

REDES NEURAIIS NA DETERMINAÇÃO DE HIPERTENSÃO ARTERIAL CRÔNICA E
DOENÇA PERIODONTAL USANDO METABÓLITOS SALIVARES COMO
INDICADORES

Adriano Gomes Sabino de Araújo

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: Luiz Pereira Calôba

Carla Martins de Oliveira

Rio de Janeiro

Setembro de 2014

REDES NEURAIS NA DETERMINAÇÃO DE HIPERTENSÃO ARTERIAL CRÔNICA E
DOENÇA PERIODONTAL USANDO METABÓLITOS SALIVARES COMO
INDICADORES

Adriano Gomes Sabino de Araújo

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM
ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Luiz Pereira Calôba, Dr.ing.

Prof.^a Carla Martins de Oliveira, D.Sc.

Prof. José Gabriel Rodriguez Carneiro Gomes, Ph.D.

Prof.^a Ivete Pomarico Ribeiro de Souza, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
SETEMBRO DE 2014

Araújo, Adriano Gomes Sabino de

Redes Neurais na determinação de Hipertensão Arterial Crônica e Doença Periodontal usando Metabólitos Salivares como Indicadores/Adriano Gomes Sabino de Araújo. – Rio de Janeiro: UFRJ/COPPE, 2014.

XIV, 88 p.: il.; 29,7 cm.

Orientadores: Luiz Pereira Calôba

Carla Martins de Oliveira

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Elétrica, 2014.

Referências Bibliográficas: p. 83-88.

1. Redes Neurais. 2. Metabólitos Salivares. 3. Hipertensão Arterial Crônica. 4. Doença Periodontal. I. Calôba, Luiz Pereira *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Agradecimentos

Agradeço primeiramente a Deus, aquele que está comigo em todos os momentos e dá-me sabedoria para alcançar todos os meus objetivos.

Agradeço à minha família, a qual sempre me apoiou e contribuiu de todas as formas para que eu chegasse aonde eu cheguei.

Agradeço a todos que contribuíram de alguma forma para a conclusão desta dissertação, em especial aos meus orientadores Luiz Pereira Calôba, o qual orientou-me na parte de Redes Neurais desde o início até o fim desta dissertação e Carla Martins de Oliveira, que me orientou na parte odontológica.

Agradeço ainda à Profa. Ivete Pomarico Ribeiro de Souza e sua equipe de estudo que autorizaram a utilização de dados, por eles coletados, para que fosse realizado o emprego de Redes Neurais e o desenvolvimento desta dissertação.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

REDES NEURAIIS NA DETERMINAÇÃO DE HIPERTENSÃO ARTERIAL CRÔNICA E
DOENÇA PERIODONTAL USANDO METABÓLITOS SALIVARES COMO
INDICADORES

Adriano Gomes Sabino de Araújo

Setembro/2014

Orientadores: Luiz Pereira Calôba

Carla Martins de Oliveira

Programa: Engenharia Elétrica

As Redes Neurais Artificiais são técnicas computacionais que se baseiam na estrutura e funcionamento do cérebro e, portanto, tentam se aproveitar de características dele tais como a capacidade de aprendizagem e reconhecimento de padrões de maneira eficaz e rápida. Estas técnicas vêm sendo muito utilizadas para realizar tarefas complexas tais como a detecção de objetos e texto em imagens, o reconhecimento voz e diversos tipos de predições. A previsão de doenças é uma aplicação das Redes Neurais Artificiais que vêm sendo constantemente abordada na literatura. Com base nisso, o objetivo deste trabalho é utilizar tais técnicas para tentar prever a Hipertensão Arterial Crônica e a Doença Periodontal a partir de metabólitos salivares de mulheres no período pós-parto identificados pela Ressonância Magnética Nuclear (RMN). Os resultados obtidos na previsão foram razoáveis para a Hipertensão Arterial Crônica e bons para a Doença Periodontal, indicando a existência de uma associação entre tais doenças e os metabólitos salivares.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

NEURAL NETWORKS IN DETERMINATION OF CHRONIC ARTERIAL
HYPERTENSION AND PERIODONTAL DISEASE USING SALIVARY METABOLITES
AS INDICATORS

Adriano Gomes Sabino de Araújo

September/2014

Advisors: Luiz Pereira Calôba

Carla Martins de Oliveira

Department: Electrical Engineering

Artificial Neural Networks are computational techniques that are based on the structure and functioning of the brain and therefore try to take advantage of features such as his ability to learn and recognize patterns quickly and efficiently. These techniques have been widely used to perform complex tasks such as object detection in images and text, voice recognition and various types of predictions. The disease forecasting is one application of neural networks that are being constantly discussed in the literature. Based on this, the objective of this work is to use such techniques to try to predict the Hypertension and Chronic Periodontal Disease from salivary metabolites of women in the postpartum period identified by Nuclear Magnetic Resonance (NMR). The prediction results were reasonable for Chronic Hypertension and good for Periodontal Disease, indicating the existence of an association between these diseases and salivary metabolites.

Sumário

CAPÍTULO 1. INTRODUÇÃO.....	1
1.1 MOTIVAÇÃO.....	1
1.2 OBJETIVO.....	2
1.3 ORGANIZAÇÃO DO TRABALHO.....	3
CAPÍTULO 2. REVISÃO LITERÁRIA.....	4
2.1 CONCEITOS MÉDICOS.....	4
2.1.1 <i>Hipertensão Arterial Crônica</i>	4
2.1.2 <i>Doença Periodontal</i>	5
2.1.3 <i>Metabólitos</i>	6
2.2 CONCEITOS TECNOLÓGICOS.....	7
2.2.1 <i>Ressonância Magnética Nuclear</i>	7
2.2.2 <i>Redes Neurais Artificiais</i>	8
2.2.2.1 <i>Funções de Ativação</i>	11
CAPÍTULO 3. OBTENÇÃO DOS DADOS.....	12
3.1.1 <i>Coleta dos Dados</i>	12
3.1.2 <i>Descrição dos Dados</i>	13
3.1.2.1 <i>Variáveis Obtidas nas Entrevistas e na Medição de Pressão Arterial</i>	13
3.1.2.2 <i>Variáveis obtidas no exame bucal</i>	15
3.1.2.3 <i>Variáveis obtidas via RMN</i>	17
CAPÍTULO 4. ETAPAS PARA A PREVISÃO	18
4.1 PRÉ-PROCESSAMENTO DOS DADOS	18
4.1.1 <i>Representação dos dados</i>	18
4.1.2 <i>Normalização dos dados</i>	19
4.1.3 <i>Preenchimento de Dados Faltantes</i>	20
4.1.4 <i>Identificação das Entradas e Saídas</i>	20
4.1.5 <i>Seleção de características</i>	22
4.1.6 <i>PCA e SVD para visualização dos dados</i>	32
4.1.7 <i>Detecção de Intrusos</i>	37

4.2	DIVISÃO DOS DADOS.....	42
CAPÍTULO 5. EXPERIMENTOS E RESULTADOS.....		44
5.1	HIPERTENSÃO ARTERIAL CRÔNICA.....	44
5.1.1	<i>Experimento 1</i>	44
5.1.2	<i>Experimento 2</i>	48
5.1.3	<i>Experimento 3</i>	52
5.1.4	<i>Experimento 4</i>	54
5.1.5	<i>Experimento 5</i>	56
5.1.6	<i>Experimento 6</i>	58
5.1.7	<i>Experimento 7</i>	62
5.1.8	<i>Experimento 8</i>	64
5.1.9	<i>Experimento 9</i>	67
5.1.10	<i>Experimento 10</i>	69
5.1.11	<i>Conclusão</i>	71
5.2	DOENÇA PERIODONTAL	72
5.2.1	<i>Experimento 1</i>	72
5.2.2	<i>Experimento 2</i>	74
5.2.3	<i>Experimento 3</i>	75
5.2.4	<i>Experimento 4</i>	77
5.2.5	<i>Conclusão</i>	79
CAPÍTULO 6. CONCLUSÃO E TRABALHOS FUTUROS.....		81
BIBLIOGRAFIA		83

Figuras

FIGURA 1: MEDIDOR DE PRESSÃO (ESQUERDA) E ESTETOSCÓPIO (DIREITA).....	5
FIGURA 2: DESENHO ILUSTRATIVO DOS ESTÁGIOS DA DOENÇA PERIODONTAL. FONTE: WWW.COLGATEPROFISSIONAL.COM.BR.....	6
FIGURA 3: ESPECTRO 1H RMN DE SALIVA DE MULHERES COM: (A) HIPERTENSÃO ARTERIAL CRÔNICA E (B) DOENÇA PERIODONTAL. FONTE: PEREIRA ET AL. 2011.....	7
FIGURA 4: (DIREITA) REPRESENTAÇÃO SIMPLES DE UM NEURÔNIO REAL. (ESQUERDA) REPRESENTAÇÃO DE NEURÔNIO NORMALMENTE UTILIZADA PELAS REDES NEURAIS ARTIFICIAIS. RETIRADO DE [FYFE, 2005].....	9
FIGURA 5: REDE MLP COM UMA CAMADA ESCONDIDA E UM ÚNICO NEURÔNIO NA CAMADA DE SAÍDA.	10
FIGURA 6: GRÁFICOS DA CORRELAÇÃO DOS METABÓLITOS SALIVARES COM: (A) A SAÍDA “HIPERTENSO”. (B) A SAÍDA “PA SISTÓLICA”. (C) A SAÍDA “PA DIASTÓLICA”.	26
FIGURA 7: GRÁFICOS DA CORRELAÇÃO DAS VARIÁVEIS DE ENTRADA DA: (A) CONFIGURAÇÃO H3 COM A SAÍDA “HIPERTENSO”. (B) CONFIGURAÇÃO H4 COM A SAÍDA “PA SISTÓLICA”. (C) CONFIGURAÇÃO H4 COM A SAÍDA “PA DIASTÓLICA”.....	27
FIGURA 8: GRÁFICOS DA CORRELAÇÃO DOS METABÓLITOS SALIVARES COM: (A) A SAÍDA “SAUDÁVEL”. (B) A SAÍDA “GENGIVITE”. (C) A SAÍDA “PERIODONTITE”.	29
FIGURA 9: GRÁFICOS DA CORRELAÇÃO DOS ÍNDICES UTILIZADOS PARA DIAGNÓSTICO DA DOENÇA PERIODONTAL COM: (A) A SAÍDA “SAUDÁVEL”. (B) A SAÍDA “GENGIVITE”. (C) A SAÍDA “PERIODONTITE”.....	30
FIGURA 10: GRÁFICOS REFERENTES À EXECUÇÃO DO ALGORITMO: (A) PCA SOBRE A CONFIGURAÇÃO H1 (B) SVD SOBRE A CONFIGURAÇÃO H1 (C) PCA SOBRE A CONFIGURAÇÃO H2 E (D) SVD SOBRE A CONFIGURAÇÃO H2	34
FIGURA 11: GRÁFICOS REFERENTES À EXECUÇÃO DO ALGORITMO: (A) PCA SOBRE A CONFIGURAÇÃO DP1 – GENGIVITE; (B) SVD SOBRE A CONFIGURAÇÃO DP1 – GENGIVITE; (C) PCA SOBRE A CONFIGURAÇÃO DP1 – PERIODONTITE; (D) SVD SOBRE A CONFIGURAÇÃO DP1 – PERIODONTITE (E) PCA SOBRE A CONFIGURAÇÃO DP1 – DOENÇA PERIODONTAL; (F) SVD SOBRE A CONFIGURAÇÃO DP1 – DOENÇA PERIODONTAL.	36
FIGURA 12: VALIDAÇÃO CRUZADA	43

Tabelas

TABELA 1: FUNÇÕES DE ATIVAÇÃO COM SUAS DERIVADAS.	11
TABELA 2: VALORES DAS VARIÁVEIS DE A ATÉ G OBTIDAS NAS ENTREVISTAS PARA OS 5 PRIMEIROS INDIVÍDUOS DO ESTUDO.....	14
TABELA 3: VALORES DAS VARIÁVEIS DE H ATÉ N OBTIDAS NAS ENTREVISTAS PARA OS 5 PRIMEIROS INDIVÍDUOS DO ESTUDO.....	14
TABELA 4: VALORES DAS VARIÁVEIS DE O ATÉ U OBTIDAS NAS ENTREVISTAS PARA OS 5 PRIMEIROS INDIVÍDUOS DO ESTUDO.....	15
TABELA 5: VALORES DAS VARIÁVEIS DE V ATÉ X OBTIDAS NAS ENTREVISTAS PARA OS 5 PRIMEIROS INDIVÍDUOS DO ESTUDO.....	15
TABELA 6: VALORES DAS VARIÁVEIS DE A ATÉ G OBTIDAS NO EXAME BUCAL PARA OS 5 PRIMEIROS INDIVÍDUOS DO ESTUDO.....	16
TABELA 7: VALORES DAS VARIÁVEIS DE H ATÉ O OBTIDAS NO EXAME BUCAL PARA OS 5 PRIMEIROS INDIVÍDUOS DO ESTUDO.....	16
TABELA 8: TRÊS PRIMEIRAS E TRÊS ÚLTIMAS VARIÁVEIS OBTIDAS APÓS A RMN PARA OS 24 INDIVÍDUOS DO ESTUDO.	17
TABELA 9: CONFIGURAÇÕES DE VARIÁVEIS DE ENTRADAS E SAÍDAS UTILIZADAS PARA PREVISÃO DA HIPERTENSÃO ARTERIAL CRÔNICA.....	21
TABELA 10: CONFIGURAÇÕES DE VARIÁVEIS DE ENTRADAS E SAÍDAS UTILIZADAS PARA PREVISÃO DA DOENÇA PERIODONTAL.	22
TABELA 11: VARIÁVEIS SELECIONADAS NA ETAPA DE “SELEÇÃO DE CARACTERÍSTICAS” PARA CADA CONFIGURAÇÃO DA TABELA 9.	31
TABELA 12: VARIÁVEIS SELECIONADAS NA ETAPA DE “SELEÇÃO DE CARACTERÍSTICAS” PARA CADA CONFIGURAÇÃO DA TABELA 10.	32
TABELA 13: ERROS E ACERTOS NOS CONJUNTOS DE TREINAMENTO, VALIDAÇÃO E TESTE PARA CADA CASO DA VALIDAÇÃO CRUZADA (CONFIGURAÇÃO H1).	38
TABELA 14: ERROS PARA CADA ENTRADA EM CADA CASO DA VALIDAÇÃO CRUZADA (CONFIGURAÇÃO H2A).	39
TABELA 15: ERROS E ACERTOS NOS CONJUNTOS DE TREINAMENTO, VALIDAÇÃO E TESTE PARA CADA CASO DA VALIDAÇÃO CRUZADA (CONFIGURAÇÃO H2A).	39
TABELA 16: ERROS PARA CADA ENTRADA EM CADA CASO DA VALIDAÇÃO CRUZADA (CONFIGURAÇÃO H2A).	40
TABELA 17: ERROS E ACERTOS NOS CONJUNTOS DE TREINAMENTO, VALIDAÇÃO E TESTE PARA CADA CASO DA VALIDAÇÃO CRUZADA (CONFIGURAÇÃO H2B).	41
TABELA 18: ERROS PARA CADA ENTRADA EM CADA CASO DA VALIDAÇÃO CRUZADA (CONFIGURAÇÃO H2B).	41

