, e, finalmente, na Figura 56 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	95.7143	4.2857
<i>Classe 2</i>	32.5000	67.5000

**Tabela 182 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.1.**

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	57.1429	42.8571
<i>Classe 2</i>	72.5000	27.5000

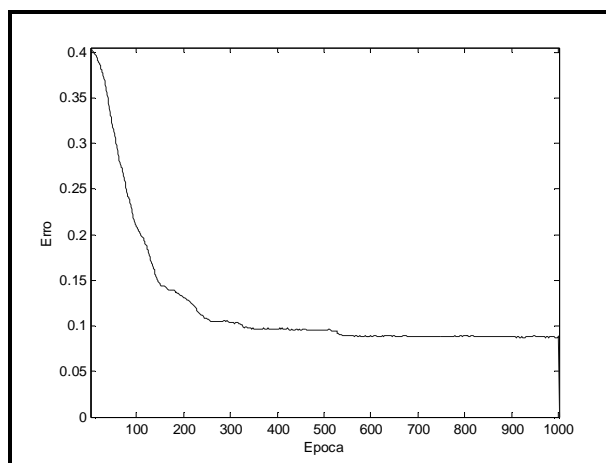
**Tabela 183 - Matriz de confusão percentual do conjunto de teste do Experimento H.1.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	85.4545
<i>Por conjunto de teste</i>	46.3636
<i>Médio por classe no conjunto de treinamento</i>	81.6071
<i>Médio por classe no conjunto de teste</i>	42.3214

**Tabela 184 - Percentuais de reconhecimento do Experimento H.1.**

<i>Classe 1</i>	<i>Classe 2</i>
0.8213	0.6217

**Tabela 185 - Vetor de competência do Experimento H.1.**



**Figura 56 - Curva conceitual de aprendizado do Experimento H.1.**

#### 4.2.8.2 Experimento H.2

O Experimento *H.2* apresenta a arquitetura da RNA análoga ao experimento anterior, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 5. Na Tabela 186 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 187, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 188, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 189 demonstra o nível de especialização através do vetor de competência no conjunto de treinamento, e, finalmente, na Figura 57 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	97.1429	2.8571
<i>Classe 2</i>	15.0000	85.0000

**Tabela 186 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.2.**

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	54.2857	45.7143
<i>Classe 2</i>	72.5000	27.5000

**Tabela 187 - Matriz de confusão percentual do conjunto de teste do Experimento H.2.**



<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	92.7273
<i>Por conjunto de teste</i>	44.5455
<i>Médio por classe no conjunto de treinamento</i>	91.0714
<i>Médio por classe no conjunto de teste</i>	40.8929

Tabela 188 - Percentuais de reconhecimento do Experimento H.2.

<i>Classe 1</i>	<i>Classe 2</i>
0.8982	0.8150

Tabela 189 - Vetor de competência do Experimento H.2.

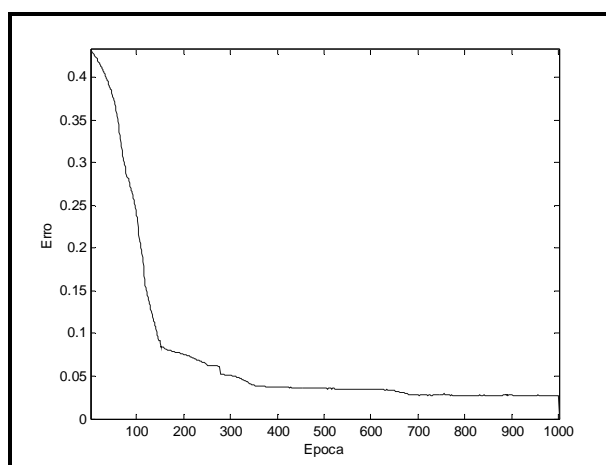


Figura 57 - Curva conceitual de aprendizado do Experimento H.2.

#### 4.2.8.3 Experimento H.3

O Experimento *H.3* apresenta a arquitetura da RNA análoga ao Experimento *H.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 7. Na Tabela 190 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 191, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 192, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 193 demonstra o nível de especialização através do vetor de competência no conjunto de treinamento, e, finalmente, na Figura 58 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	97.1429	2.8571
<i>Classe 2</i>	10.0000	90.0000

Tabela 190 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.3.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	52.8571	47.1429
<i>Classe 2</i>	75.0000	25.0000

Tabela 191 - Matriz de confusão percentual do conjunto de teste do Experimento H.3.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	94.5455
<i>Por conjunto de teste</i>	42.7273
<i>Médio por classe no conjunto de treinamento</i>	93.5714
<i>Médio por classe no conjunto de teste</i>	38.92.86

Tabela 192 - Percentuais de reconhecimento do Experimento H.3.

<i>Classe 1</i>	<i>Classe 2</i>
0.9214	0.8667

Tabela 193 - Vetor de competência do Experimento H.3.

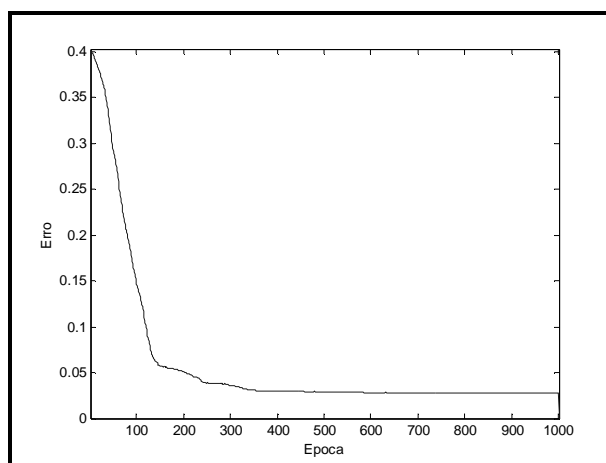


Figura 58 - Curva conceitual de aprendizado do Experimento H.3.

#### 4.2.8.4 Experimento H.4

O Experimento *H.4* apresenta a arquitetura da RNA análoga ao Experimento *H.1*, exceto com uma variação no número de neurônios na camada oculta, que de 3 passou para 10. Na Tabela 194 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 195, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 196, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 197 demonstra o nível de especialização através do vetor de competência no conjunto de treinamento, e, finalmente, na Figura 59 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	95.7143	4.2857
<i>Classe 2</i>	12.5000	87.5000

Tabela 194 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.4.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	62.8571	37.1429
<i>Classe 2</i>	60.0000	40.0000

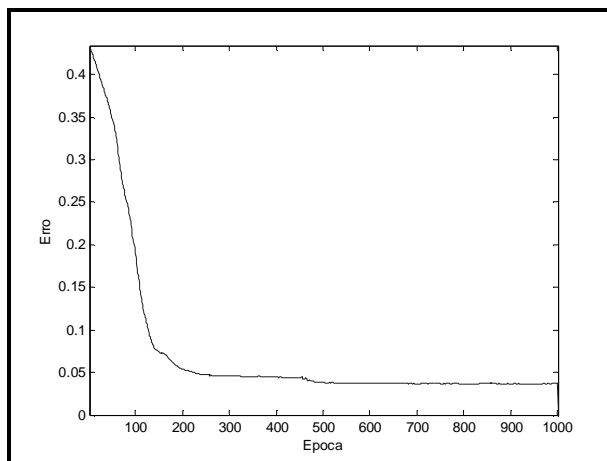
Tabela 195 - Matriz de confusão percentual do conjunto de teste do Experimento H.4.

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	92.7273
<i>Por conjunto de teste</i>	54.5455
<i>Médio por classe no conjunto de treinamento</i>	91.6071
<i>Médio por classe no conjunto de teste</i>	51.4286

Tabela 196 - Percentuais de reconhecimento do Experimento H.4.

<i>Classe 1</i>	<i>Classe 2</i>
0.9002	0.8217

Tabela 197 - Vetor de competência do Experimento H.4.