TABELA 19: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 1.	45
TABELA 20: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 1.	45
TABELA 21: RESULTADO DO EXPERIMENTO 1 UTILIZANDO A REDE E1R1 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	46
TABELA 22: RESULTADO DO EXPERIMENTO 1 UTILIZANDO A REDE E1R2 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	47
TABELA 23: RESULTADO DO EXPERIMENTO 1 UTILIZANDO A REDE E1R3 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	48
TABELA 24: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 2.	49
TABELA 25: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 2.	49
TABELA 26: RESULTADO DO EXPERIMENTO 2 UTILIZANDO A REDE E2R1 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	50
TABELA 27: RESULTADO DO EXPERIMENTO 2 UTILIZANDO A REDE E2R2 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	51
TABELA 28: RESULTADO DO EXPERIMENTO 2 UTILIZANDO A REDE E2R3 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	51
TABELA 29: RESULTADO DO EXPERIMENTO 3 UTILIZANDO A REDE E1R1 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	53
TABELA 30: RESULTADO DO EXPERIMENTO 3 UTILIZANDO A REDE E1R2 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	53
TABELA 31: RESULTADO DO EXPERIMENTO 3 UTILIZANDO A REDE E1R3 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	54
TABELA 32: RESULTADO DO EXPERIMENTO 4 UTILIZANDO A REDE E2R1 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	55
TABELA 33: RESULTADO DO EXPERIMENTO 4 UTILIZANDO A REDE E2R2 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	55
TABELA 34: RESULTADO DO EXPERIMENTO 4 UTILIZANDO A REDE E2R3 PARA DIFERENTES NÚMEROS DE NEURÔNIOS.	56
TABELA 35: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 5.	57
TABELA 36: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 5.	57
TABELA 37: RESULTADOS PARA A REDE E5R1 DO EXPERIMENTO 5.	58
TABELA 38: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 6.	59
TABELA 39: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 6.	59

TABELA 40: RESULTADOS PARA A REDE E6R1 DO EXPERIMENTO 6.....	60
TABELA 41: RESULTADOS PARA A REDE E6R2 DO EXPERIMENTO 6.....	60
TABELA 42: RESULTADOS PARA A REDE E6R3 DO EXPERIMENTO 6.....	61
TABELA 43: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 7.	62
TABELA 44: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 6.	62
TABELA 45: RESULTADOS PARA A REDE E7R1 DO EXPERIMENTO 7.....	63
TABELA 46: RESULTADOS PARA A REDE E7R2 DO EXPERIMENTO 7.....	63
TABELA 47: RESULTADOS PARA A REDE E7R3 DO EXPERIMENTO 7.....	63
TABELA 48: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 8.	65
TABELA 49: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 8.	65
TABELA 50: RESULTADOS PARA A REDE E8R1 DO EXPERIMENTO 8.....	65
TABELA 51: RESULTADOS PARA A REDE E8R2 DO EXPERIMENTO 8.....	66
TABELA 52: RESULTADOS PARA A REDE E8R3 DO EXPERIMENTO 8.....	66
TABELA 53: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 9.	67
TABELA 54: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 8.	67
TABELA 55: RESULTADOS PARA A REDE E9R1 DO EXPERIMENTO 9.....	68
TABELA 56: RESULTADOS PARA A REDE E9R2 DO EXPERIMENTO 9.....	68
TABELA 57: RESULTADOS PARA A REDE E9R3 DO EXPERIMENTO 9.....	68
TABELA 58: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 10.	69
TABELA 59: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NAS REDES DO EXPERIMENTO 10.	70
TABELA 60: RESULTADOS PARA A REDE E10R1 DO EXPERIMENTO 10.....	70
TABELA 61: RESULTADOS PARA A REDE E10R2 DO EXPERIMENTO 10.....	70
TABELA 62: RESULTADOS PARA A REDE E10R3 DO EXPERIMENTO 10.....	71
TABELA 63: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 1.	73
TABELA 64: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NA REDE DO EXPERIMENTO 1.....	73
TABELA 65: RESULTADOS DO EXPERIMENTO 1.....	73
TABELA 66: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 2.	74
TABELA 67: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NA REDE DO EXPERIMENTO 2.....	74
TABELA 68: RESULTADOS DO EXPERIMENTO 2.....	75
TABELA 69: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 3.	76

TABELA 70: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NA REDE DO EXPERIMENTO 3.....	76
TABELA 71: RESULTADOS DO EXPERIMENTO 3.....	76
TABELA 72: DETALHES DAS REDES UTILIZADAS NO EXPERIMENTO 4.	77
TABELA 73: VARIÁVEIS DE ENTRADA E SAÍDA UTILIZADAS NA REDE DO EXPERIMENTO 4.....	77
TABELA 74: RESULTADOS DA REDE E4R1 DO EXPERIMENTO 4.	78
TABELA 75: RESULTADOS DA REDE E4R2 DO EXPERIMENTO 4.	78
TABELA 76: RESULTADOS DA REDE E4R3 DO EXPERIMENTO 4.	78

Algoritmos

ALGORITMO 2: PRIMEIRO ALGORITMO PARA SELEÇÃO DE CARACTERÍSTICAS.	24
ALGORITMO 3: SEGUNDO ALGORITMO PARA SELEÇÃO DE CARACTERÍSTICAS.	25

Capítulo 1. Introdução

1.1 Motivação

A hipertensão arterial crônica é uma doença que pode trazer grandes prejuízos à saúde dos indivíduos que a apresentam. O acidente vascular cerebral (AVC), as doenças da retina, arritmias, infarto do miocárdio, doenças vasculares periféricas, insuficiência cardíaca e insuficiência renal são algumas das complicações causadas por esta enfermidade. Quando ela está presente durante a gravidez, pode ocorrer restrição do crescimento fetal, prematuridade, descolamento prematuro da placenta, além do aumento do risco de morte materna ([SASS et al., 2002]). Além disso, a hipertensão arterial crônica pode trazer outra complicação que afeta tanto a saúde da mulher como a do feto: a pré-eclâmpsia. De acordo com SASS et al. (2002), a tendência de mulheres optarem por ter filhos em idades mais avançadas, muitas vezes próximo aos quarenta anos, tem levado ao maior registro da hipertensão arterial crônica intercorrente na gravidez. Em vista dos diversos males que esta doença pode causar à grávida e ao seu filho e devido ao crescente aumento dos casos, é extremamente importante o diagnóstico desta doença e o acompanhamento constante da gravidez das mulheres que sofrem deste mal.

Outra doença que pode causar sérios danos à saúde é a doença periodontal, a qual pode levar à perda de osso alveolar, afrouxamento do tecido conjuntivo, formação de bolsas ao redor dos dentes afetados e à perda destes. Mesmo sendo uma doença bucal, estudos têm revelado a associação da doença periodontal com outras doenças, tais como doença cardíaca coronariana, diabetes mellitus e o AVC ([LÖE, 1993], [JOSHIPURA et al., 2003], [GRAU et al., 2004] e [HUMPHREY *et al.*, 2008]). Em relação ao diabetes mellitus, a doença periodontal é considerada a sexta complicadora, uma vez que diversos estudos mostram que ela ocorre mais e é mais grave em pessoas que apresentam diabete mellitus ([LÖE, 1993]). Já com relação à doença cardíaca coronariana, a doença periodontal é um fator de risco que independe dos outros fatores de risco tradicionais ([HUMPHREY *et al.*, 2008]). Além disso, quando presente em mulheres grávidas, a doença periodontal aumenta o risco de pré-eclâmpsia ([VARSHNEY *et al.*, 2014]), nascimento de bebês prematuros ou com baixo peso ([HAERIAN-ARDAKANI, 2013], [JACOB *et al.*, 2013]).

Dadas as graves consequências da hipertensão arterial crônica e da doença periodontal, é de extrema importância o diagnóstico destas. Pensando nisso, este trabalho utiliza as Redes Neurais Artificiais para tentar prever estas duas doenças a partir da análise de componentes salivares. A saliva total (como é utilizada neste trabalho) é composta não só pelas secreções das glândulas salivares como também por substâncias oriundas do fluido crevicular gengival, secreções brônquicas ou nasais, células epiteliais descamadas, restos de alimentos, microrganismos e o produto de seus metabolismos ([MOURA et al., 2008]), o que a torna um fluido rico e que levanta a possibilidade de ser útil para o diagnóstico não só de doenças localizadas na boca mas como também uma série de outras doenças. Diversos trabalhos anteriores buscaram diagnosticar doenças a partir da saliva ([KIRSCHBAUM et al., 1992], [STRECKFUS et al., 2000], [GRIGOR'EV et al., 2002], [MOURA, 2004], [GOTOH et al., 2005], [MBULAITEYE et al., 2006]). A busca por biomarcadores salivares para diagnóstico e monitoramento de doenças bucais e sistêmicas tem despertado interesse de pesquisadores pelo fato da coleta deste biofluido não ser invasiva e, conseqüentemente, indolor, trazendo mais conforto para o paciente. Ademais, a coleta salivar é de baixo custo e fácil de ser realizada. Doenças que afetam a função das glândulas salivares (cirrose alcoólica, fibrose cística, sarcoidose, diabetes mellitus e doenças do córtex adrenal), doenças com o envolvimento de agentes biológicos infecciosos (cárie dentária, doença periodontal, rubéola, raiva e HIV), doenças neoplásicas malignas (câncer de mama, doenças causadas pelo vírus Epstein-Barr e o carcinoma de células escamosas da cabeça e pescoço), doenças psiquiátricas e doenças auto-imunes (síndrome de Sjögren) já foram objetos de estudo de pesquisadores a fim de diagnosticá-las a partir de biomarcadores encontrados na saliva ([MOURA et al., 2008], [JACOB et al., 2014]).

1.2 Objetivo

Utilizar as Redes Neurais Artificiais do tipo Perceptron Multicamadas para tentar prever a Hipertensão Arterial Crônica e a Doença Periodontal a partir de metabólitos salivares de mulheres no período pós-parto identificados pela Ressonância Magnética Nuclear (RMN).

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma: no Capítulo 2 é apresentada a revisão literária realizada durante esse trabalho. Essa revisão fornece a base necessária para que o leitor tenha uma visão geral sobre as técnicas utilizadas e o contexto dos problemas propostos. No Capítulo 3 são descritos o processo de obtenção dos dados utilizados e a descrição deles, a fim de esclarecer a forma como os dados foram adquiridos e o que cada dado significa. No Capítulo 4 são detalhadas todas as etapas efetuadas a fim de realizar a previsão da hipertensão arterial crônica e da doença periodontal utilizando Redes Neurais Artificiais. No Capítulo 5 são descritos os experimentos efetuados a fim de realizar esta previsão e os resultados destes também são apresentados. Por fim, o Capítulo 6 apresenta as conclusões e os possíveis trabalhos futuros.

Capítulo 2. Revisão Literária

Este capítulo visa descrever os principais conceitos necessários para o entendimento deste trabalho. São descritos os conceitos referentes à parte médica do trabalho (hipertensão arterial crônica, doença periodontal, metabólitos e Ressonância Magnética Nuclear) e um conceito referente à parte tecnológica (Redes Neurais Artificiais). Outros conceitos tecnológicos são descritos no Capítulo 4.

2.1 Conceitos Médicos

São descritos os conceitos referentes à parte médica deste trabalho.

2.1.1 Hipertensão Arterial Crônica

Um indivíduo sofre de hipertensão arterial crônica se a sua pressão sistólica está acima de 140 mmHg ou sua pressão diastólica está acima de 90 mmHg avaliados por um médico em pelo menos dois momentos distintos e com um intervalo maior que quatro horas. Segundo SIBAI (2002), a hipertensão arterial crônica na gravidez é definida como a presença de hipertensão antes da vigésima semana de gestação. Se a hipertensão se manifestar depois da vigésima semana de gestação e não tiver sido constatada antes, ela é chamada de hipertensão gestacional e não hipertensão arterial crônica.

Se não for tratada, a hipertensão arterial crônica pode causar problemas no cérebro, no coração, nos olhos e nos rins. Exemplos destes problemas são: AVC, infarto do miocárdio, arritmias, insuficiência cardíaca, doenças vasculares periféricas, doença da retina e insuficiência renal. De acordo com SIBAI (2002), a hipertensão arterial crônica na gravidez incrementa o risco de pré-eclâmpsia sobreposta, descolamento prematuro da placenta, edema pulmonar, encefalopatia hipertensiva, retinopatia, hemorragia cerebral e insuficiência renal aguda.

Atualmente, a pressão arterial (PA) é medida com o uso de um aparelho chamado esfigmomanômetro (localizado na esquerda da Figura 1) e com o auxílio de um estetoscópio (localizado na direita da Figura 1). Tal método de medição da PA não é simples de ser utilizado e exige que a pessoa seja treinada para que possa realizar a medição. Por esse motivo, uma pessoa sem treinamento não consegue acompanhar a sua pressão arterial

diariamente, a não ser que seja acompanhada por uma pessoa treinada. Os medidores digitais de pressão surgiram a fim de facilitar esse acompanhamento, mas a maioria não é tão eficaz como o medidor manual.



Figura 1: Medidor de pressão (esquerda) e estetoscópio (direita)

2.1.2 Doença Periodontal

Segundo LÖE (1993), a doença periodontal é uma condição inflamatória crônica caracterizada pela destruição dos tecidos periodontais, resultando na perda do ligamento do tecido conjuntivo, perda de osso alveolar e na formação de bolsas patológicas ao redor dos dentes afetados (ilustrados na Figura 2).

Estudos anteriores falam sobre a associação da doença periodontal com outras doenças, tais como a diabetes mellitus ([LÖE, 1993]), doenças cardiovasculares ([BECK *et al.*, 1996], [SCANNAPIECO *et al.*, 2003] e [HUMPHREY *et al.*, 2008]) e AVC ([SCANNAPIECO *et al.*, 2003], [GRAU *et al.*, 2004]). Esta doença afeta grande parte da população nacional e mundial, principalmente a população de regiões mais pobres ([PETERSEN *et al.*, 2005]).

Tendo em vista a gravidade da hipertensão arterial crônica e também da doença periodontal e as vantagens de se utilizar a saliva como meio de diagnóstico para algumas doenças, este trabalho utiliza os metabólitos (seção 2.1.3) presentes na saliva para tentar prevêê-las.

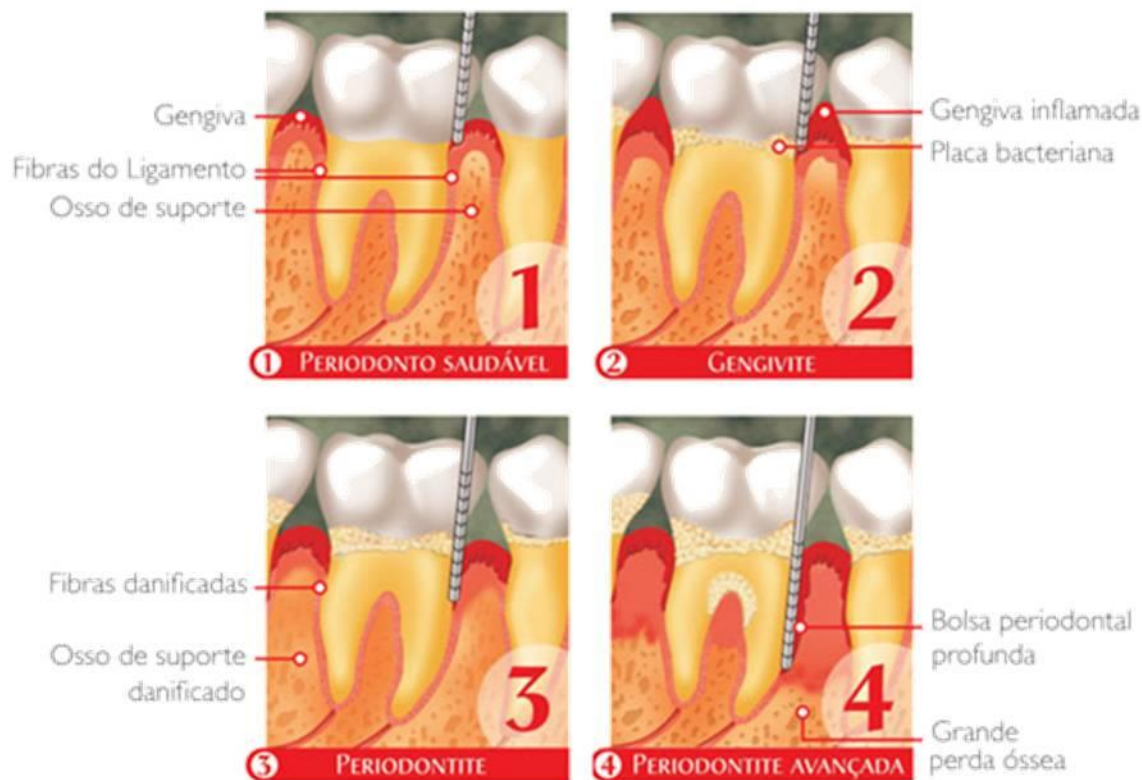


Figura 2: Desenho ilustrativo dos estágios da doença periodontal. Fonte: www.colgateprofissional.com.br.

2.1.3 Metabólitos

Metabólitos são os produtos de reações enzimáticas que ocorrem naturalmente dentro das células e estão presentes nos biofluidos corporais, como: sangue, urina, lágrima, suor e saliva. Este trabalho faz o uso de metabólitos salivares, os quais são metabólitos obtidos da saliva. Diversos trabalhos mostram a relação entre os metabólitos (salivares ou não) e algumas doenças ([MOURA et al., 2008], [MATTA *et al.*, 2010], [JOU *et al.*, 2010], [YEH *et al.*, 2011], [PFAFFE et al., 2011]). MOURA et al. (2008) faz uma revisão da literatura e cita trabalhos que relacionam os metabólitos salivares com doenças, tais como: doença periodontal, câncer de mama, diabetes mellitus, HIV e síndrome de Sjögren. Os metabólitos salivares, identificados por RMN, foram utilizados neste trabalho a fim de prever a hipertensão arterial crônica e a doença periodontal.

2.2 Conceitos Tecnológicos

São descritos os conceitos referentes à parte tecnológica deste trabalho.

2.2.1 Ressonância Magnética Nuclear

De acordo com RMN INFOPÉDIA, a Ressonância Magnética Nuclear é uma técnica analítica para o estudo estrutural de compostos orgânicos e organometálicos, que se baseia na transição entre os diversos níveis energéticos produzida pela absorção de energia de alta frequência (radiofrequência) que tem lugar quando um núcleo atômico com spin não nulo é submetido a um campo magnético estacionário de intensidade própria.

Segundo ALMEIDA (2013), aparelhos de Ressonância Magnética Nuclear (RMN ^1H) utilizam ímãs supercondutores com campos magnéticos muito intensos e pulsos curtos de radiação eletromagnética de alta frequência. Estes pulsos são capazes de provocar a absorção de energia pelos núcleos de ^1H e, como consequência, é gerado um fluxo de pequena corrente numa bobina receptora que envolve a amostra. A corrente é amplificada e o sinal (um pico ou uma série de picos) é apresentado no computador, permitindo a exibição dos sinais do espectro. A Figura 3 ilustra dois espectros resultantes da Ressonância Magnética Nuclear. O espectro (A) é referente a uma mulher com hipertensão arterial crônica e o (B) é referente a uma mulher saudável. O eixo representado na figura representa o deslocamento químico e é medido em ppm.

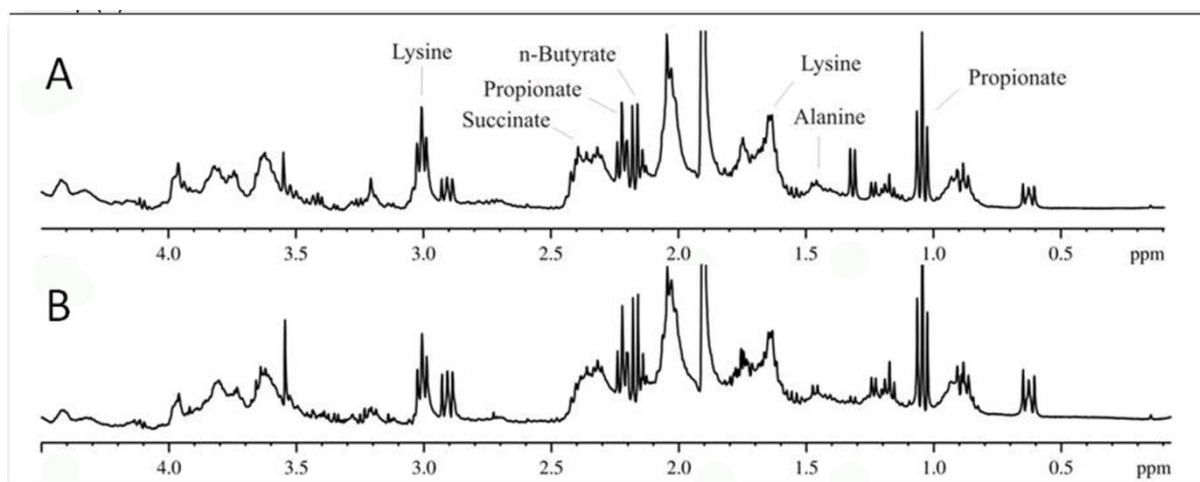


Figura 3: Espectro ^1H RMN de saliva de mulher com Hipertensão Arterial Crônica (A) e de mulher saudável (B). Fonte: Pereira et al. 2011.

Este trabalho faz uso de RNA (técnica descrita na seção 2.2.2) a fim de tentar prever a hipertensão arterial crônica e a doença periodontal através de metabólitos salivares identificados por Ressonância Magnética Nuclear.

2.2.2 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) é um método computacional baseado no funcionamento do cérebro. Este método tenta emular o funcionamento do cérebro e se baseia no fato de que este é capaz de executar tarefas complexas, tais como o reconhecimento de padrões em imagens, de forma eficaz e rápida, enquanto as máquinas não conseguem realizar tais tarefas ou não conseguem ser tão eficientes quanto o cérebro nestas. Duas definições interessantes de RNA são dadas por HAYKIN (1994) e POMMERANZENBAUM (2014). HAYKIN (1994) define redes neurais da seguinte forma: *“Uma rede neural é um processador maciçamente paralelamente distribuído constituído de elementos de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. Ela se assemelha ao cérebro em dois aspectos: (1) o conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem; (2) forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido”*. POMMERANZENBAUM (2014) define RNAs como: *“sistemas compostos por elementos de processamento operando de forma paralela, chamados neurônios, cuja função é determinada pela estrutura da rede, ou melhor, pelos pesos de conexão entre os elementos de processamento, chamados pesos sinápticos, e pelas funções aplicadas por estes elementos”*. A Figura 4 mostra uma representação do neurônio do cérebro (lado esquerdo) e a representação comumente usada pelas RNA (lado direito).

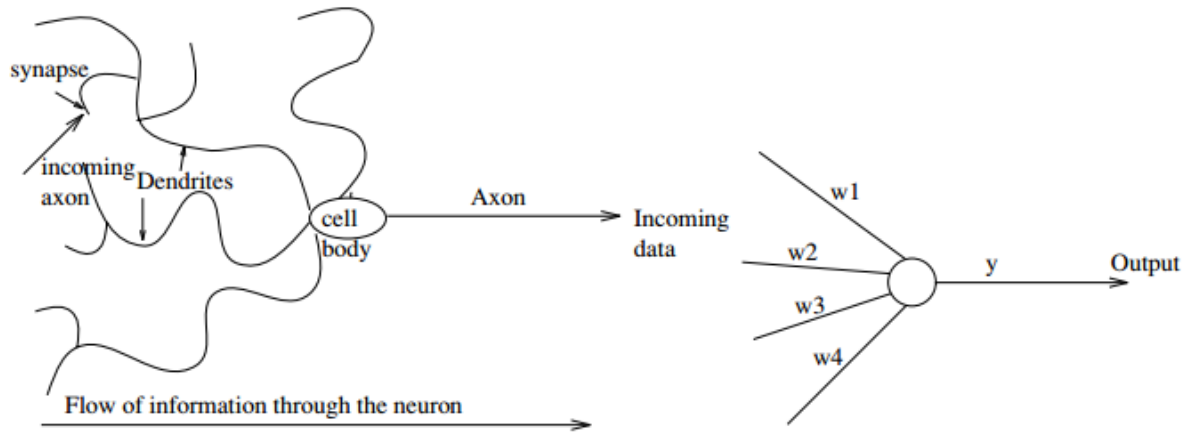


Figura 4: (Direita) Representação simples de um neurônio real. (Esquerda) Representação de neurônio normalmente utilizada pelas redes neurais artificiais. Retirado de [FYFE, 2005]

Em suas definições, HAYKIN (1994) e POMMERANZENBAUM (2014) mencionam os pesos sinápticos e as funções aplicadas pelos elementos de processamento (neurônios). Tais funções são chamadas funções de ativação e as mais utilizadas estão representadas na tabela 1. Para obter o valor da saída do neurônio, a equação 1 é utilizada. Nela, \mathbf{w} representa o vetor de pesos sinápticos, \mathbf{x} representa o vetor de entrada e o valor \mathbf{y} obtido é o valor da saída do neurônio.

$$y = f\left(\sum_i w_i x_i\right) = f(\mathbf{w} \cdot \mathbf{x}) = f(\mathbf{w}^t \mathbf{x}) \quad (1)$$

Existem diversos tipos de redes neurais artificiais, as quais diferem entre si pela forma como os neurônios estão organizados e pela maneira como o sinal é propagado. Neste trabalho, a rede utilizada é a Perceptron Multi-Camadas (MLP, do inglês, *Multi-Layer Perceptron*). Ela é uma RNA em que os neurônios são organizados em camadas e os dados são propagados sempre para frente (*feedforward*), ou seja, os neurônios não formam ciclos. Um exemplo de MLP é apresentado na Figura 5, onde estão embutidos os conceitos de camada de entrada, camada escondida ou oculta e camada de saída. A camada de entrada é formada pelos sinais de entrada, as camadas escondidas ou ocultas são as camadas de neurônios localizadas entre a camada de entrada e a de saída e a camada de saída é formada pelos neurônios que fornecem as saídas da rede.

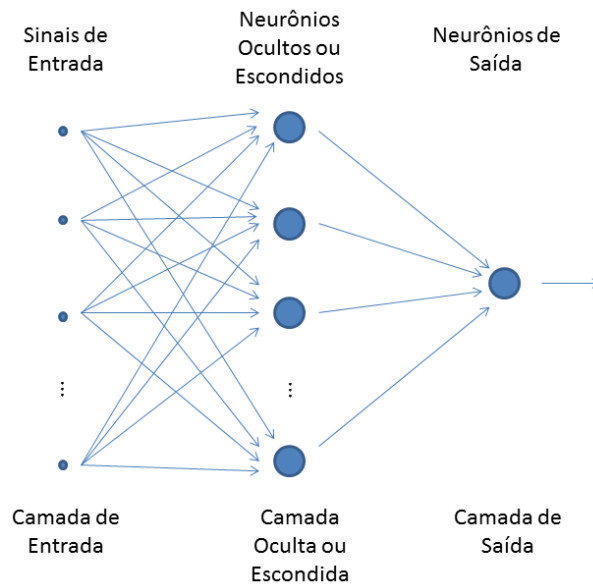


Figura 5: Rede MLP com uma camada escondida e um único neurônio na camada de saída.

As RNA, semelhantemente ao cérebro, são capazes de aprender sobre o ambiente onde estão inseridas. Este processo de aprendizagem é possibilitado com os chamados algoritmos de treinamento, os quais utilizam exemplos de entrada para ajustar os pesos sinápticos da rede aos dados. Com o uso desse processo, as RNA conseguem generalizar os dados e, portanto, fornecer saídas para entradas não vistas anteriormente. O processo de treinamento pode ser supervisionado ou não supervisionado. No treinamento supervisionado, as entradas dos exemplos são apresentadas à RNA e as saídas verdadeiras são comparadas com as saídas fornecidas pela RNA e utilizadas para ajustar os pesos sinápticos, realizando então um mapeamento entrada-saída. Já no treinamento não supervisionado, apenas as entradas são apresentadas e o algoritmo de treinamento é capaz de agrupar os dados de acordo com os padrões presentes nos dados. Neste trabalho, é utilizado o treinamento supervisionado retropropagação do erro (*error backpropagation* ou simplesmente *backpropagation*), o qual é um dos mais utilizados na literatura, é simples de ser implementado e comprovadamente eficaz. Tal algoritmo retropropaga os erros e se baseia no algoritmo gradiente descendente a fim de calcular os valores necessários para ajustar os pesos da rede ([HAYKIN, 1994]).

Quando se constrói uma rede MLP, surge a seguinte pergunta: qual a quantidade de camadas ideal e o número de neurônios nestas camadas? Em relação à quantidade de camadas

ocultas necessárias em uma rede MLP, HAYKIN (1994) afirma que uma única camada oculta é suficiente para uma MLP computar uma aproximação para um dado conjunto de treinamento. Já em relação ao número de neurônios necessários nesta camada oculta, a melhor maneira de determiná-lo é realizando testes e vendo qual a quantidade que fornece os melhores resultados. Quando se trata da quantidade de neurônios da camada de saída, esta sempre é conhecida, já que ela é igual ao número de saídas que se deseja prever.

2.2.2.1 Funções de Ativação

Como visto anteriormente, as saídas dos neurônios são computadas utilizando a equação 1. Nela, é utilizada a função de ativação. As mais utilizadas são a tangente hiperbólica (2), a logística (3), a linear (4) e a função degrau (5), as quais são apresentadas abaixo.

$$f(x) = \tanh(bx) = \frac{1 - e^{-2bx}}{1 + e^{-2bx}} \quad (2)$$

$$f(x) = \frac{1}{1 + \exp(-ax)} \quad (3)$$

$$f(x) = x \quad (4)$$

$$f(x) = 0, \quad \text{se } x < 0 \quad (5)$$

$$f(x) = 1, \quad \text{se } x \geq 0$$

Como no algoritmo *backpropagation* é necessário utilizar a derivada da função de ativação, a tabela 1 apresenta as quatro funções de ativação apresentadas acima e as suas respectivas derivadas.