**Figura 59 - Curva conceitual de aprendizado do Experimento H.4.**

#### 4.2.8.5 Experimento H.5

O Experimento *H.5* apresenta a arquitetura da RNA análoga ao Experimento *H.4*, exceto por apresentar a função de transferência *tansig* na camada oculta. Na Tabela 198 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 199, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 200, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 201 demonstra o nível de especialização através do vetor de competência no conjunto de treinamento, e, finalmente, na Figura 60 apresenta-se a curva percentual do aprendizado para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	91.4286	8.5714
<i>Classe 2</i>	10.0000	90.0000

**Tabela 198 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.5.**

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	55.7143	44.2857
<i>Classe 2</i>	70.0000	30.0000

**Tabela 199 - Matriz de confusão percentual do conjunto de teste do Experimento H.5.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	90.9091
<i>Por conjunto de teste</i>	46.3636
<i>Médio por classe no conjunto de treinamento</i>	90.7143
<i>Médio por classe no conjunto de teste</i>	42.8571

Tabela 200 - Percentuais de reconhecimento do Experimento H.5.

<i>Classe 1</i>	<i>Classe 2</i>
0.8671	0.7933

Tabela 201 - Vetor de competência do Experimento H.5.

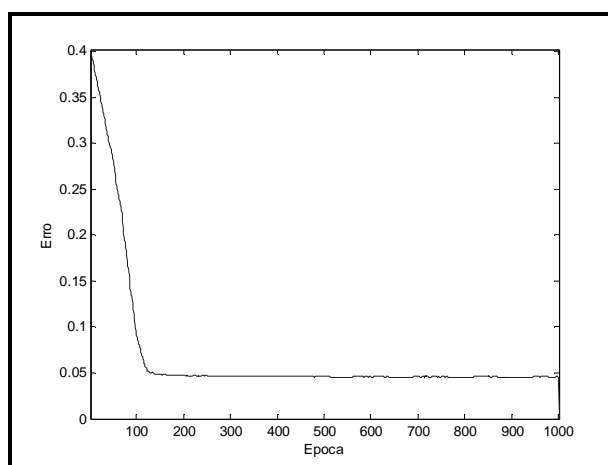


Figura 60 - Curva conceitual de aprendizado do Experimento H.5.

#### 4.2.8.6 Experimento H.6

O Experimento *H.6* apresenta a arquitetura da RNA análoga ao Experimento *H.4*, exceto por apresentar duas camadas ocultas com 10 neurônios cada. Isto também foi feito para se verificar o efeito na performance da rede. Na Tabela 202 pode-se observar a matriz de confusão percentual do conjunto de treinamento. Na Tabela 203, pode-se ver a mesma matriz, só que relacionada ao conjunto de teste. Na Tabela 204, observa-se o percentual de reconhecimento por conjunto de treinamento e teste, e a taxa de reconhecimento médio por classe, tanto do conjunto de treinamento, como também no conjunto de teste. A Tabela 205 demonstra o nível de especialização através do vetor de competência no conjunto de

treinamento, e, finalmente, na Figura 61 apresenta-se a curva percentual do aprendizado, para este experimento.

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	95.7143	4.2857
<i>Classe 2</i>	10.0000	90.0000

**Tabela 202 - Matriz de confusão percentual do conjunto de treinamento do Experimento H.6.**

	<i>Classe 1</i>	<i>Classe 2</i>
<i>Classe 1</i>	58.5714	41.4286
<i>Classe 2</i>	75.0000	25.0000

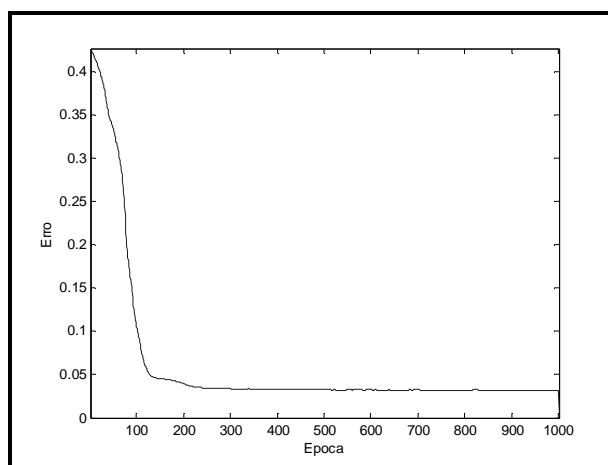
**Tabela 203 - Matriz de confusão percentual do conjunto de teste do Experimento H.6.**

<i>Reconhecimento</i>	<i>Percentual</i>
<i>Por conjunto de treinamento</i>	93.6364
<i>Por conjunto de teste</i>	46.3636
<i>Médio por classe no conjunto de treinamento</i>	92.8571
<i>Médio por classe no conjunto de teste</i>	41.7857

**Tabela 204 - Percentuais de reconhecimento do Experimento H.6.**

<i>Classe 1</i>	<i>Classe 2</i>
0.9099	0.8400

**Tabela 205 - Vetor de competência do Experimento H.6.**



**Figura 61 - Curva conceitual de aprendizado do Experimento H.6.**

### 4.3 COMENTÁRIOS DOS EXPERIMENTOS

#### 4.3.1 Quadro Resumo dos Experimentos

Segue na Tabela abaixo, um quadro com o resumo das variações aplicadas aos experimentos realizados neste trabalho, e nas subseções seguintes os comentários de cada experimento.

<i>EXPERIMENTO</i>	<i>BASE DE DADOS</i>		<i>DADOS DE ENTRADA</i>		<i>DADOS DE SAÍDA</i>		<i>VARIAÇÃO NO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA</i>	<i>VARIAÇÃO NA FUNÇÃO DE TRANSFERÊNCIA</i>	<i>COMITE DE REDES</i>	<i>VARIAÇÃO NO N° DE CAMADAS OCULTAS</i>	<i>RDP</i>
	<i>SANGUE</i>	<i>ETNIA</i>	<i>NÃO ESTRATIF.</i>	<i>ESTRATIF.</i>	<i>NÃO ESTRATIF.</i>	<i>ESTRATIF.</i>					
<i>A</i>	X		X		X		X	X		X	
<i>B</i>		X	X		X		X	X		X	
<i>C</i>		X		X	X		X	X		X	
<i>D</i>		X	X			X	X	X		X	
<i>E</i>		X		X	X				X		X
<i>F</i>		X		X (SEM ETNIA BRANCA)	X		X	X		X	
<i>G</i>		X	X (BINARIZADA)		X		X	X		X	
<i>H</i>		X	X (SEM ETNIA BRANCA)		X		X	X		X	

Tabela 206 - Resumo das variações de técnicas aplicadas aos experimentos.

### 4.3.2 Experimento A

O *Experimento A* foi utilizado para se aferir a estabilidade e robustez da base de dados. Para isto, foi verificada a eficácia na classificação da tipagem sanguínea correta mediante o conjunto de reagentes sanguíneos apresentados. Os resultados dos percentuais de reconhecimento dos conjuntos de treinamento e teste podem ser vistos na Tabela 207. O melhor resultado foi para o experimento A.3, que tinha como característica 7 neurônios na camada oculta. Estes experimentos também obtiveram excelentes resultados em relação ao vetor de competência, que pode ser visto na Tabela 208. Praticamente todas as classes obtiveram valores próximos de 1 o que significa que a generalização (Seção 4.1) está satisfatória.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento A.1</i>	87,7573	81,9070	84,6740	84,9779
<i>Experimento A.2</i>	96,9737	97,3488	94,6208	95,0000
<i>Experimento A.3</i>	98,9766	99,9070	97,1311	97,5000
<i>Experimento A.4</i>	98,9766	99,8605	97,1311	97,4779
<i>Experimento A.5</i>	98,9474	99,8140	97,0944	97,4223
<i>Experimento A.6</i>	97,8655	97,3953	97,0902	97,4779

**Tabela 207 - Comparação dos resultados do Experimento A. Percentuais de reconhecimento.**



	<i>Vetor de Competências</i>			
	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Experimento A.1</i>	0.6972	0.7093	0.8798	0.8477
<i>Experimento A.2</i>	0.9960	0.8159	0.8697	0.9988
<i>Experimento A.3</i>	0.9955	0.8852	0.9701	0.9996
<i>Experimento A.4</i>	0.9960	0.8852	0.9944	0.9823
<i>Experimento A.5</i>	0.9955	0.8837	0.9683	0.9996
<i>Experimento A.6</i>	0.9718	0.9361	0.9000	1.0000

Tabela 208 - Comparação dos resultados do Experimento A. Vetor de Competência.