Tabela 1: Funções de Ativação com suas derivadas.

Nome	Função	Derivada
Tangente hiperbólica	$f(x) = \tanh(bx) = \frac{1 - e^{-2bx}}{1 + e^{-2bx}}$	$f'(x) = b(1 - f(x) * f(x))$
Logística	$f(x) = \frac{1}{1 + \exp(-ax)}$	$f'(x) = af(x)(1 - f(x))$
Linear	$f(x) = x$	$f'(x) = 1$
Degrau	$f(x) = 0, \quad \text{se } x < 0$ $f(x) = 1, \quad \text{se } x \geq 0$	$f'(x) = +\infty, \quad \text{se } x = 0$ $f'(x) = 0, \quad \text{se } x \neq 0$

Capítulo 3. Obtenção dos Dados

Os dados utilizados para o presente trabalho foram coletados por Pereira L. (2011) e fazem parte da dissertação de mestrado em Odontologia (área de concentração: odontopediatria) da Faculdade de Odontologia da Universidade Federal do Rio de Janeiro. O estudo inicial, realizado por Pereira L. (2011), foi aprovada pelo Comitê de Ética em Pesquisa da Secretaria Municipal de Saúde e Defesa Civil do Rio de Janeiro (número 173/09).

Tais dados foram disponibilizados pelo pesquisador responsável para serem processados através de RNA a fim de tentar prever a Hipertensão Arterial Crônica e a Doença Periodontal.

3.1.1 Coleta dos Dados

De acordo com Pereira L. (2011), para a coleta dos dados, foram selecionadas mulheres saudáveis e hipertensas que encontravam-se no período pós-parto. Estas mulheres foram entrevistadas com a finalidade de obter informações sobre a saúde, medicação, gestação, dieta e hábitos de higiene bucal. Além disso, foram medidos a pressão arterial, o peso e a altura de cada mulher e o índice de massa corporal (IMC) foi calculado ([DEURENBERG et al, 1999]).

Após as entrevistas e medições, foi realizada a coleta das amostras salivares das mulheres selecionadas, as quais foram orientadas para que não comessem nem bebessem por 2 horas antes da coleta. Para realização da coleta salivar, foi pedido para as mulheres cuspirem dentro de um tubo de coleta graduado.

As amostras salivares coletadas foram processadas para separação do *pellet* (onde se encontram resíduos alimentares, células descamadas e bactérias) e sendo aproveitado apenas o sobrenadante salivar (porção da saliva que apresenta os componentes bioquímicos, como os metabólitos, por exemplo). As amostras de sobrenadante salivar foram armazenadas a -80°C até o momento da análise por RMN, a qual foi realizada no Centro Nacional de Ressonância Magnética Nuclear Jiri Jonas (CCS/UFRJ), utilizando o espectrômetro 400MHz (Bruker, Alemanha).

Vale ressaltar que também foi realizado exame bucal das mulheres com a finalidade de avaliar a presença de doença cárie e a condição periodontal das mulheres. Para isso, foi

realizado o exame periodontal que compreendeu o índice de placa (IP), índice gengival (IG), presença de cálculo dental, de supuração, de sangramento à sondagem (BOP, do inglês, *bleeding on probing*), medição da profundidade de bolsa (PPD, do inglês, *periodontal pocket depth*) e nível clínico de inserção (NCI). Todas estas avaliações foram feitas em 6 diferentes sítios de cada dente permanente, exceto nos terceiros molares. Os dados obtidos através do exame periodontal permitiram o diagnóstico periodontal dos sujeitos do estudo. Para isso, foi aplicada a classificação sugerida por SILVA-BOGHOSSIAN et al (2011) onde o sujeito pode ser diagnosticado com saúde periodontal, gengivite, periodontite crônica e periodontite agressiva.

3.1.2 Descrição dos Dados

Este capítulo descreve todas as variáveis coletadas seja através das entrevistas, do exame bucal e da medição de pressão arterial, de peso, altura, IMC ou através da RMN da saliva dos indivíduos. As variáveis são descritas em lotes de 7 e os valores de cada variável do lote são mostrados em uma tabela para os 5 primeiros indivíduos do estudo.

3.1.2.1 Variáveis Obtidas nas Entrevistas e na Medição de Pressão Arterial

- a. “Código”: o código que identifica a amostra.
- b. “Hipertenso”: identifica se o indivíduo sofre de hipertensão arterial crônica ou não.
- c. “Idade”: a idade do indivíduo em anos.
- d. “Dias pós-parto”: o tempo (em dias) decorrido desde o parto até o momento da entrevista e coleta da amostra salivar.
- e. “Tempo de diagnóstico”: tempo (em meses) decorrido desde o momento do diagnóstico da hipertensão arterial crônica.
- f. “Início da hipertensão”: se a hipertensão arterial crônica se iniciou antes, durante ou após a gestação.
- g. “Pré-eclâmpsia”: presença desta complicação.

Tabela 2: Valores das variáveis de a até g obtidas nas entrevistas para os 5 primeiros indivíduos do estudo.

COD	HIPERTENSA	IDADE	DIAS PÓS-PARTO	TEMPO DE DIAGNÓSTICO (Meses)	INÍCIO DA HIPERTENSAO	PRÉ-ECLAMPSIA
814	SIM	23	30	-----	ANTES	SIM
815	SIM	23	30	-----	ANTES	NÃO
816	SIM	38	26	-----	ANTES	NÃO
818	SIM	35	20	36	ANTES	NÃO
819	SIM	31	51	-----	ANTES	NÃO

- h. “Eclâmpsia”: presença desta complicação.
- i. “Uso de medicamentos”: se a pessoa faz uso de medicamento.
- j. “Med pré”: qual medicamento usou antes da gravidez.
- k. “Med durante”: qual medicamento usou durante a gravidez.
- l. “Med pós”: qual medicamento usou após a gravidez.
- m. “Diabetes”: se a mulher, além da hipertensão arterial crônica, também apresentava diabetes mellitus.
- n. “Hipoglicemiante”: se a mulher com diabetes mellitus fazia uso de hipoglicemiantes orais (medicamento).

Tabela 3: Valores das variáveis de h até n obtidas nas entrevistas para os 5 primeiros indivíduos do estudo.

ECLAMPSIA	USO DE MEDICAMENTOS	MED PRÉ	MED DURANTE	MED PÓS	DIABETES	HIPOGLICEMIANTES
NÃO	SIM	-----	Metildopa	Captopril	NÃO	-----
NÃO	NÃO	-----	-----	-----	NÃO	-----
NÃO	SIM	-----	Metildopa, Predinison	Captopril	NÃO	-----
NÃO	SIM	Captopril	Metildopa,	Captopril	NÃO	-----

- o. “Insulina”: se a mulher com diabetes mellitus fazia uso de insulina.
- p. “Tempo de gestação”: tempo (em semanas) de duração da gestação.
- q. “Intercorrência”: se houve algum problema de saúde durante a gestação.
- r. “Parto”: se o parto foi normal ou cesariana.
- s. “PA sistólica”: valor da pressão arterial sistólica.
- t. “PA diastólica”: valor da pressão arterial diastólica.
- u. “Peso pós-parto”: peso após o parto (em quilogramas).

Tabela 4: Valores das variáveis de o até u obtidas nas entrevistas para os 5 primeiros indivíduos do estudo.

INSULINA	TEMPO DE GESTAÇÃO	INTERCORRÊNCIA	PARTO	PA SISTÓLICA	PA DIASTÓLICA	PESO PÓS-PARTO
-----	-----	NÃO	CESARIANA	120	80	58.2
-----	-----	NÃO	CESARIANA	130	60	122.0
-----	-----	NÃO	CESARIANA	140	80	88.8
-----	40	NÃO	CESARIANA	160	120	63.7
-----	-----	NÃO	CESARIANA	130	80	97.0

- v. “Altura”: A altura do indivíduo (em metros).
- w. “IMC”: Valor do índice de massa corporal do indivíduo.
- x. “Classificação IMC”: Classificação do IMC de acordo com DEURENBERG et al (1999). Abaixo do peso (abaixo de 18,5), peso normal (18,5 – 24,9), sobrepeso (25 – 29,9), obesidade Grau 1 (30 – 34,9), obesidade Grau 2 (35 – 39,9) e obesidade Grau 3 (a partir de 40).

Tabela 5: Valores das variáveis de v até x obtidas nas entrevistas para os 5 primeiros indivíduos do estudo.

ALTURA	IMC	CLASSIFICAÇÃO IMC
1.55	18.78	NORMAL
1.73	35.26	OBESIDADE 2
1.63	27.24	SOBREPESO
1.5	21.23	NORMAL
1.69	28.7	SOBREPESO

3.1.2.2 Variáveis Obtidas no Exame Bucal

As variáveis de “a” até “l” são resultados do exame periodontal e algumas delas são utilizadas para realizar o diagnóstico periodontal. Já as variáveis “n” e “o” estão relacionadas à cárie dentária. As variáveis de “g” até “l” se baseiam na mesma classificação realizada em KOSHY et al. (2005), ou seja, a PBS e o NCI são classificados em rasos, médios e avançados.

- a. “Profundidade de bolsa à sondagem” (PBS): Profundidade de bolsa (medida em milímetros).
- b. “Nível clínico de inserção” (NCI): Nível clínico de inserção (medido em milímetros).

- c. “Placa”: presença de placa dentária (porcentagem de sítios).
- d. “Sangramento à sondagem”: Presença de sangramento à sondagem (porcentagem de sítios).
- e. “Supuração”: Presença de supuração (porcentagem de sítios).
- f. “Cálculo”: Presença de cálculo dentário (porcentagem de sítios).
- g. “PBS rasa”: Porcentagem de sítios com $PBS < 4$.

Tabela 6: Valores das variáveis de a até g obtidas no exame bucal para os 5 primeiros indivíduos do estudo.

PROFUNDIDADE DE BOLSA (mm)	NÍVEL DE PERDA DE INSERÇÃO (mm)	PLACA (% de sítios)	SANGRAMENTO A SONDAÇÃO (% de sítios)	SUPURAÇÃO (% de sítios)	CÁLCULO (% de sítios)	PBS RASA (% de sítios)
1.87	1.87	50.0	13.7	0.0	0.0	100.0
2.04	2.04	51.2	11.9	0.0	4.2	95.2
2.35	2.35	91.7	15.5	0.0	42.9	88.1
2.08	2.07	42.9	3.6	0.0	20.2	94.0
1.83	1.89	4.2	7.1	0.0	1.8	97.0

- h. “PBS média”: Porcentagem de sítios com $4 \leq PBS \leq 6$.
- i. “PBS avançada”: Porcentagem de sítios com $PBS > 6$.
- j. “NCI raso”: Porcentagem de sítios com $NCI < 4$.
- k. “NCI médio”: Porcentagem de sítios com $4 \leq NCI \leq 6$.
- l. “NCI avançado”: Porcentagem de sítios com $NCI > 6$.
- m. “Diagnóstico periodontal”: O diagnóstico periodontal (saúde gengival, gengivite, periodontite crônica ou periodontite agressiva).
- n. “CPOD”: Índice de cárie (CPOD – dentes cariados, perdidos e obturados).
- o. “Dentes cariados”: Número de dentes cariados.

Tabela 7: Valores das variáveis de h até o obtidas no exame bucal para os 5 primeiros indivíduos do estudo.

PBS MÉDIA (% de sítios)	PBS AVANÇADA (% de sítios)	NCI RASO (% de sítios)	NCI MÉDIO (% de sítios)	NCI AVANÇADO (% de sítios)	DIAGNÓSTICO PERIODONTAL	CPOD	DENTES CARIADOS
0.0	0.0	100.0	0.0	0.0	GENGIVITE	7.0	5.0
4.8	0.0	95.2	4.8	0.0	PERIODONTITE	2.0	2.0
11.9	0.0	88.1	11.9	0.0	PERIODONTITE	0.0	0.0
6.0	0.0	94.0	6.0	0.0	PERIODONTITE	9.0	0.0
3.0	0.0	95.8	4.2	0.0	PERIODONTITE	4.0	1.0

3.1.2.3 Variáveis Obtidas via RMN

As variáveis obtidas na ressonância magnética nuclear são referentes aos deslocamentos químicos entre 0.495 ppm e 4.485 ppm de 0.03 em 0.03 ppm, totalizando 134 variáveis. As variáveis obtidas na RMN foram nomeadas como x1, x2,..., x134. A tabela 8 mostra os dados relativos aos valores das 3 primeiras e 3 últimas variáveis relativos às amostras salivares dos 24 indivíduos (mulheres) do estudo. A primeira linha identifica as variáveis e, entre parênteses, estão os deslocamentos químicos relacionados àquelas variáveis.

Tabela 8: Três primeiras e três últimas variáveis obtidas após a RMN para os 24 indivíduos do estudo.

CÓDIGOS	x1 (4.48500)	x2 (4.45500)	x3 (4.42500)	...	X132 (0.55500)	x133 (0.52500)	x134 (0.49500)
saliva_814	0.01947746	0.01842793	0.01831410	...	0.00107923	0.00091566	0.00086208
saliva_815	0.00880174	0.00834230	0.00803223	...	0.00001435	-0.00001476	-0.00000839
saliva_816	0.00260813	0.00232026	0.00310151	...	0.00006408	0.00009643	0.00012178
saliva_818	0.00291637	0.00198740	0.00470462	...	0.00004694	0.00000895	0.00002494
saliva_819	0.03178617	0.02923092	0.02719360	...	-0.00172510	-0.00171206	-0.00157728
saliva_820	0.01444100	0.01406902	0.01409966	...	-0.00004200	-0.00005870	-0.00007973
saliva_822	0.00749969	0.00573081	0.00495015	...	-0.00088056	-0.00069849	-0.00055611
saliva_824	0.00689063	0.00569404	0.00752911	...	0.00145235	0.00125326	0.00118393
saliva_826	0.02031015	0.01944251	0.01907764	...	0.00019837	0.00017446	0.00018024
saliva_829	0.00174891	0.00145793	0.00243755	...	0.00023503	0.00019848	0.00019776
saliva_831	0.00252341	0.00179115	0.00336133	...	0.00018104	0.00017323	0.00018430
saliva_833	0.00158136	0.00188576	0.00570236	...	0.00007194	0.00007639	0.00004074
saliva_836	-0.00019119	-0.00030473	0.00083421	...	0.00331971	0.00328582	0.00328233
saliva_838	0.01148866	0.00954615	0.00928261	...	0.00151822	0.00134437	0.00125083
saliva_817	0.00602828	0.00492229	0.00593467	...	0.00007595	0.00003312	0.00005751
saliva_821	-0.00014933	-0.00003228	0.00274970	...	0.00003699	-0.00002118	-0.00003511
saliva_830	0.00098743	0.00015972	0.00243907	...	0.00011087	-0.00007184	-0.00009761
saliva_340	0.00364286	0.00394462	0.00520603	...	-0.00047263	-0.00028351	-0.00019273
saliva_343	0.00044076	0.00072326	0.00338411	...	-0.00032119	-0.00022316	-0.00015000
saliva_344	0.00033032	0.00265146	0.00736347	...	-0.00007307	-0.00005604	-0.00001775
saliva_347	0.00261519	0.00291997	0.00430917	...	0.00020154	0.00003072	-0.00003320
saliva_349	-0.00253518	-0.00257325	-0.00257837	...	0.00026005	0.00028424	0.00032808
saliva_357	0.02864434	0.02236209	0.01949352	...	-0.00161830	-0.00120914	-0.00087942
saliva_358	0.00759773	0.00617072	0.00775497	...	-0.00015770	-0.00019134	-0.00017931

Capítulo 4. Etapas para a Previsão

Diversas etapas foram executadas a fim de tentar prever a hipertensão arterial crônica e a doença periodontal através de metabólitos salivares identificados por RMN. Cada uma destas etapas é descrita neste Capítulo e os experimentos utilizando tais etapas são descritos no capítulo posterior.