### 4.3.3 Experimento B

O Experimento B foi utilizado para se verificar a eficácia da RNA em classificar corretamente a etnia simplificada levando em consideração o conjunto de dados de treinamento com os dados não estratificados. A quantidade de padrões da etnia branca superava em até quase 5 vezes a quantidade dos outros padrões. Isto levou a rede a apresentar resultados na faixa de 50 % de reconhecimento quando do submetimento do conjunto de treinamento à rede. Porém o reconhecimento médio por classe se manteve muito baixo mesmo no conjunto de teste, na ordem de 33 %. O *experimento B.3* apresentou o melhor resultado dos testes deste experimento, mas ainda se mostram bem insatisfatórios. Como se pode ver nas matrizes confusões nas Tabelas 46, 47, 50, 51, 54, 55, 58, 59, 62, 63, 66 e 67, tem-se a quase totalidade dos padrões convergindo e generalizando para a classe branca. Os resultados do *Experimento B* em relação aos percentuais de reconhecimento dos conjuntos de treinamento e teste podem ser vistos na Tabela 209. Na Tabela 210, o vetor de competência dos experimentos demonstra, claramente, o nível de especialização da classe branca. As curvas conceituais de aprendizado de todos os testes do experimento B demonstram níveis de erro muito elevados, o que não é satisfatório.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento B.1</i>	67,2661	53,7674	33,6198	33,2184
<i>Experimento B.2</i>	67,2222	53,9535	33,5417	33,3333
<i>Experimento B.3</i>	67,4269	53,8140	33,9626	33,2893
<i>Experimento B.4</i>	67,6316	53,6744	34,3084	33,1609
<i>Experimento B.5</i>	67,3977	53,7209	33,8813	33,1897
<i>Experimento B.6</i>	67,4561	53,5814	34,0042	33,1456

Tabela 209 - Comparação dos resultados do Experimento B. Percentuais de reconhecimento.

	<i>Vetor de Competências</i>		
	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Experimento B.1</i>	0	0.6721	0.0086
<i>Experimento B.2</i>	0	0.6718	0.0063
<i>Experimento B.3</i>	0	0.6732	0.0194
<i>Experimento B.4</i>	0	0.6746	0.0296
<i>Experimento B.5</i>	0.0010	0.6731	0.0156
<i>Experimento B.6</i>	0.0010	0.6734	0.0195

Tabela 210 - Comparação dos resultados do Experimento B. Vetor de Competência.

#### 4.3.4 Experimento C

O *Experimento C* foi utilizado para se verificar a eficácia da RNA em classificar corretamente a etnia simplificada levando em consideração o conjunto de dados de treinamento com os dados estratificados. Estes testes foram feitos para se verificar o efeito da entrada de dados com um conjunto estratificado e seu efeito no classificador. Os resultados demonstrados nas matrizes de confusão nos experimentos C.1 até C.6, tanto de treinamento quanto de teste revelam uma melhor distribuição só que ao invés de se ter resultados especializados em branco, o erro é distribuído pelas classes pardas e negras. Os resultados se

mostraram não substanciais e podem ser observados em resumo na Tabela 211. Na Tabela 212, apresentam-se os vetores de competência atingidos durante os experimentos e, como se pode observar, nenhuma classe apresenta um nível de especialização adequado. As curvas conceituais de aprendizado de todos os testes do *experimento C* também demonstram níveis de erro muito elevados.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento C.1</i>	35,4327	25,1163	34,4661	32,4747
<i>Experimento C.2</i>	38,6951	27,9070	38,4696	33,3144
<i>Experimento C.3</i>	38,9560	27,7674	38,7698	33,2542
<i>Experimento C.4</i>	40,4258	28,5581	40,4784	33,1418
<i>Experimento C.5</i>	37,9052	28,2326	37,9311	32,8282
<i>Experimento C.6</i>	38,8942	27,4884	38,5491	34,2004

Tabela 211 - Comparação dos resultados do Experimento C. Percentuais de reconhecimento.

	<i>Vetor de Competências</i>		
	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Experimento C.1</i>	0.0853	0.0990	0.2823
<i>Experimento C.2</i>	0.2537	0.1796	0.2415
<i>Experimento C.3</i>	0.2562	0.1824	0.2353
<i>Experimento C.4</i>	0.3010	0.2116	0.2141
<i>Experimento C.5</i>	0.2544	0.1926	0.1909
<i>Experimento C.6</i>	0.2521	0.1515	0.2438

Tabela 212 - Comparação dos resultados do Experimento C. Vetor de Competência.

### 4.3.5 Experimento D

O *Experimento D* foi utilizado para se verificar a eficácia da RNA em classificar corretamente a etnia simplificada levando em consideração o conjunto de dados de teste com os dados estratificados. Os dados de entrada não estavam estratificados. Os resultados foram muito parecidos com os alcançados nos *experimentos B*. Os resultados do *Experimento D* em relação aos percentuais de reconhecimento dos conjuntos de treinamento e teste podem ser vistos na Tabela 213. Na Tabela 214 têm-se os vetores de competência que demonstram ainda a alta especialização da classe branca. As curvas conceituais de aprendizado apresentam níveis de erros ainda não aceitáveis.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento D.1</i>	67,3830	36,8153	33,8657	33,2318
<i>Experimento D.2</i>	67,3977	36,8790	33,9668	33,2893
<i>Experimento D.3</i>	67,4415	36,6879	33,9699	33,1034
<i>Experimento D.4</i>	67,5877	36,8790	34,2761	33,2893
<i>Experimento D.5</i>	67,5292	36,8153	34,0885	33,2318
<i>Experimento D.6</i>	67,4854	36,8153	34,0271	33,2184

Tabela 213 - Comparação dos resultados do Experimento D. Percentuais de reconhecimento.

	<i>Vetor de Competências</i>		
	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Experimento D.1</i>	0	0.6730	0.0163
<i>Experimento D.2</i>	0	0.6728	0.0201
<i>Experimento D.3</i>	0	0.6732	0.0195
<i>Experimento D.4</i>	0.0010	0.6746	0.0280
<i>Experimento D.5</i>	0	0.6740	0.0226
<i>Experimento D.6</i>	0.0020	0.6739	0.0187

Tabela 214 - Comparação dos resultados do Experimento D. Vetor de Competência.

### 4.3.6 Experimento E

O *Experimento E* foi utilizado para se verificar a eficácia da RNA em classificar corretamente a etnia simplificada levando em consideração o conjunto de dados de treinamento com os dados estratificados num total de 1456 padrões. Foi utilizada a técnica de RDP inicialmente com fator 5 e variando-se nos experimentos posteriores até o fator 25. A técnica serviu para alternar o conjunto de treinamento a fim de se utilizar o comitê de redes neurais por média. Tentava-se atingir resultados que fossem satisfatórios, pois os resultados dos outros experimentos relacionados à etnia foram insatisfatórios. Os resultados demonstrados nas matrizes de confusão nos *experimentos E.1* até *E.4*, tanto de treinamento quanto de teste revelam uma melhor distribuição só que o resultado do comitê quando deveria classificar para a classe parda estava, ao contrário aumentando o resultado em relação à etnia negra. Isto pode ser devido a erros de avaliação por parte do entrevistador que fez a coleta de sangue ou, realmente, não ser possível realizar a classificação com este conjunto de atributos. De qualquer modo, os resultados não são substanciais, pois como pode ser visto na Tabela 215, tem-se taxas de reconhecimento do conjunto e por classe em um patamar muito baixo. Na Tabela 216, apresentam-se os vetores de competência atingidos durante os experimentos e, como se pode observar, nenhuma classe apresenta um nível de especialização adequado. Apesar das curvas de aprendizado terem apresentado um melhor desempenho do que os apresentados nos experimentos relacionados a etnia até agora, tivemos erros próximo a zero antes de atingir o número total de épocas, a generalização não estava satisfatória e nem mesmo a convergência.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento E.1</i>	38,4066	27,3488	38,2249	33,2295
<i>Experimento E.2</i>	33,0907	33,2093	33,2742	32,5337
<i>Experimento E.3</i>	33,5027	35,5814	33,9441	33,6906
<i>Experimento E.4</i>	33,4547	36,1860	33,8924	33,4408
<i>Experimento E.5</i>	NÃO SE OBTIVE RESULTADOS POR FALTA DE RECURSOS COMPUTACIONAIS			
<i>Experimento E.6</i>				

Tabela 215 - Comparação dos resultados do Experimento E. Percentuais de reconhecimento.

	<i>Vetor de Competências</i>		
	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Experimento E.1</i>	0.2589	0.1650	0.2202
<i>Experimento E.2</i>	0.1353	0.1839	0.1245
<i>Experimento E.3</i>	0.1932	0.1370	0.0790
<i>Experimento E.4</i>	0.1848	0.1542	0.0780
<i>Experimento E.5</i>	NÃO SE OBTIVE RESULTADOS POR FALTA DE RECURSOS COMPUTACIONAIS		
<i>Experimento E.6</i>			

Tabela 216 - Comparação dos resultados do Experimento E. Vetor de Competência.

#### 4.3.7 Experimento F

O *Experimento F* foi utilizado para se verificar a eficácia da RNA em classificar corretamente a etnia simplificada levando em consideração o conjunto de dados de treinamento com os dados estratificados num total de 243 padrões sem a etnia branca. Estes experimentos foram feitos para se verificar o efeito na classificação de uma RNA usando um classificador individual sem a classe dominante. Para se tentar obter melhores resultados foram feitas alterações na arquitetura da rede em relação ao número de neurônios na camada oculta, número de camadas e função de transferência. Os resultados demonstrados nas matrizes de confusão nos *experimentos F.1* até *F.6*, tanto de treinamento quanto de teste

revelam uma forte tendência a classificar resultados que deveriam ser negros a serem erroneamente classificados como pardos. Isto novamente remete ao questionamento se houve erros de avaliação por parte do entrevistador que fez a coleta de sangue ou, realmente, não ser possível realizar a classificação com este conjunto de atributos. Mesmo observando, na Tabela 217, valores mais eficientes dos que os já encontrados até agora, temos nos gráficos das curvas conceituais de aprendizado a constatação que o nível de erro está na faixa de 0,2 a 0,25 o que representa uma taxa insatisfatória. A Tabela 218 demonstra que a classe parda está por demais especializada.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento F.1</i>	56,0082	47,0370	49,9384	49,9174
<i>Experimento F.2</i>	56,0082	48,3951	51,5074	50,5080
<i>Experimento F.3</i>	55,7613	49,1358	51,9075	50,9608
<i>Experimento F.4</i>	56,9959	49,3827	53,2265	51,2699
<i>Experimento F.5</i>	57,6955	49,5062	53,2172	51,7381
<i>Experimento F.6</i>	57,2016	51,9753	55,4262	52,8244

**Tabela 217 - Comparação dos resultados do Experimento F. Percentuais de reconhecimento.**

	<i>Vetor de Competências</i>	
	<i>Classe 1</i>	<i>Classe 2</i>
<i>Experimento F.1</i>	0.0208	0.0407
<i>Experimento F.2</i>	0.1103	0.5153
<i>Experimento F.3</i>	0.1396	0.4958
<i>Experimento F.4</i>	0.1499	0.5032
<i>Experimento F.5</i>	0.1437	0.5376
<i>Experimento F.6</i>	0.2634	0.4590

**Tabela 218 - Comparação dos resultados do Experimento F. Vetor de Competência.**

### 4.3.8 Experimento G

O *Experimento G* foi utilizado para se verificar a eficácia da RNA em classificar corretamente a etnia simplificada levando em consideração o conjunto de dados de treinamento com os dados não estratificados. O conjunto de atributos utilizados como entrada da RNA sofreu uma binarização; isto é, cada reagente foi representado por 3 dígitos binários dando um total de 15 neurônios utilizados na entrada. Isto foi feito a fim de se tentar alcançar melhores resultados com os dados de entrada com uma melhor separação. Isto levou a rede a apresentar resultados na faixa de 65 % de reconhecimento quando do submissão do conjunto de treinamento à rede. Porém o reconhecimento médio por classe se manteve muito baixo mesmo no conjunto de teste, estando na ordem de 33 %. A rede ainda se mostrou extremamente especializada na etnia branca, como pode ser observado na tabela 220. Pode-se observar nas matrizes confusões a forte tendência em classificar somente a etnia branca. Os resultados do *Experimento G* em relação aos percentuais de reconhecimento dos conjuntos de treinamento e teste podem ser vistos na Tabela 219. As curvas conceituais de aprendizado de todos os testes do *Experimento G* continuam demonstrando níveis de erro muito elevados.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento G.1</i>	67,9377	54,6667	35,8378	33,0139
<i>Experimento G.2</i>	67,6261	55,1111	35,0174	33,1950
<i>Experimento G.3</i>	68,0415	54,5778	36,0839	32,9178
<i>Experimento G.4</i>	68,1306	54,5778	36,0744	32,8750
<i>Experimento G.5</i>	68,0712	54,8000	35,9022	32,9174
<i>Experimento G.6</i>	68,0267	54,5778	35,9768	32,8750

Tabela 219 - Comparação dos resultados do Experimento G. Percentuais de reconhecimento.



	<i>Vetor de Competências</i>		
	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>
<i>Experimento G.1</i>	0.0061	0.6746	0.0449
<i>Experimento G.2</i>	0.0092	0.6765	0.0678
<i>Experimento G.3</i>	0.0193	0.6770	0.0654
<i>Experimento G.4</i>	0.0173	0.6777	0.0660
<i>Experimento G.5</i>	0.0183	0.6772	0.0595
<i>Experimento G.6</i>	0.0163	0.6771	0.0646

Tabela 220 - Comparação dos resultados do Experimento G. Vetor de Competência.

#### 4.3.9 Experimento H

O *Experimento H* foi utilizado para se verificar a eficácia da RNA em classificar corretamente a etnia simplificada levando em consideração o conjunto de dados de treinamento com os dados não estratificados sem a etnia branca. Foram utilizados como dados de entrada só conjuntos disjuntos que tivessem as saídas agrupadas. As classes foram agrupadas em pardos e negros perfazendo um total de 22 padrões para toda a base de dados. O gráfico de dispersão dos atributos do problema em questão está representado na Figura 62, onde cada cor representa um padrão diferente. O melhor resultado foi encontrado no experimento H.4 com aproximadamente 51 % de média de reconhecimento por classe no conjunto de teste. Observa-se que mesmo com uma taxa alta de reconhecimento no conjunto de treinamento (o que representa uma boa convergência da rede) não significa que se teria um bom resultado no conjunto de teste (generalização), pois 51 % também não é um resultado satisfatório, ou melhor confiável. Nota-se pelas matrizes confusões que o maior erro está quando o resultado do classificador deveria ser pardo está generalizando para negro. Volta-se novamente para o questionamento: houve erro de avaliação por parte do entrevistador que fez a coleta de sangue sobre a etnia do paciente ou, realmente, não é possível realizar a classificação com este conjunto de atributos. Observa-se nas Tabelas 221 e 222 os percentuais de reconhecimento e os vetores de competência dos *experimentos H.1* até *H.6*. Obteve-se nestes experimentos boas taxas de erro que podem ser observadas nas curvas conceituais de aprendizado.

	<i>Percentual de Reconhecimento do Conjunto de Treinamento</i>	<i>Percentual de Reconhecimento do Conjunto de Teste</i>	<i>Reconhecimento Médio por Classe do Conjunto de Treinamento</i>	<i>Reconhecimento Médio por Classe do Conjunto de Teste</i>
<i>Experimento H.1</i>	85,4545	46,3636	81,6071	42,3214
<i>Experimento H.2</i>	92,7273	44,5455	91,0714	40,8929
<i>Experimento H.3</i>	94,5455	42,7273	93,5714	38,9286
<i>Experimento H.4</i>	92,7273	54,5455	91,6071	51,4286
<i>Experimento H.5</i>	90,9091	46,3636	90,7143	42,8571
<i>Experimento H.6</i>	93,6364	46,3636	92,8571	41,7857

Tabela 221 - Comparação dos resultados do *Experimento H*. Percentuais de reconhecimento.