4.1 Pré-processamento dos Dados

É extremamente importante a realização do pré-processamento dos dados, uma vez que, após coletados, os dados podem estar incompletos, não normalizados ou incorretos. Outro fator a ser levado em consideração no pré-processamento dos dados é a dimensionalidade dos dados, ou seja, a quantidade de variáveis utilizadas na entrada da técnica de reconhecimento de padrões utilizada (no caso deste trabalho, as Redes Neurais Artificiais). As técnicas de seleção de características ou seleção de variáveis são utilizadas para reduzir as dimensões dos dados, já que dados com muitas dimensões atrapalham o funcionamento da técnica de reconhecimento de padrões.

Este trabalho faz uso das técnicas para preenchimento de dados faltantes, normalização dos dados e seleção de características, os quais são especificados nas seções posteriores.

4.1.1 Representação dos Dados

Os dados podem apresentar a forma de variáveis contínuas e variáveis discretas. As variáveis contínuas são aquelas que podem assumir valores numéricos reais. As variáveis discretas podem ser divididas em ordinais e categóricas. As ordinais são aquelas que possuem uma ordenação natural e as variáveis categóricas são aquelas que não possuem uma ordenação natural [(BISHOP, 1995)].

Não é necessário se preocupar com a representação de variáveis contínuas, uma vez que seus valores são numéricos (reais). As variáveis ordinais podem simplesmente ser transformadas para valores contínuos. Um exemplo é a idade (em anos), a qual é uma variável ordinal e pode ser considerada como se fosse uma variável contínua. Já as variáveis categóricas precisam ser tratadas. Um exemplo é a variável cor, podendo assumir os valores

vermelho, verde e azul, a qual poderia ser representada pelos valores -1, 0 e 1, o que introduziria uma ordem nesta variável. Como a variável cor é categórica e não apresenta uma ordem, isso poderia acarretar problemas, já que a rede neural poderia estar considerando-a quando na verdade esta não existe. A fim de solucionar este problema, a variável pode ser representada por três variáveis distintas, chamada notação maximamente esparsa. A primeira diz se a cor é vermelha ou não, a segunda diz se é verde ou não e a terceira diz se a cor é azul ou não. Então, o valor vermelho poderia ser representado pelo vetor (1, -1, -1), o verde pelo vetor (-1, 1, -1) e o azul pelo vetor (-1, -1, 1). Se a variável fosse binária, ou seja, pudesse assumir apenas dois valores, ela poderia ser representada por uma única variável.

Neste trabalho, as variáveis foram representadas seguindo este modelo. Um exemplo é a variável “parto”, a qual é uma variável categórica que pode assumir os valores “normal” ou “cesariana” e foi representada por uma única variável, onde o valor 1 representa o parto “normal” e -1 representa o parto “cesariana”. Outros exemplos de variáveis categóricas e que foram representadas de forma semelhante foram as variáveis “diagnóstico periodontal”, “uso de medicamentos”, “eclampsia”, “intercorrência”, “insulina”, “hipertensa”, “med pré”, “med durante”, “med pós”, “início da hipertensão” e “classificação IMC”. Exemplos de variáveis ordinais são: “dentes cariados”, “dias pós-parto”, “tempo de diagnóstico” e “idade”.

4.1.2 Normalização dos Dados

Uma forma de pré-processamento necessária e muito comum é a normalização dos dados, ou seja, o reescalonamento das variáveis de entrada. A normalização é importante, uma vez que as variáveis de entrada podem apresentar valores muito distintos em termos de magnitude e a rede pode associar a magnitude à importância da variável. Então, se a normalização não for efetuada, certas variáveis podem ser consideradas mais importantes que outras e, como consequência, a eficiência da RNA pode ser prejudicada.

Neste trabalho a normalização utilizada é a *z-score* ou normalização estatística, a qual leva a média para zero e o desvio padrão para um. A equação referente à normalização *z-score* está representada pela equação 6

$$x_i^n = \frac{x_i^n - \mu_{x_i}}{\sigma_{x_i}} \quad (6)$$

Onde x_i^n é o valor da variável i da entrada n normalizada, x_i^n é o valor da variável i da entrada n antes de ser normalizada, μ_{x_i} é a média da variável i computada sobre todas as

entradas disponíveis e σ_{x_i} é o desvio padrão da variável i também computada sobre todas as entradas disponíveis.

4.1.3 Preenchimento de Dados Faltantes

O problema dos dados faltantes é muito comum e duas abordagens podem ser tomadas a fim de solucioná-lo. A primeira é o descarte de dados incompletos, a qual é uma solução que deve ser tomada apenas no caso de existirem uma grande quantidade de dados completos, pois, caso contrário, a eliminação pode piorar os resultados da técnica de reconhecimento de padrões utilizada. A segunda solução é o preenchimento dos dados faltantes com um determinado valor. Diversas heurísticas podem ser utilizadas para o preenchimento destes dados faltantes. Este trabalho utiliza a média das variáveis (dimensões) para este preenchimento, ou seja, a média de cada variável é calculada sobre os dados disponíveis e cada dado incompleto é substituído pela média da sua respectiva variável.

Os dados utilizados neste trabalho foram divididos em duas partes: (1) a parte referente aos metabólitos salivares e; (2) os dados que não são referentes aos metabólitos salivares (dados referentes ao exame bucal e aos dados obtidos via entrevista, conforme detalhado no capítulo anterior). Os dados referentes aos metabólitos salivares estavam completos, mas os outros dados apresentavam valores incompletos e foram completados utilizando a média (conforme explicado acima). Como neste trabalho foi utilizada a normalização *z-score*, a média de cada variável é zero, então os dados faltantes foram completados com zeros.

4.1.4 Identificação das Entradas e Saídas

Uma das etapas necessárias para preparar os dados para serem utilizados em uma RNA é a seleção de quais as variáveis serão utilizadas para compor a sua entrada e quais serão utilizadas para compor a sua saída. Como este trabalho tem a finalidade de prever a hipertensão arterial crônica e a doença periodontal, nove configurações distintas de conjuntos de entradas e conjuntos de saída foram utilizados (seis para a hipertensão arterial crônica, representadas na tabela 9, e três para a doença periodontal, representadas na tabela 10).

As tabelas 9 e 10 apresentam as configurações utilizadas e é importante ressaltar que as variáveis de entradas apresentadas na tabela ainda passarão pelo processo de seleção de características a fim de selecionar o subconjunto de variáveis realmente úteis para prever as saídas. Outro ponto a ser considerado é que a configuração DP3 utiliza os índices de

diagnóstico da doença periodontal para prever a doença. Como a doença é prevista a partir destes índices utilizando regras específicas, a RNA que utilizar esta configuração apenas fará o mapeamento desta regra e, portanto, espera-se que o desempenho desta seja alto.

Tabela 9: Configurações de variáveis de entradas e saídas utilizadas para previsão da hipertensão arterial crônica.

Configurações	Variáveis de entrada	Variáveis de saída	Descrição
H1	Variáveis metabólicas (seção 3.1.3.3)	“hipertenso”	Permite prever se um indivíduo sofre de hipertensão ou não apenas utilizando as variáveis metabólicas.
H2	Variáveis metabólicas (seção 3.1.3.3)	“PA sistólica” e “PA diastólica”	Permite prever as pressões sistólica e diastólica apenas utilizando as variáveis metabólicas.
H3	Todas (exceto variáveis “PA sistólica”, “PA diastólica”, “hipertenso” e variáveis metabólicas da seção 3.1.3.3).	“hipertenso”	Permite prever se um indivíduo sofre de hipertensão ou não utilizando apenas as variáveis não metabólicas.
H4	Todas (exceto variáveis “PA sistólica”, “PA diastólica”, “hipertenso” e variáveis metabólicas da seção 3.1.3.3).	“PA sistólica” e “PA diastólica”	Permite prever as pressões sistólica e diastólica utilizando apenas as variáveis não metabólicas.
H5	Todas (exceto variáveis “PA sistólica”, “PA diastólica” e “hipertenso”).	“hipertenso”	Permite prever se um indivíduo sofre de hipertensão ou não utilizando as variáveis não metabólicas e as variáveis metabólicas.
H6	Todas (exceto variáveis “PA sistólica”, “PA diastólica” e “hipertenso”).	“PA sistólica” e “PA diastólica”	Permite prever as pressões sistólica e diastólica utilizando as variáveis não metabólicas e as variáveis metabólicas.

Tabela 10: Configurações de variáveis de entradas e saídas utilizadas para previsão da doença periodontal.

Configurações	Variáveis de entrada	Variáveis de saída	Descrição
DP1	Variáveis metabólicas (seção 3.1.3.3).	“diagnóstico periodontal”	Permite prever se um indivíduo sofre de alguma doença periodontal apenas utilizando as variáveis metabólicas.
DP2	Índices de diagnóstico da doença periodontal (variáveis de “a” à “l” especificadas na seção 3.1.3.2) e variáveis metabólicas (seção 3.1.3.3).	“diagnóstico periodontal”	Permite prever se um indivíduo sofre de alguma doença periodontal utilizando os índices de diagnóstico da DP e utilizando as variáveis metabólicas.
DP3	Índices de diagnóstico da doença periodontal (variáveis de “a” à “l” especificadas na seção 3.1.3.2).	“diagnóstico periodontal”	Permite prever se um indivíduo sofre de alguma doença periodontal utilizando apenas os índices de diagnóstico da DP.

4.1.5 Seleção de Características

Dados que possuem uma dimensão elevada sofrem da Maldição da Dimensionalidade. Esta se refere ao fato de as técnicas de reconhecimento de padrões terem o seu desempenho prejudicado quando aplicadas sobre dados com muitas variáveis. Em contrapartida, um número muito pequeno de variáveis pode também causar o mesmo efeito, ou seja, prejudicar o desempenho dessas técnicas. As técnicas de seleção de características consistem em selecionar as variáveis mais importantes para o reconhecimento dos padrões, a fim de que a dimensionalidade dos dados seja diminuída e, como consequência, o desempenho da técnica de reconhecimento de padrões seja melhorado.

Neste trabalho, a técnica de seleção de características utilizada se baseia na ideia de que as variáveis de entrada devem ser pouco correlacionadas entre si, já que se elas possuírem uma correlação alta umas com as outras, grande parte da informação presente em uma variável estará contida em outra e, portanto, uma delas poderá ser desnecessária. Além disso, as variáveis de entrada devem ser altamente correlacionadas com as variáveis de saída, já que, se elas tiverem pouca correlação com as saídas, indicarão que a partir delas dificilmente será possível prever as saídas. Como em POMMERANZENBAUM (2014), sendo n o número de amostras existentes, o valor C definido pela Equação 7 e baseado no intervalo de confiança de 95% foi utilizado para decidir se uma variável X é muito ou pouco

correlacionada com uma variável Y, ou seja, se a correlação for inferior a C, as variáveis são consideradas pouco correlacionadas e, se for superior a este valor, elas são consideradas muito correlacionadas.

$$C = \frac{2}{\sqrt{n}} \quad (7)$$

Neste trabalho, é utilizada a correlação de Pearson a fim de verificar se duas variáveis X e Y são muito ou pouco correlacionadas. A Equação 8 é utilizada para calcular o coeficiente de correlação de Pearson.

$$\rho_{xy} = \frac{COV(X, Y)}{\sqrt{VAR(X) \cdot VAR(Y)}} = \frac{\sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_X)^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \mu_Y)^2}} \quad (8)$$

Quando os dados estão normalizados utilizando a normalização *z-score* (como é o caso deste trabalho), a Equação 8 pode ser reescrita como

$$\rho_{xy} = \frac{COV(X, Y)}{\sqrt{VAR(X) \cdot VAR(Y)}} = \frac{\sum_{i=1}^n x_i y_i}{(n - 1)} \quad (9)$$

Sabe-se que a correlação de Pearson mede o grau de correlação linear entre duas variáveis. Sabendo disso, é possível descorrelacionar a variável X da variável Y utilizando as equações 10 e 11.

$$x = ay + x' \rightarrow x' = x - ay \quad (10)$$

$$a = \frac{\sigma_x}{\sigma_y} \rho_{xy} \quad (11)$$

Os valores do coeficiente de correlação de Pearson variam entre -1 e 1 e, quanto maior em módulo, maior é a correlação entre as variáveis. Então, se o coeficiente de correlação de Pearson entre duas variáveis X e Y é igual a -1 ou +1, existe uma correlação linear perfeita entre X e Y. Já, se a correlação é igual a zero, as duas variáveis não dependem linearmente uma da outra.

Utilizando o coeficiente de correlação de Pearson dado pela equação 9 e o valor definido na equação 7, uma variável é considerada muito correlacionada com a outra se $|\rho| > C$ e é considerada pouco correlacionada se $|\rho| < C$. Com base nessa informação, o algoritmo de seleção de características utilizado neste trabalho é descrito abaixo:

- 1) *Dados o conjunto de entradas E , o conjunto S de saídas, o conjunto de variáveis VE do conjunto de entrada, o conjunto de variáveis VS do conjunto de saídas e o conjunto O vazio de variáveis obtidas pelo algoritmo. Para cada variável vs de VS faça
 - 1.1) *Calcule a correlação de cada variável de VE em relação à variável vs , utilizando o conjunto de entradas E e o conjunto de saídas S . Obtenha o conjunto de correlações de Pearson P .*
 - 1.2) *Selecione as variáveis de VE que possuem valor em P com módulo maior do que C e obtenha o conjunto VE' de variáveis.*
 - 1.3) *Ordene o conjunto VE' em ordem decrescente de acordo com os seus respectivos valores em P .*
 - 1.4) *Utilize o conjunto E para descorrelacionar as demais variáveis de VE' da primeira variável de VE' . Obtenha o conjunto E' com os valores obtidos após a descorrelação.*
 - 1.5) *Retire a primeira variável de VE' e coloque no conjunto de variáveis obtidas O .*
 - 1.6) *Utilize o conjunto E' para obter os coeficientes de correlação de Pearson de cada variável de VE' em relação a vs . Obtenha o conjunto P' .*
 - 1.7) *Remova de VE' todas as variáveis com valor em P' menor do que C .*
 - 1.8) *Se VE' não estiver vazio, volte ao passo 1.4.**
- 2) *Retorne o conjunto O que conterà todas as variáveis selecionadas pelo algoritmo.*

Algoritmo 2: Primeiro algoritmo para seleção de características.

Ao utilizar este algoritmo para selecionar características, poucas foram as variáveis selecionadas e, através de experimentos, viu-se que a inclusão de outras variáveis aumentou o desempenho dos experimentos. Por esse motivo, a maioria dos experimentos realizados utilizam todas aquelas variáveis que estavam muito correlacionadas com alguma das saídas, ou seja, aquelas em que algum dos coeficientes de Pearson em relação a alguma das saídas tinha valor maior do que o valor especificado na equação 7. Tal algoritmo é representado pelo algoritmo 3.