	<i>Vetor de Competências</i>	
	<i>Classe 1</i>	<i>Classe 2</i>
<i>Experimento H.1</i>	0.8213	0.6217
<i>Experimento H.2</i>	0.8982	0.8150
<i>Experimento H.3</i>	0.9214	0.8667
<i>Experimento H.4</i>	0.9002	0.8217
<i>Experimento H.5</i>	0.8671	0.7933
<i>Experimento H.6</i>	0.9099	0.8400

Tabela 222 - Comparação dos resultados do *Experimento H*. Vetor de Competência.

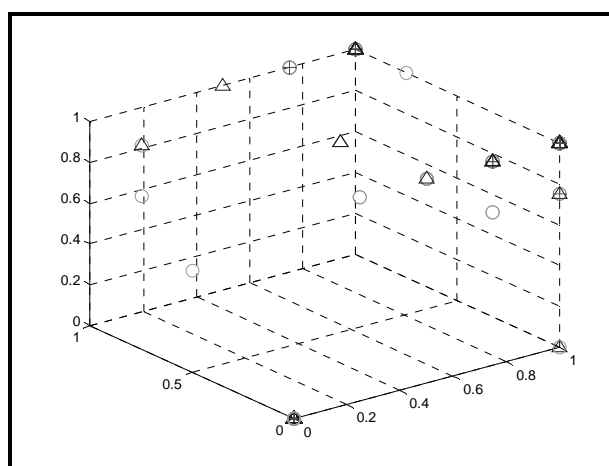


Figura 62 - Gráfico de dispersão dos atributos da base de dados de etnia .

## 5 CONCLUSÕES E TRABALHOS FUTUROS

### 5.1 CONCLUSÕES

Pelos resultados apresentados, verificou-se que, com o atual conjunto de treinamento, não temos resultados significativos que abonem a utilização desta técnica, para a determinação, através do conjunto de reagentes sanguíneos, da etnia simplificada dos indivíduos. Isto aconteceu, mesmo com a aplicação de uma grande variação de esquemas classificadores e de configurações de rede.

O conjunto de dados originais era composto de 911 pacientes, sendo que 575 eram da etnia branca, 175 da etnia parda e 149 da etnia negra. Neste trabalho foram utilizadas técnicas para se estratificar este conjunto e, também, para se tentar conseguir melhores resultados, utilizando-se de conceitos como *RDP*, *Comitês de Redes Neurais* e variações da arquitetura da rede (número de neurônios em cada camada, e, até mesmo, o número de camadas utilizadas).

No primeiro experimento, verificou-se a consistência e estabilidade da classificação da tipagem sanguínea, utilizando-se a base de dados acima. Os resultados se mostraram bastante eficientes. No experimento B se observa uma grande especialização da base de dados na etnia branca, classificando, erroneamente, todos os padrões das outras etnias como sendo brancos. Nos experimentos subsequentes, diversas tentativas foram feitas para se minimizar este problema. Em diversos experimentos (*Experimentos C, D, E e F*) foram utilizadas bases de dados estratificadas, mas, mesmo assim, os resultados não foram melhores. Em outro

experimento (*Experimento F*), a classe referente à etnia branca foi eliminada da base de dados. Isto foi feito para se verificar o efeito, na classificação, em relação às outras duas classes. O resultado continuou sendo inconsistente, apesar de uma melhora no vetor de competência do experimento. Obteve-se, então, uma taxa de reconhecimento entre 50 e 55 %, tanto no conjunto de treinamento, quanto no de teste. Contudo, percebeu-se uma especialização na classe parda. Prevalendo a dúvida se o classificador utilizado com este conjunto de dados é incapaz de atingir uma taxa de generalização aceitável, ou se o atendente que colheu o sangue e entrevistou o paciente classificou a etnia erradamente.

Os resultados dos experimentos são inconclusivos, mas não se pode afirmar que foram “ruins”; alguns fatores devem ser levados em consideração, tais como: i) para  $n$  classes o índice de acertos não ficou abaixo do sorteio ( $1/n$ ); ii) o conjunto de dados podia não ser representativo o suficiente para o problema; iii) os experimentos visavam investigar a possibilidade de correlação (com poucos fatores de entrada) entre sangue e etnia; iv) a classificação étnica simplificada foi realizada de forma empírica, e fugia ao controle e análise deste trabalho.

Mesmo com a aplicação de todas as técnicas apresentadas neste trabalho, foi demonstrado, através dos resultados dos experimentos, que não houve melhora no resultado do classificador final. A média de reconhecimento por classe foi muito baixa e variou de 30 a 40 %. Quando ela atingiu 50 % no conjunto de teste (experimento H), apresentou uma boa taxa relacionada ao conjunto de treinamento, o que indicou que a rede estava convergindo bem para duas classes e poucos padrões, mas a generalização não foi significativa.

Como foi sugerido por AGUIAR (2003), foi feito um estudo mais detalhado do conjunto de atributos utilizados como entrada nas RNA, levando-se em consideração a distribuição espacial destes padrões, em relação às respectivas saídas, e se chegou à conclusão de que só havia 22 padrões realmente disjuntos dentro do universo possível, já que são 5 atributos e cada um podendo variar de 0 a 4; ter-se-ia, então, um total de 3125 padrões, que poderiam compor uma futura base para teste. Analisando a base de dados, percebe-se, pelas informações demonstradas no experimento H, que só 22 conjuntos disjuntos estariam representando as etnias parda e negra.

### 5.1.1 TRABALHOS FUTUROS

Como sugestão para trabalhos futuros, seria interessante avaliar a aplicação de outras técnicas, para se trabalhar o conjunto da base de dados, como, por exemplo, *bootstrap* e *arc-x4*, que poderiam melhorar a participação das classes com pior desempenho (SANTOS, 2001).

Outro fator, que pode ter contribuído de forma negativa para a inconclusividade dos resultados, está na hipótese de ter havido interpretação errônea, ou não metódica, da etnia dos pacientes, por parte dos técnicos que realizaram a coleta de sangue. Uma avaliação visual, não muito precisa, pode concluir que uma determinada pessoa seja considerada branca, ou parda, ou até mesmo negra. A fim de se resolver isto, precisa-se levar em conta fatores relacionados à regionalização, como visto na Seção 2.3. Erros de avaliação, por parte de quem realiza a entrevista, para se definir qual a etnia de uma pessoa, ou por quem se *auto-classifica*, também podem levar a respostas não verdadeiras. Além do mais, a correta classificação da cor na sociedade brasileira é bastante complexa, dada a grande miscigenação das etnias que formaram este povo. Sendo assim, exames morfológicos complexos, ou até mesmo exames de DNA, poderiam ser utilizados, mas ambos mostram-se ainda de custo muito elevado.

Uma direção interessante a seguir, seria o levantamento de uma nova base de dados, que levaria em conta o conjunto de atributos de reagentes, mas relacionados com uma divisão em uma escala mais bem distribuída da etnia. Como exemplo, ao invés de se utilizar somente branco, pardo e negro, poderia se utilizar branco, pardo, negro, pardo-negro (pardo tendendo a negro), pardo-branco (pardo tendendo a branco). O atendente, no banco de sangue que estivesse envolvido na pesquisa, deveria receber um treinamento básico, a fim de que pudesse reconhecer e discernir características morfológicas e fenotípicas básicas de cada etnia. Outra possibilidade seria a de realizar o estudo sobre uma base de dados de uma população de outro país (ou de uma determinada região do Brasil), de variada composição étnica, porém com menor taxa de miscigenação.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABRANCHES, Sérgio. Retrato falado do Brasil. *Veja*, Rio de Janeiro: Editora Abril, n. 46, p. 27, nov 2003.

AGUIAR, Elzencleber F., LIMA, Priscila M. V., OCHI, Luiz S., SANTOS, Claudia, SANTOS, Rafael O. V. Determinação da Influência da Reatividade dos Grupos Sangüíneos ABO e RH na Etnia Simplificada. In: CONGRESSO BRASILEIRO DE COMPUTAÇÃO, 3., 2003, Santa Catarina. Anais do III Workshop de Informática aplicada à saúde, Santa Catarina, 2003. Seção Relato de Pesquisa. 2003.p. 2171-2181.

AURÉLIO, Marco; VELLASCO, Marley; LOPES, Carlos H. Descoberta de Conhecimento e Mineração de Dados. Rio de Janeiro, 1999. Apostila - ICA – Laboratório de Inteligência Computacional Aplicada, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica, Rio de Janeiro. 1999.

BARBOZA, Jorge L. L. *Estudo de Algoritmos de Catgorização Visando Reconhecimento Automático de caracteres manuscritos*. Rio de janeiro, 2002. Projeto Final – Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2002.

BRANCOS têm cinco vezes mais acesso ao ensino superior do que outros brasileiros, aponta IBGE. *O Globo*, Rio de Janeiro, 02 dez 2003. Disponível em: <<http://www.oglobo.com.br/>>. Acesso em: 02 dez 2003.

BERKELEY. In University of California, Berkeley [on line]. Seção Museum of Paleontology. Disponível na Internet: <http://www.ucmp.berkeley.edu/history/lamarck.html>. Capturado em 10 de Outubro de 2003. On Line.

CARICATTI, A. M.; WEIGANG, L. Reconhecimento de Locutores em Língua Portuguesa com Modelos de Redes Neurais e Gaussianos. Proceedings of V Brazilian Conference on Neural Networks, p. 25-30, Abril 2-5, 2001, Rio de Janeiro, RJ, Brasil.

CARVALHO, Afrânio. Observações sobre a organização e execução do Recenseamento Geral do Brasil em 1940. f. 124 e 125. In: Subcomissão do Censo Demográfico: “O quesito cor no censo de 1970”, Coleção Ibegeana, Fundação Instituto Brasileiro de Geografia e Estatística – IBGE, Rio de Janeiro, 1970.

## BIOMANIA.

CENSO 2000: cresce o número de famílias multirraciais. *O Globo*, Rio de Janeiro, 26 dez 2003. Disponível em: <<http://www.oglobo.com.br/>>. Acesso em: 26 dez 2003.

COSTA, Tereza C. A. O princípio classificatório ‘cor’, sua complexidade e implicações para um estudo censitário. *In*: PETRUCCELLI, José Luiz. *A Cor Denominada*: Um estudo do suplemento da Pesquisa Mensal de Emprego de julho de 1998. Rio de Janeiro, 2000. 59p. Textos para discussão – Departamento de População e Indicadores Sociais, Diretoria de Pesquisas, Instituto Brasileiro de Geografia e Estatística.

DEMUTH, H., BEALE, M. Neural Network Toolbox – For Use With Matlab. The Mathworks Inc, Version 2.0a, 1994.

DEMUTH, H., BEALE, M. Neural Network Toolbox – For Use With Matlab. The Mathworks Inc, Version 3.0, 1997.

DERR, Nancy; IRONY, Ilan. Collection of Race and Ethnicity Data in Clinical Trials. U.S Department of Health and Human Services, Guidance for Industry, jan 2003.

FRUDAKIS, Thomas, DNA Witness 2.0 – Validation Studies. Florida: DNAPrint genomics Incorporation, 17 p. 2003.

FILHO, B. D. B.; CABRAL, E. L. L. Um Novo Conceito de Redes Neurais Artificiais para Controle de Sistemas de Reatores Nucleares. Proceedings of V Brazilian Conference on Neural Networks, p. 283-288, Abril 2-5, 2001, Rio de Janeiro, RJ, Brasil.

FREIRE-MAIA, Eleidi A. C. O Genoma Humano – Diversidade e Semelhança. *Jornal de Ciência e Fé*. Rio de Janeiro, dez 2000. Disponível em: <<http://www.cienciaefe.org.br/jornal/arquivo/newton/genoma2.htm>>. Acesso em 06 jun 2003.

FUNDAÇÃO IBGE. Divisão De Documentação. O quesito cor no censo de 1970.

GENÉTICA. *In* BioMédica [on line]. Seção Genética. Disponível na Internet: <http://www.biomania.com.br>. Capturado em 10 jan 2003. On Line.

GTI faz radiografia dos negros nas universidades. *Educação*: Notícias. Brasília: MEC. Boletim 89, 06 jan 2004.

HOMEM moderno “nasceu” negro e africano. *O Globo*, Rio de Janeiro, 11 jun 2003. Disponível em: <<http://www.oglobo.com.br/>>. Acesso em: 11 jun 2003.

KENSKI, Rafael. Afinal, do que é feito o racismo? *Super Interessante*, São Paulo: Editora Abril, ano 17, nº 4, 187º edição, p. 42-50, abr. 2003.

KOVÁCS, Z. L. Redes Neurais Artificiais – Fundamentos e Aplicações. Editora Collegium Cognitio, 2ª Edição, 1996.

LEITE, Marcelo. Segundo Dna Brasileiro, Raça É Só Conceito Social. *Folha de São Paulo*, São Paulo, 17 Dez. 2002. In BioMédica [on line]. Seção Bio News. Disponível na Internet: <http://www.biomania.com.br>. Capturado em 10 jan 2003. On Line.

LIMA, Leila S. Um país em branco e preto. *O Dia*, Rio de Janeiro, 01 dez. 2002. Empregos & Concursos, p.1-2.

MAIO, Marcos C.; SANTOS, Ricardo V. Injetando Sangue no Mito da Democracia Racial? Genética, Relações Raciais e Política no Brasil Contemporâneo. In PENA, Sergio D. J. *Homo Brasilis: Aspectos Genéticos, Lingüísticos, Históricos e Socioantropológicos da formação do Povo Brasileiro*. São Paulo: Editora Funpec, 2002.192 p.

MATLAB. MATLAB, The Language of Technical Computing – Getting Started with Matlab. The Math Works Inc., Version 5.1, 1997.

MARAR, João F., RODRIGUES, Rodrigo F., COSTA, Danilo N. Detecção de falsificações em assinaturas humanas baseada em redes neurais artificiais. In: CONGRESSO BRASILEIRO DE COMPUTAÇÃO, 2., 2002, Santa Catarina. Seção Inteligência Artificial. 2002.

MEDLINEPLUS. In Biblioteca Nacional de Medicina de E.E. U.U. [on line]. Seção Enciclopédia. Disponível na Internet: <http://www.nlm.nih.gov/medlineplus/spanish/ency/article/001166.htm>. Capturado em 10 de Outubro de 2003. On Line.

MEROLA, Ediane, UERJ busca um futuro para o sistema de cotas. *O Globo*, Rio de Janeiro, 20 jul 2003. Rio, p. 33.

PENA, Sergio D. J.; Silva, Denise R. Carvalho; Silva, Juliana Alves. Retrato Molecular do Brasil. *Ciência Hoje*, São Paulo, v. 27, n. 159, p.16-25, 2000.

PENA, Sergio D. J. *Homo Brasilis: Aspectos Genéticos, Lingüísticos, Históricos e Socioantropológicos da formação do Povo Brasileiro*. São Paulo: Editora Funpec, 2002.192 p.

PETRUCCELLI, José Luiz. A Cor Denominada: Um estudo do suplemento da Pesquisa Mensal de Emprego de julho de 1998. Rio de Janeiro, 2000. 59p. Textos para discussão – Departamento de População e Indicadores Sociais, Diretoria de Pesquisas, Instituto Brasileiro de Geografia e Estatística.

RACISMO não tem base na genética. *Jornal do Brasil*, Rio de Janeiro, 13 fev 2001, Seção Ciência. Disponível em <<http://jbonline.terra.com.br/jb/papel/ciencia/2001/02/12/jorcie20010212004.html>>. Acesso em: 06 jun 2003.

RUSSEL, Stuart J.; NORVIG, Peter. Artificial Intelligence: A Modern Approach. Editora Prentice Hall, 1995. 932 p.

SANTOS, Rafael de Oliveira Valle dos. Combinação de Redes Neurais MLP em Problemas de Classificação. Rio de Janeiro, 2001. 106 p. Dissertação de Mestrado – Faculdade de Engenharia, Pontifícia Universidade Católica do Rio de Janeiro.