- 1) *Dados o conjunto de entradas E, o conjunto S de saídas, o conjunto de variáveis VE do conjunto de entrada, o conjunto de variáveis VS do conjunto de saídas e o conjunto O vazio de variáveis obtidas pelo algoritmo. Para cada variável vs de VS faça*
 - 1.1) *Calcule a correlação de cada variável de VE em relação à variável vs, utilizando o conjunto de entradas E e o conjunto de saídas S. Obtenha o conjunto de correlações de Pearson P.*
 - 1.2) *Selecione as variáveis de VE que possuem valor em P com módulo maior do que C e adicione no conjunto O.*
- 2) *Retorne o conjunto O que conterá todas as variáveis selecionadas pelo algoritmo.*

Algoritmo 3: Segundo algoritmo para seleção de características.

Os gráficos a seguir (Figuras 6, 7 e 8) mostram os valores dos coeficientes de Pearson referentes à correlação entre as variáveis de entrada e as saídas das configurações apresentadas nas Tabelas 9 e 10 da seção 4.1.4. O eixo X dos gráficos representa números sequenciais usados apenas para poder desenhar o gráfico e o eixo Y representa o módulo do coeficiente de correlação de Pearson, o qual mede a correlação entre uma determinada entrada e uma determinada saída.

O gráfico A da Figura 6 foi plotado a fim de selecionar os metabólitos salivares a serem utilizados para prever a saída “hipertenso” da configuração H1. Nela, pode-se perceber que cinco metabólitos ficaram acima da linha cheia azul, que representa aquele valor mencionado anteriormente que serve para identificar quais metabólitos estão muito correlacionados com a saída (valor definido pela equação 7). Então, cinco metabólitos devem ser selecionados para compor a entrada da rede para a configuração H1. São eles: x13, x14, x15, x16 e x25. Os gráficos B e C da Figura 6 foram plotados a fim de selecionar os metabólitos salivares a serem utilizados para prever as saídas “PA sistólica” e “PA diastólica” da configuração H2. O gráfico B representa a correlação dos metabólitos salivares com a saída “PA sistólica” e pode-se perceber que vinte e um metabólitos ficaram acima da linha cheia azul. São eles (ordenado pela maior correlação): x124, x131, x125, x132, x133, x123, x88, x134, x97, x89, x126, x98, x90, x99, x52, x127, x38, x34, x100, x55 e x114. O gráfico C representa a correlação dos metabólitos salivares com a saída “PA sistólica” e vinte metabólitos ficaram acima da linha cheia azul. São eles (ordenados pela maior correlação): x89, x90, x98, x99,

x103, x100, x88, x123, x97, x124, x104, x125, x91, x102, x101, x126, x55, x93, x94 e x52. Então, as variáveis selecionadas para a configuração H2 são formadas pela união das variáveis selecionadas no gráfico B com as selecionadas no gráfico C. São elas (totalizando 29): x124, x131, x125, x132, x133, x123, x88, x134, x97, x89, x126, x98, x90, x99, x52, x127, x38, x34, x100, x55, x114, x103, x88, x104, x91, x102, x101, x93 e x94.

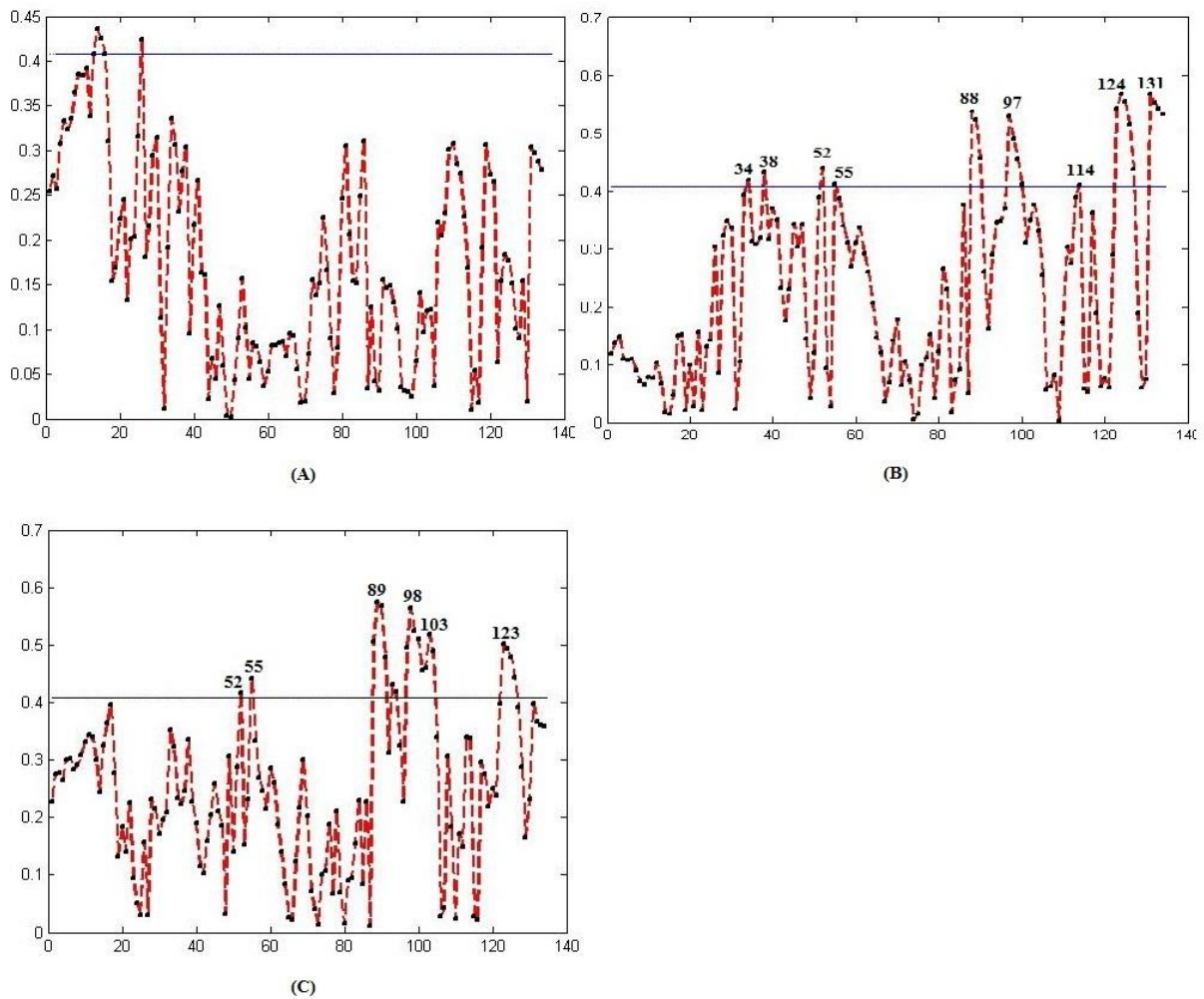


Figura 6: Gráficos da correlação dos metabólitos salivares com: (A) a saída “hipertenso”. (B) a saída “PA sistólica”. (C) a saída “PA diastólica”.

O gráfico A da Figura 7 foi plotado a fim de selecionar as entradas da configuração H3 a serem utilizadas para prever a saída “hipertenso”. Nele, pode-se perceber que cinco variáveis ficaram acima da linha cheia azul. Então, as cinco variáveis devem ser selecionadas para compor a entrada da rede para a configuração H3. São elas (ordenadas por correlação): “IMC”, “peso pós-parto”, “idade”, “parto” e “sangramento à sondagem”. Os gráficos B e C

da Figura 7 foram plotados a fim de selecionar as entradas da configuração H4 a serem utilizados para prever as saídas “PA sistólica” e “PA diastólica”. No gráfico B, pode-se perceber que duas variáveis ficaram acima da linha cheia azul. Então, as duas variáveis devem ser selecionadas. São elas (ordenadas por correlação): “IMC” e “peso pós-parto”. No gráfico C, pode-se perceber que nenhuma variável ficou acima da linha cheia azul. Então, nenhuma variável foi selecionada. Como a saída da configuração H4 é formada pelas saídas “PA sistólica” e “PA diastólica”, as variáveis selecionadas para esta configuração são as variáveis selecionadas nos gráficos B e C. São elas: “IMC” e “peso pós-parto”.

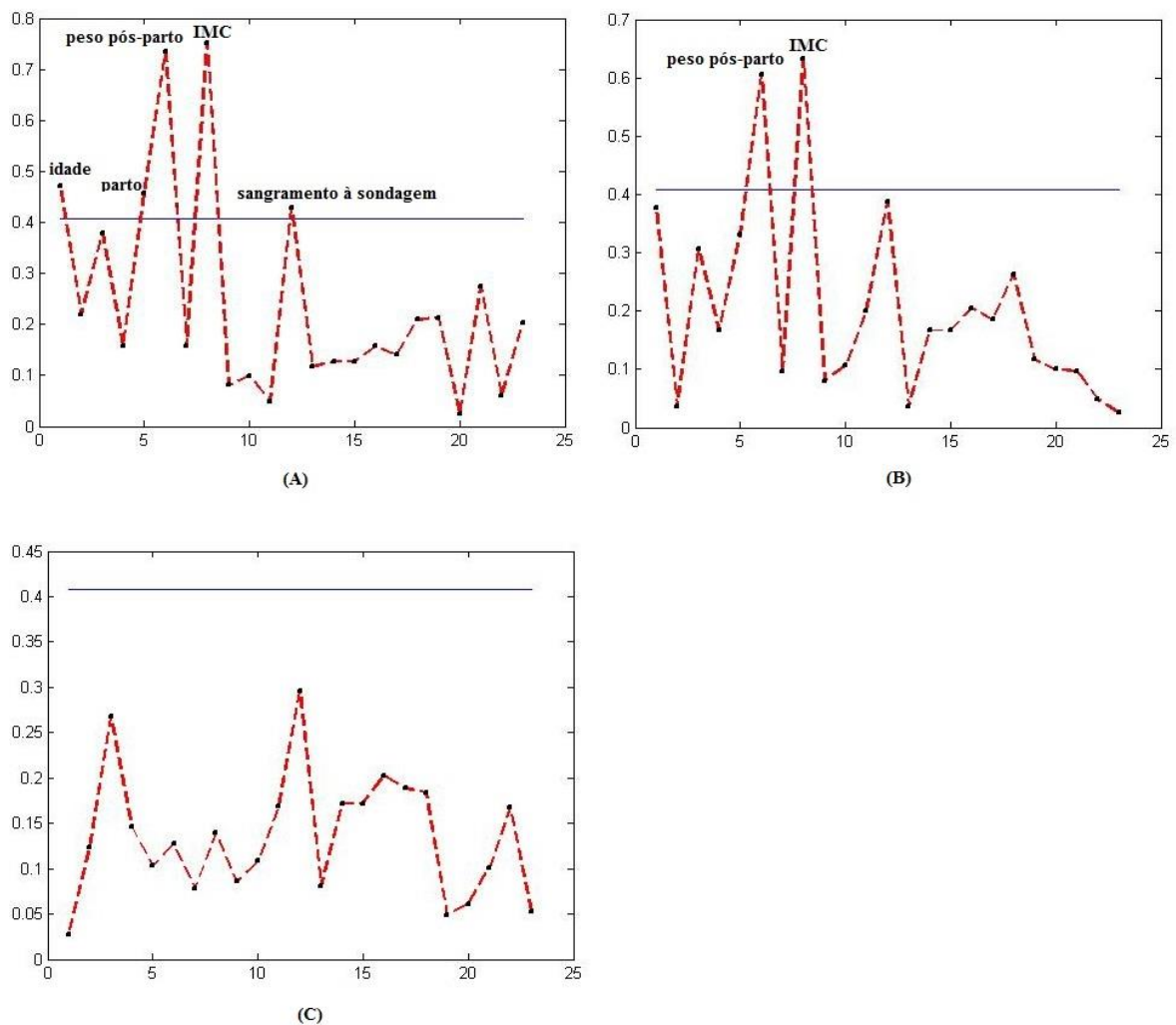


Figura 7: Gráficos da correlação das variáveis de entrada da: (A) configuração H3 com a saída “hipertenso”. (B) configuração H4 com a saída “PA sistólica”. (C) configuração H4 com a saída “PA diastólica”.

Os gráficos da Figura 8 foram plotados a fim de selecionar as entradas da configuração DP1 a serem utilizadas para prever as saídas da variável “diagnóstico periodontal”. Essa variável foi dividida em três, já que a notação utilizada foi a maximamente esparsa, ou seja, são três variáveis, onde a primeira indica se o indivíduo é saudável (valor 1 se for saudável e -1 caso contrário), a segunda indica se o indivíduo sofre de gengivite (valor 1 se sofrer de gengivite e -1 caso contrário) e a terceira indica se o indivíduo sofre de periodontite (valor 1 se sofrer de periodontite e -1 caso contrário). O gráfico A da Figura 8 mostra a correlação das entradas com a variável que indica se o indivíduo apresenta ou não doença periodontal (saudável ou não). Nela, pode-se perceber que uma variável ficou acima da linha cheia azul. Então, a variável deve ser selecionada. É ela: x128. O gráfico B da Figura 8 mostra a correlação das entradas com a variável que indica se o indivíduo sofre de gengivite. Nela, pode-se perceber que sete variáveis ficaram acima da linha cheia azul. Então, as sete variáveis devem ser selecionadas. São elas (ordenadas pela correlação): x54, x71, x130, x129, x72, x86 e x80. O gráfico C da Figura 8 mostra a correlação das entradas com a variável que indica se o indivíduo sofre de periodontite. Nela, pode-se perceber que oito variáveis ficaram acima da linha cheia azul. Então, as oito variáveis devem ser selecionadas. São elas (ordenadas pela correlação): x129, x128, x54, x130, x53, x71, x127 e x72. Como a configuração DP1 está relacionada às três saídas (relativas aos gráficos A, B e C da Figura 8), as variáveis selecionadas para esta configuração são resultantes da união das variáveis selecionadas para cada uma destas saídas. São elas (totalizando 10): x128, x54, x71, x130, x129, x72, x86, x80, x53 e x127.

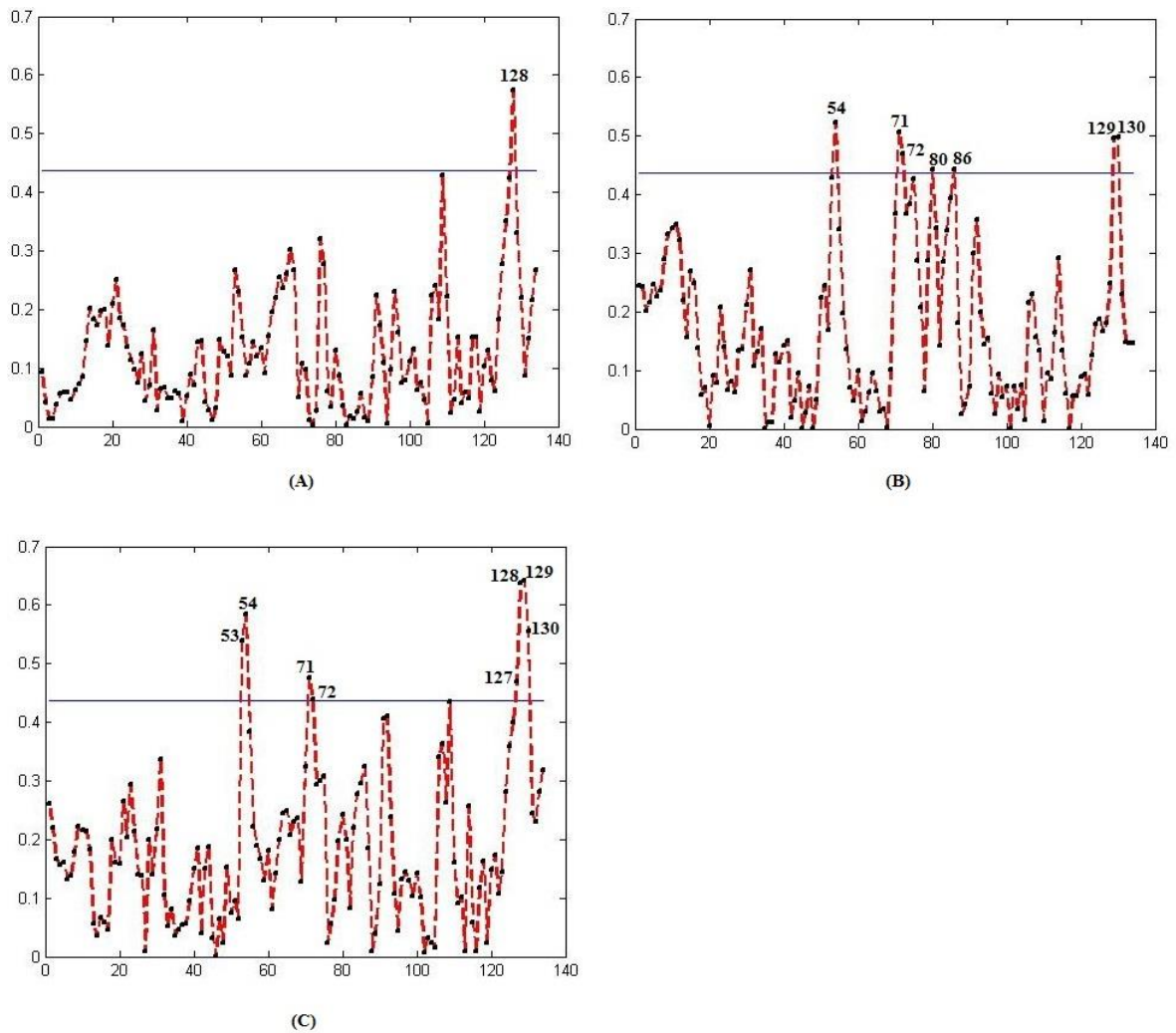


Figura 8: Gráficos da correlação dos metabólitos salivares com: (A) a saída “saudável”. (B) a saída “gengivite”. (C) a saída “periodontite”.

Os gráficos da Figura 9 foram plotados a fim de selecionar as entradas da configuração DP3 a serem utilizadas para prever as saídas da variável “diagnóstico periodontal”. Conforme citado anteriormente, essa variável foi dividida em três, já que a notação utilizada para variáveis categóricas foi a notação maximamente esparsa, ou seja, serão três variáveis, onde a primeira indica se o indivíduo é saudável (valor 1 se for saudável e -1 caso contrário), a segunda indica se o indivíduo sofre de gengivite (valor 1 se sofrer de gengivite e -1 caso contrário) e a terceira indica se o indivíduo sofre de periodontite (valor 1 se sofrer de periodontite e -1 caso contrário). O gráfico A da Figura 9 mostra a correlação das entradas com a variável que indica se o indivíduo apresenta ou não doença periodontal (saudável ou não). Nele, pode-se perceber que sete variáveis ficaram acima da linha cheia azul. Então, as

variáveis devem ser selecionadas. São elas (ordenadas pela correlação): “Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”. O gráfico B da Figura 9 mostra a correlação das entradas com a variável que indica se o indivíduo apresenta ou não gengivite. Nele, pode-se perceber que nenhuma variável ficou acima da linha cheia azul. Então, nenhuma variável foi selecionada. O gráfico C da Figura 9 mostra a correlação das entradas com a variável que indica se o indivíduo apresenta ou não periodontite. Nele, pode-se perceber que seis variáveis ficaram acima da linha cheia azul. Então, as variáveis devem ser selecionadas. São elas (ordenadas pela correlação): “Profundidade de bolsa”, “Nível de perda de inserção”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”. A união das variáveis selecionadas através dos gráficos A, B e C da Figura 9 são as variáveis selecionadas para a configuração DP3. São elas: “Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”.

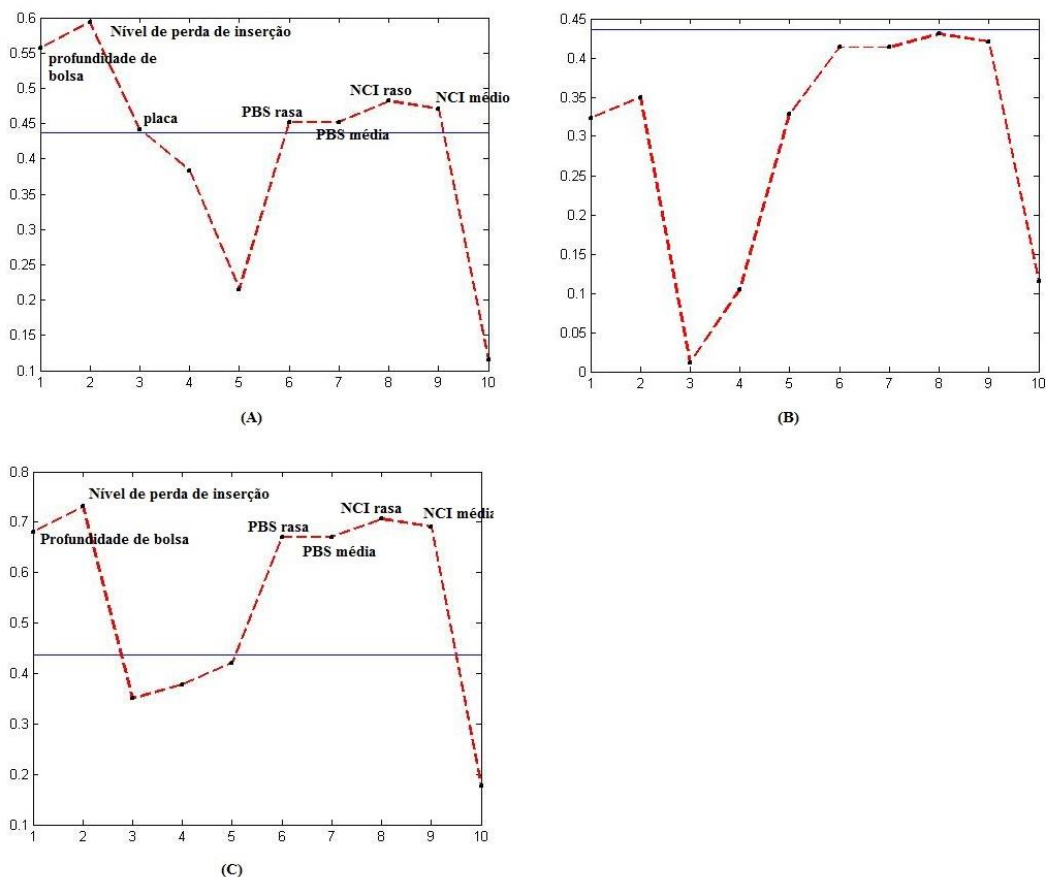


Figura 9: Gráficos da correlação dos índices utilizados para diagnóstico da doença periodontal com: (A) a saída “saúde”. (B) a saída “gengivite”. (C) a saída “periodontite”.

Para selecionar as variáveis de entrada para a configuração DP2, basta unir as variáveis selecionadas nas configurações DP1 e DP3, visto que as variáveis de entrada da configuração DP3 são resultantes da união das variáveis de entrada das configurações DP1 e DP3. Então, as variáveis selecionadas para a configuração DP2 são (totalizando 17): x128, x54, x71, x130, x129, x72, x86, x80, x53, x127, “Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”.

As tabelas 11 e 12 resumem as variáveis selecionadas para cada uma das configurações especificadas nas tabelas 9 e 10 da seção 4.1.4.

Tabela 11: variáveis selecionadas na etapa de “seleção de características” para cada configuração da Tabela 9.

Configurações	Variáveis de entrada selecionadas	Total de variáveis selecionadas	Variáveis de saída
H1	x13, x14, x15, x16 e x25	5	“hipertenso”
H2	x124, x131, x125, x132, x133, x123, x88, x134, x97, x89, x126, x98, x90, x99, x52, x127, x38, x34, x100, x55, x114, x103, x88, x104, x91, x102, x101, x93 e x94	29	“PA sistólica” e “PA diastólica”
H3	“IMC”, “peso pós-parto”, “idade”, “parto” e “sangramento á sondagem”.	5	“hipertenso”
H4	“IMC” e “peso pós-parto”.	2	“PA sistólica” e “PA diastólica”
H5	x13, x14, x15, x16, x25, “IMC”, “peso pós-parto”, “idade”, “parto” e “sangramento á sondagem”.	10	“hipertenso”
H6	x124, x131, x125, x132, x133, x123, x88, x134, x97, x89, x126, x98, x90, x99, x52, x127, x38, x34, x100, x55, x114, x103, x88, x104, x91, x102, x101, x93, x94, “IMC” e “peso pós-parto”.	31	“PA sistólica” e “PA diastólica”

Tabela 12: Variáveis selecionadas na etapa de “seleção de características” para cada configuração da Tabela 10.

Configurações	Variáveis de entrada selecionadas	Total de variáveis selecionadas	Variáveis de saída
DP1	x128, x54, x71, x130, x129, x72, x86, x80, x53 e x127	10	“diagnóstico periodontal”
DP2	x128, x54, x71, x130, x129, x72, x86, x80, x53, x127, “Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”.	17	“diagnóstico periodontal”
DP3	“Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”.	7	“diagnóstico periodontal”

4.1.6 PCA e SVD para Visualização dos Dados

A Análise de Componentes Principais (PCA, do inglês Principal Component Analysis) e a Decomposição de Valores Singulares (SVD, do inglês Singular Value Decomposition) são dois algoritmos utilizados para redução de dimensionalidade e estão dentro do conjunto de algoritmos de extração de características. Os algoritmos de extração de características diferem dos algoritmos de seleção de características (seção 4.1.5), pois enquanto os primeiros têm como finalidade combinar linearmente ou não linearmente as variáveis originais do conjunto de dados para obter um conjunto com um menor número de dimensões, os algoritmos de seleção de características apenas selecionam as variáveis mais relevantes e não fazem qualquer tipo de combinação entre elas ([BISHOP, 1995]).

Segundo RINGNÉR (2008), a PCA é um algoritmo matemático que reduz a dimensionalidade dos dados enquanto retém a maior parte da variação no conjunto de dados. A PCA reduz a dimensionalidade identificando direções, chamadas de componentes principais, ao longo das quais a variação dos dados é máxima. Dessa forma, a PCA é capaz de resumir os dados com um alto número de variáveis utilizando um número de variáveis reduzido. A PCA pode ser desenvolvida utilizando o algoritmo SVD. Este permite a representação de uma matriz em três outras (U, S e V) e, sozinho (sem o uso da PCA), também permite eliminar de forma fácil as partes menos importantes desta representação a fim de produzir uma representação aproximada da matriz original com um número menor de variáveis ([RAJARAMAN & ULLMAN, 2012]).

Neste trabalho, as técnicas PCA e SVD foram utilizadas apenas para reduzir a dimensionalidade dos dados para duas dimensões a fim de permitir a visualização deles, ou seja, não foram utilizados com a finalidade de reduzir a dimensão dos conjuntos de entrada das RNAs utilizadas no Capítulo 5. Esse procedimento foi realizado apenas para as configurações da seção 4.1.5 em que os conjuntos de entrada são formados apenas pelos metabólitos salivares, ou seja, para as configurações H1, H2 e DP1. Os gráficos A, B, C e D da Figura 10 apresentam os resultados da aplicação das técnicas PCA e SVD sobre os dados referentes aos 24 indivíduos do estudo e considerando apenas os metabólitos salivares selecionados na seleção de características para as configurações H1 e H2. Nesses gráficos, os pontos relativos aos indivíduos hipertensos são círculos vermelhos e os relativos aos indivíduos não hipertensos são x's azuis. Neles, percebe-se que a maior parte dos indivíduos saudáveis ficou agrupada na região delimitada pela elipse verde. Apesar de dentro destas elipses existirem também indivíduos hipertensos e fora existirem indivíduos saudáveis, a existência desta região indica que existe a possibilidade de separação dos indivíduos saudáveis dos hipertensos, mesmo porque com a redução da dimensionalidade parte da informação foi perdida.