SHARKEY, A.J.C. Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems. In: SANTOS, Rafael de Olivaes Valle dos. Combinação de Redes Neurais MLP em Problemas de Classificação. Rio de Janeiro, 2001. 106 p. Dissertação de Mestrado – Faculdade de Engenharia, Pontifícia Universidade Católica do Rio de Janeiro..

SISTEMA de cotas em universidades vai corrigir desigualdade histórica, diz ministra. O Globo, Rio de Janeiro, 13 mai 2003, Seção Plantão. Disponível em <<http://oglobo.globo.com/oglobo/plantão/107807238.htm>>. Acesso em: 13 mai 2003.

SUBCOMISSÃO DO CENSO DEMOGRÁFICO: “O quesito cor no censo de 1970”, Coleção Ibegeana, Fundação Instituto Brasileiro de Geografia e Estatística – IBGE, Rio de Janeiro, 1970.

TAFNER, M. A. Reconhecimento de palavras faladas isoladas usando redes neurais artificiais. Santa Catarina, 1996. Monografia da Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina.

ZURADA, J. M. Introduction to Artificial Neural Systems. PWS Publishing Company, 1995.

## ANEXOS

## ANEXO 1 - DISTRIBUIÇÃO DAS CATEGORIAS ESPONTÂNEAS DO QUESITO COR.

	<i>Cor/Raça</i>	<i>Abs</i>	<i>%</i>	<i>% ac</i>
1	AFRICANA	1103	0,00	0,00
2	ALEMAO	11136	0,03	0,04
3	ALOURADO	187	0,00	0,04
4	ALVA	6219	0,02	0,05
5	AMARELA	369192	1,08	1,14
6	AMARELOCLARO	182	0,00	1,14
7	AZUL/BRANCA	789	0,00	1,14
8	BAIANA	182	0,00	1,14
9	BEGE	187	0,00	1,14
10	BEMLOIRA	789	0,00	1,15
11	BOMBOM	542	0,00	1,15
12	BRANCA	18443834	54,17	55,32
13	BRANCMARELA	416	0,00	55,32
14	BRANCAVERMELHADA	729	0,00	55,32
15	BRANCABRASILEIRA	981	0,00	55,33
16	BRANCACLARA	738	0,00	55,33
17	BRANCAEPARDA	789	0,00	55,33
18	BRANCAESCURA	576	0,00	55,33
19	BRANCALEITE	187	0,00	55,33
20	BRANCAMORENA	3707	0,01	55,35
21	BRANCAOUMULATA	542	0,00	55,35
22	BRANCOAZEDO	542	0,00	55,35
23	BRANCOMEDIO	182	0,00	55,35
24	BRANCOMORENOCLARO	2367	0,01	55,36
25	BRANQUINHA	1088	0,00	55,36
26	BRASILEIRA	41933	0,12	55,48
27	BRONZEADA	416	0,00	55,48
28	BUGRE	1536	0,00	55,49
29	CABOCLA	5331	0,02	55,50
30	CABOVERDE	6370	0,02	55,52
31	CAFUSO	2668	0,01	55,53
32	CANELA	2592	0,01	55,54
33	CANELAESCURA	208	0,00	55,54

Tabela 223 - Distribuição das categorias espontâneas do quesito Cor (PETRUCCELLI, p. 46) - (parte 1) .

	<i>Cor/Raça</i>	<i>Abs</i>	<i>%</i>	<i>% ac</i>
34	CANELINHA	182	0,00	55,54
35	CASTANHA	3491	0,01	55,55
36	CASTANHOCLARA	1113	0,00	55,55
37	CEARENSE	192	0,00	55,55
38	CHOCOLATE	556	0,00	55,55
39	CINZA	208	0,00	55,56
40	CLARA	264089	0,78	56,33
41	CLARABRANCA	187	0,00	56,33
42	CLARAPARDA	768	0,00	56,33
43	CLARINHA	916	0,00	56,34
44	CLAROBRAZILEIRO	192	0,00	56,34
45	CORDECANELA	572	0,00	56,34
46	CORDECUIA	192	0,00	56,34
47	CRIOULO	542	0,00	56,34
48	DESCASCADO	192	0,00	56,34
49	EDIFICILDEDIZER	789	0,00	56,34
50	ENCARDIDA	390	0,00	56,35
51	ESCURA	129291	0,38	56,72
52	ESCUROCABOVERDE	182	0,00	56,73
53	FRANCES	192	0,00	56,73
54	GALEGO	3736	0,01	56,74
55	GALEGOBRANCO	208	0,00	56,74
56	INDIA	3779	0,01	56,75
57	INDIANEGRACAFUZA	789	0,00	56,75
58	INDIGENA	39815	0,12	56,87
59	ITALIANA	960	0,00	56,87
60	JAMBO	3662	0,01	56,88
61	JAPONESA	7101	0,02	56,90
62	LATINAAMERICANA	374	0,00	56,90
63	LEITE	182	0,00	56,90
64	LOIRA	15413	0,05	56,95
65	LOIRA CLARA	187	0,00	56,95
66	MARROM	8104	0,02	56,97
67	MARRONCANELA	182	0,00	56,97
68	MEIABRANCA	1331	0,00	56,98
69	MEIOTERMO	192	0,00	56,98
70	MEL	187	0,00	56,98
71	MESTICA	20281	0,06	57,04
72	MESTICA(MORENA CLARA)	187	0,00	57,04
73	MESTICAPOSTIÇA	208	0,00	57,04
74	MISTA	5558	0,02	57,06
75	MISTURADA	1098	0,00	57,06
76	MOREN@OCAFE COM LEITE	182	0,00	57,06
77	MORENA	7097472	20,85	77,91
78	MORENA-CLARA-JAMBO	187	0,00	77,91
79	MORENABEMCLARA	551	0,00	77,91
80	MORENACABOCLA	374	0,00	77,91
81	MORENACABOVERDE	546	0,00	77,91
82	MORENACAFE	182	0,00	77,91
83	MORENACANELA,	416	0,00	77,91
84	MORENACASTANHA	3794	0,01	77,92
85	MORENA CLARA	990607	2,91	80,83
86	MORENA ESCURA	151900	0,45	81,28
87	MORENA JAMBO	4186	0,01	81,29
88	MORENA MAIS PARA AMARELA	182	0,00	81,29

Tabela 224 - Distribuição das categorias espontâneas do quesito Cor (PETRUCCELLI, p. 46) - (parte 2) .

	<i>Cor/Raça</i>	<i>Abs</i>	<i>%</i>	<i>% ac</i>
89	MORENAMEDIA	374	0,00	81,29
90	MORENAMESTICA	395	0,00	81,30
91	MORENAO	187	0,00	81,30
92	MORENAPALIDA	561	0,00	81,30
93	MORENAPARDA	956	0,00	81,30
94	MORENAPRETA	208	0,00	81,30
95	MORENASARA	374	0,00	81,30
96	MORENATRIGUEIRO	208	0,00	81,30
97	MOREMNHA	5912	0,02	81,32
98	MORENINHO-BRANQUINHO	187	0,00	81,32
99	MORENOBRANCO	1765	0,01	81,33
100	MORENOMEDIO	748	0,00	81,33
101	MORENOMULATO	187	0,00	81,33
102	MORENONORMAL	187	0,00	81,33
103	MORENOPALIDO	187	0,00	81,33
104	MORENOQUEIMADO	182	0,00	81,33
105	MULATA	268124	0,79	82,12
106	MULATACLARA	5173	0,02	82,13
107	MULATAESCURA	364	0,00	82,13
108	MULATAMORENA	182	0,00	82,13
109	MULATINHA	1084	0,00	82,14
110	MULATOMEDIO	182	0,00	82,14
111	MULATONEGRA	542	0,00	82,14
112	NEGÃO	724	0,00	82,14
113	NEGRA	1067894	3,14	85,28
114	NEGRAMORENO	187	0,00	85,28
115	NEGRINHO	182	0,00	85,28
116	NEGROCLARO	187	0,00	85,28
117	NEGROMORENO	208	0,00	85,28
118	NEGROPARDO	182	0,00	85,28
119	NEGUINHO	182	0,00	85,28
120	PALIDA	1205	0,00	85,29
121	PARDA	3528735	10,36	95,65
122	PARDA(MORENAESCURA)	2168	0,01	95,66
123	PARDA/MORENA	789	0,00	95,66
124	PARDACLARA	2990	0,01	95,67
125	PARDAESCURA	546	0,00	95,67
126	PARDAMORENACLARA	724	0,00	95,67
127	PARDAO	182	0,00	95,67
128	PARDINHA	542	0,00	95,67
129	PARDOCABOCLO	182	0,00	95,67
130	PELODURO	192	0,00	95,68
131	POLONESA	576	0,00	95,68
132	PORTUGUESA	1523	0,00	95,68
133	POUCOMORENO	208	0,00	95,68
134	PRETA	1448878	4,26	99,94
135	PRETA-NEGRA	182	0,00	99,94
136	PRETINHA	542	0,00	99,94
137	ROXA	561	0,00	99,94
138	RUIVA	5125	0,02	99,96
139	SARARA	12358	0,04	99,99
140	SARARAZADA	182	0,00	99,99
141	SAXAO	187	0,00	99,99
142	TOSTADA	192	0,00	99,99
143	VERMELHA	1892	0,01	100,00
	TOTAL	34045265	100,00	

Tabela 225 - Distribuição das categ. espontâneas do quesito Cor (PETRUCCELLI, 2000, p. 46) - (parte 3)

**ANEXO 2 – CRITÉRIOS DE AGREGAÇÃO DA VARIÁVEL COR PARA AS CATEGORIAS AGRUPADAS.**

<b>CATEGORIA</b>	<b>COMPOSIÇÃO</b>
AMARELA	Amarela, Amarelada, Amarela Clara, Japonesa
BRANCA	Branca, Branquinha, Leite, Alva, Latino-americana, Polonesa, Saxão, Portuguesa, Italiana, Alemã, Francesa
BRANCA+	Branca Amarela, Branca Morena, Branca Brasileira, Branca Avermelhada, Branca Clara, Branca Parda, Branca Escura, Branca Leite, Branca Morena, Branca Mulata, Branca Azeda, Branca Media, Branca Morena, Branca Morena Clara, Meia Branca, Azul Branca
CABO VERDE	Cabo Verde, Morena Cabo Verde, Escura Cabo Verde
CANELA	Canela, Canela Escura, Canelinha, Cor de Canela, Marrom Canela
CASTANHA	Castanha, Castanha Clara
CLARA	Clara, Clara Branca, Clara Parda, Clarinha, Clara Brasileira
ESCURA	Escura, Escura Morena, Escurinha
GALEGA	Galega, Galega Branca
INDÍGENA	Indígena, Índia
JAMBO	Jambo, Morena Clara Jambo, Morena Jambo
LOIRA	Loira, Alourada, Bem Loira, Loira Clara, Loura
MARROM CHOCOLATE	Marrom, Chocolate
MISTIÇA/MISTA	Mestiça, Mista, Mestiça Morena Clara, Mestiça Morena, Mestiça Postiça, Misturada
MORENA	Morena, Bombom, Morena Media, Morena Café com Leite, Morena Cabocla, Morena Canela, Morena Castanha, Morena Mestiça, Morena Trigueiro, Moreninha, Morena Normal, Bronzeada, Pouco Morena, Morena mais para Amarela
MORENA CLARA	Morena Clara, Morena bem Clara, Morena Pálida, Moreninha Branquinha, Morena Branca
MORENA ESCURA	Morena Escura, Morena Parda, Morena Café, Morena Preta, Morena Mulata, Morena Queimada
MULATA	Mulata, Mulata Clara, Mulata Escura, Mulata Morena, Mulatinha, Mulata Media, Mulata Negra
NEGRA	Negra, Negão, Negra Morena, Negra Parda, Negra Clara, Neguinha
PARDA	Parda, Pálida, Parda Morena Clara, Parda Morena Escura, Parda Clara, Parda Morena, Pardinha, Pardão
PRETA	Preta, Preta Negra, Pretinha, Africana
SARARA	Sarara, Morena Sarara
VERMELHA	Vermelha, Ruiva

**Tabela 226 - Agregação da variável Cor por Categoria (PETRUCCELLI, 2000, p. 47).**

**ANEXO 3 – COR ABERTA POR COR FECHADA NA REGIÃO METROPOLITANA DO RIO DE JANEIRO.**

<i>RJ RES. ESP.</i>	<i>Branca</i>	<i>Preta</i>	<i>Amarela</i>	<i>Parda</i>	<i>Indígena</i>	<i>TOTAL</i>
AMARELA	4,35	4,35	69,57	21,74	0,00	100,00
BRANCA	99,12	0,12	0,07	0,65	0,04	100,00
BRANCA	75,00	0,00	0,00	0,00	25,00	100,00
BRASILEIRA	81,82	9,09	0,00	9,09	0,00	100,00
CABOCLA	0,00	0,00	0,00	50,00	50,00	100,00
CAPUSO	0,00	0,00	0,00	100,00	0,00	100,00
CASTANHO	0,00	0,00	0,00	100,00	0,00	100,00
CLARA	87,21	0,00	0,00	12,79	0,00	100,00
ESCURA	0,00	81,55	0,00	16,50	1,94	100,00
INDÍGENA	0,00	0,00	0,00	5,56	94,44	100,00
JAMBO	0,00	0,00	0,00	100,00	0,00	100,00
LOIRA	100,00	0,00	0,00	0,00	0,00	100,00
MARROM/CHOCO	0,00	0,00	0,00	100,00	0,00	100,00
MISTIÇA/MISTA	25,00	0,00	0,00	66,67	8,33	100,00
MORENA	14,91	3,47	0,57	77,40	3,65	100,00
MORENACLA	25,60	1,60	0,00	72,00	0,80	100,00
MORENAESC	0,00	21,43	0,00	64,29	14,29	100,00
MULATA	1,54	16,67	0,00	80,86	0,93	100,00
NEGRA	0,00	91,94	0,00	8,06	0,00	100,00
PARDA	0,97	1,90	0,04	96,86	0,23	100,00
PRETA	0,38	97,72	0,09	1,61	0,19	100,00
SARARA	0,00	0,00	0,00	100,00	0,00	100,00
VERMELHA	0,00	0,00	0,00	0,00	100,00	100,00
TOTAL	58,64	10,73	0,23	29,64	0,77	100,00

**Tabela 227 - Cor Aberta X Cor Fechada - Rio de Janeiro (PETRUCCELLI, 2000, p. 50).**

**ANEXO 4 - DISTRIBUIÇÃO DA POPULAÇÃO POR COR SEGUNDO AS REGIÕES METROPOLITANAS. (PETRUCCELLI, P. 54)**

<i>RM</i>	<i>Branca</i>	<i>Preta</i>	<i>Amarela</i>	<i>Parda</i>	<i>Indígena</i>	<i>Total</i>
Recife	41.83	10.83	1.06	43.94	2.34	100.00
Salvador	21.26	20.62	0.53	56.56	1.03	100.00
Belo Horizonte	48.78	11.59	0.37	37.89	1.36	100.00
Rio de Janeiro	58.63	10.72	0.23	29.65	0.77	100.00
São Paulo	65.43	6.34	2.47	25.19	0.57	100.00
Porto Alegre	83.76	7.56	0.18	7.66	0.85	100.00
TOTAL	54.60	10.90	0.85	32.54	1.10	100.00

**Tabela 228 - Cor X Região Metropolitana (PETRUCCELLI, 2000, P. 54).**

**APÊNDICE – TRABALHO PUBLICADO PELO AUTOR**

AGUIAR, Elzenclever F., LIMA, Priscila M. V., OCHI, Luiz S., SANTOS, Claudia, SANTOS, Rafael O. V. Determinação da Influência da Reatividade dos Grupos Sangüíneos ABO e RH na Etnia Simplificada. In: CONGRESSO BRASILEIRO DE COMPUTAÇÃO, 3., 2003, Santa Catarina. Anais do III Workshop de Informática aplicada à saúde, Santa Catarina, 2003. Seção Relato de Pesquisa. 2003.p. 2171-2181.