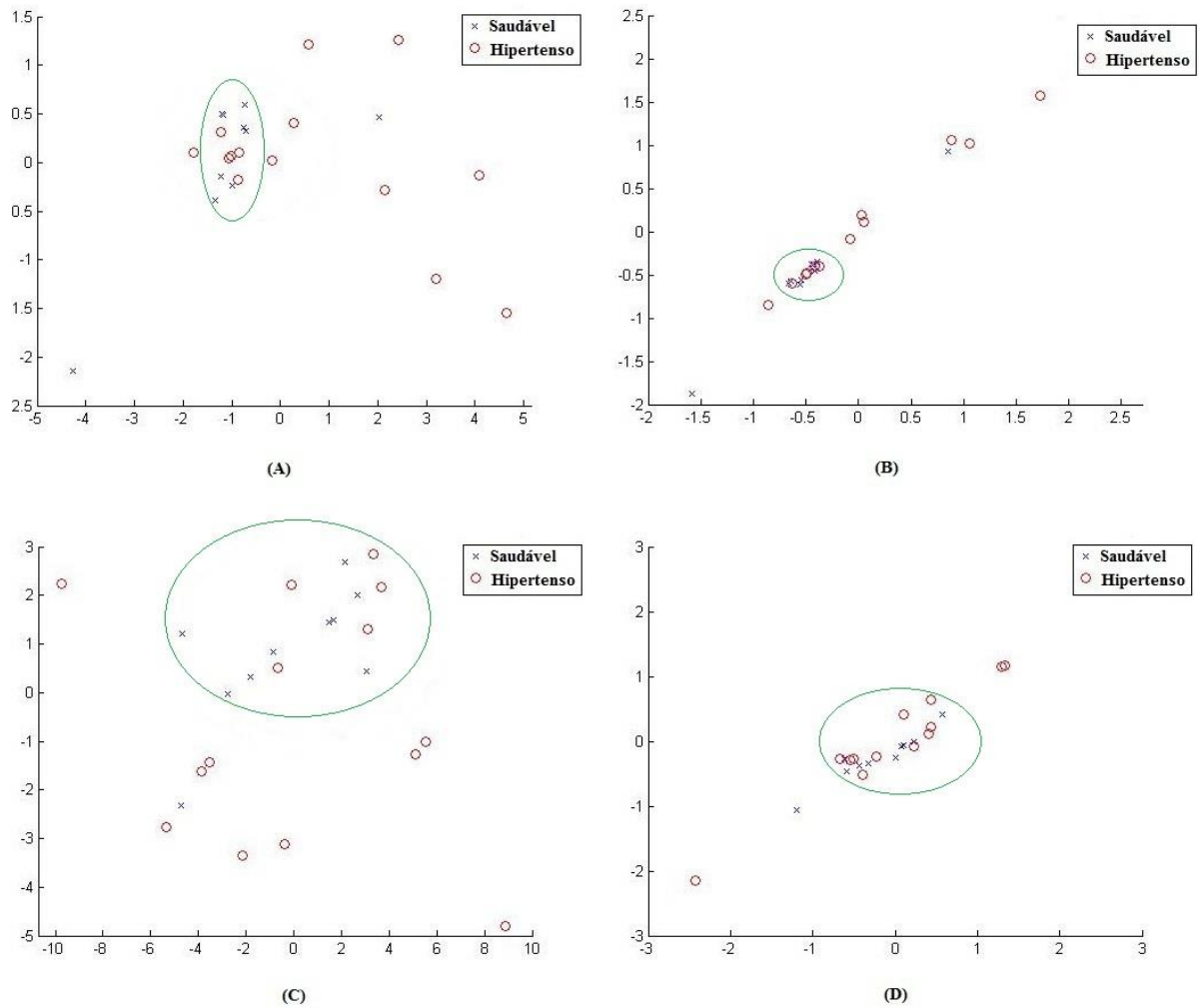


Figura 10: Gráficos referentes à execução do algoritmo: (A) PCA sobre a configuração H1 (B) SVD sobre a configuração H1 (C) PCA sobre a configuração H2 e (D) SVD sobre a configuração H2

Os gráficos A, B, C, D, E e F da Figura 11 apresentam os resultados da aplicação das técnicas PCA e SVD sobre os dados referentes aos 24 indivíduos do estudo e considerando apenas os metabólitos salivares selecionados na seleção de características para a configuração DP1. Os gráficos A e B são relativos à aplicação do PCA e SVD sobre os metabólitos selecionados em relação à saída “gingivite” (metabólitos relevantes obtidos na seção anterior para separar os indivíduos com gengivite dos outros). Os gráficos C e D são relativos à aplicação da PCA e SVD sobre os metabólitos selecionados em relação à saída “periodontite” (metabólitos relevantes obtidos na seção anterior para separar os indivíduos com periodontite dos outros) e os gráficos E e F são relativos à aplicação do PCA e SVD sobre os metabólitos selecionados em relação à saída “doença periodontal” (metabólitos relevantes obtidos na seção anterior para separar os indivíduos de acordo com o diagnóstico periodontal – saudável,

com gengivite e com periodontite). Nesses gráficos, os pontos relativos aos indivíduos saudáveis são verdes, os relativos aos indivíduos com gengivite são azuis e os relativos aos indivíduos com periodontite são roxos. Nos gráficos, percebe-se que a maior parte dos indivíduos consegue ser separados de acordo com o diagnóstico periodontal, conforme elipses presentes nas imagens. Deve-se levar em conta que parte da informação foi perdida no processo de redução de dimensionalidade e, portanto, a separação realizada pela rede neural será mais eficaz do que a separação visual sobre os dados obtidos após a aplicação do PCA e do SVD.

Através da comparação das Figuras 10 e 11, é possível ver que os grupos foram mais claramente separados na Figura 11, a qual está associada à doença periodontal, do que na Figura 10, a qual está associada à hipertensão arterial crônica. Na Figura 10, apesar de existirem regiões que separam os indivíduos hipertensos dos não hipertensos, um número razoável de indivíduos ficam localizados na região errada. Já na Figura 11, são poucos os indivíduos que estão situados na região errada, o que nos fornece fortes indícios de que a doença periodontal é mais facilmente detectável a partir da saliva do que a hipertensão arterial crônica (o que realmente foi constatado nos experimentos detalhados no Capítulo 5).

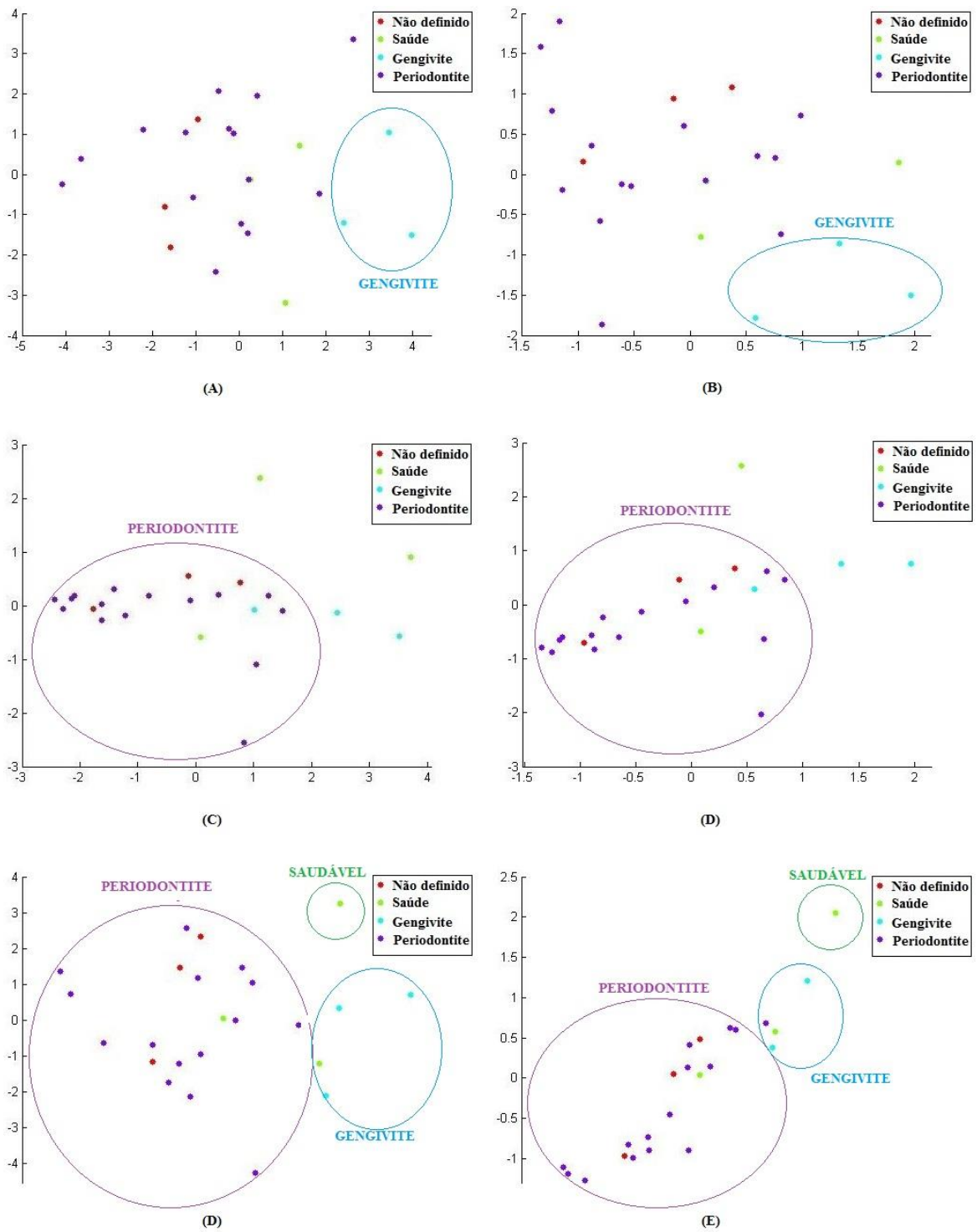


Figura 11: Gráficos referentes à execução do algoritmo: (A) PCA sobre a configuração DP1 – gengivite; (B) SVD sobre a configuração DP1 – gengivite; (C) PCA sobre a configuração DP1 – periodontite; (D) SVD sobre a configuração DP1 – periodontite (E) PCA sobre a configuração DP1 – doença periodontal; (F) SVD sobre a configuração DP1 – doença periodontal.

4.1.7 Detecção de Intrusos

Um intruso é uma observação que parece desviar de outros membros da amostra na qual ocorre ([Grubbs, 1969]). Eles atrapalham as técnicas de detecção de padrões, já que estas observações desviantes podem ser resultado de erros introduzidos nos dados, como, por exemplo, um erro de medição. Em vista disso, é extremamente importante a utilização de alguma técnica de identificação de intrusos.

Este trabalho detecta as observações desviantes em relação aos metabólitos salivares com a finalidade de identificar possíveis erros introduzidos durante o processo de obtenção destes. Para isto, foram criadas redes neurais *feedforward* e, após a normalização dos dados e seleção de características, as redes foram treinadas utilizando o algoritmo *backpropagation* e a validação cruzada (detalhada na seção 4.2). Foram criadas redes para as configurações H1 e H2, mas a configuração H2 foi dividida em duas: H2a e H2b. A H2a tem como entrada apenas os metabólitos selecionados em relação à saída “PA sistólica” e como saída esta mesma variável. A configuração H2b tem como entrada apenas os metabólitos selecionados em relação à saída “PA diastólica” e como saída esta mesma variável. Só foram utilizadas estas configurações, pois o objetivo é detectar os intrusos em relação aos metabólitos salivares e, portanto, estas três configurações bastam para que este objetivo seja alcançado. Na verdade, seria necessária apenas uma destas configurações, mas, com o objetivo de reafirmar os resultados, as três foram utilizadas.

Para detectar os intrusos, os acertos (porcentagem) e erros (erro médio quadrado) nos conjuntos de treinamento, validação e teste foram calculados para cada caso da validação cruzada, os casos com maiores erros foram identificados e os grupos que os introduziram foram selecionados (tabelas 13, 15 e 17). Além disso, cada entrada de cada caso da validação cruzada teve o seu erro calculado, as entradas pertencentes aos grupos selecionados nas tabelas 13, 15 e 17 foram analisadas e as amostras com os maiores erros foram detectadas e selecionadas como sendo possíveis intrusos (tabelas 14, 16 e 18).

A partir da análise das tabelas 13 e 14, as quais são referentes à configuração H1, é possível ver que as amostras 20, 7 e 8 do grupo 3 introduzem erros grandes e que, portanto, elas foram selecionadas como possíveis intrusos. Nas tabelas 15 e 16, as quais são referentes à configuração H2a, é possível ver que as amostras 24 e 14 do grupo 5 introduzem erros grandes e que, portanto, elas também foram selecionadas como possíveis intrusos. Em relação à análise das tabelas 17 e 18, as quais são referentes à configuração H2b, é possível

ver que as amostras 9 e 8 do grupo 3 introduzem erros grandes e que, portanto, elas também foram selecionadas como possíveis intrusos. Como resultado da detecção, seis entradas foram identificadas como intrusos. São elas: 7, 8, 9, 14, 20 e 24.

Após a detecção dos possíveis intrusos, foram realizados experimentos para verificar se a eliminação destes do conjunto de dados provocaria uma melhora nos resultados. Como pode ser visto nos experimentos 7 e 8 do Capítulo 5, a eliminação dos possíveis intrusos não ocasionou melhora significativa nos resultados, o que indicou que as entradas selecionadas ou parte delas não eram intrusos. Por isso, preferiu-se não eliminá-los nos experimentos posteriores.

Tabela 13: Erros e acertos nos conjuntos de treinamento, validação e teste para cada caso da validação cruzada (configuração H1).

CONFIGURAÇÃO H1 (SAÍDA HIPERTENSO)										
CASO	GRUPOS TREINAMENTO	GRUPOS TESTE	GRUPOS VALIDAÇÃO	ACERTO TREINAMENTO	ERRO TREINAMENTO	ACERTO TESTE	ERRO TESTE	ACERTO VALIDAÇÃO	ERRO VALIDAÇÃO	
1	345	2	1	71,43	10,23	80,00	3,14	80,00		2,22
2	245	3	1	85,71	8,43	40,00	5,30	100,00		1,46
3	235	4	1	78,57	11,41	80,00	3,51	100,00		2,20
4	234	5	1	60,00	13,85	75,00	3,85	100,00		2,66
5	345	1	2	71,43	11,55	100,00	2,67	60,00		3,94
6	145	3	2	85,71	9,39	40,00	5,06	60,00		3,99
7	135	4	2	71,43	10,72	80,00	4,07	60,00		3,73
8	134	5	2	66,67	11,39	75,00	3,26	60,00		3,93
9	245	1	3	71,43	13,50	40,00	4,88	60,00		4,91
10	145	2	3	50,00	14,19	40,00	5,37	60,00		4,92
11	125	4	3	85,71	9,29	60,00	4,10	40,00		5,02
12	124	5	3	60,00	14,05	50,00	4,00	60,00		4,76
13	235	1	4	71,43	11,66	80,00	2,67	80,00		3,76
14	135	2	4	85,71	8,97	60,00	3,52	80,00		3,27
15	125	3	4	85,71	8,12	40,00	5,19	80,00		3,81
16	123	5	4	80,00	10,97	100,00	2,61	80,00		3,84
17	234	1	5	53,33	12,45	80,00	2,74	50,00		3,31
18	134	2	5	80,00	11,44	80,00	3,61	100,00		2,82
19	124	3	5	80,00	8,38	40,00	5,13	100,00		2,13
20	123	4	5	66,67	11,53	60,00	4,73	75,00		3,66

Tabela 14: Erros para cada entrada em cada caso da validação cruzada (configuração H2a).

ERROS HIPERTENSO																			
CASO 1		CASO 2		CASO 3		CASO 4		CASO 5		CASO 6		CASO 7		CASO 8		CASO 9		CASO 10	
4	2	7	1	10	1	13	1	10	1	7	1	10	1	13	1	1	1	4	1
5	0,21322	8	1	11	1	14	1	2	0,30176	8	1	11	1	14	1	2	1	5	1,2569
6	0,043589	9	0,04645	12	0,57507	23	-0,97862	3	0,25708	9	0,2749	12	0,61259	23	-1,0202	3	1	6	1
17	-1	19	-1	21	-0,75348	24	-0,93271	15	-0,86479	19	-1	21	-1	24	-0,95745	15	-1,0446	17	-1
18	-0,38596	20	-2	22	-1,26E-01	9	0,27774	16	-1	20	-2	22	-0,37209	8	1	16	-1	18	-1,2285
24	-1	13	0,24957	8	1	21	-1	10	0,68076	10	0,89787	14	1	1	0	10	1	3	1
22	-0,04377	14	1	13	0,39211	4	1	13	0	22	0	7	1	22	0	22	-1	10	0,80462
21	-1	18	-0,2558	5	0,54776	10	1	24	-1	13	0,64587	13	0,73979	7	1,0636	13	0,95512	24	-1
9	0	24	0	6	0,1465	12	1	8	1	16	-0,90705	15	-1	16	-0,10197	4	1	11	0,59563
14	0,5293	5	0,34925	20	-1,7636	20	-1	9	0	15	-0,76672	23	-0,89933	2	0	14	0,95118	16	-1
7	1	23	-0,615	4	1,48E+00	8	0,94838	19	-1,0881	12	0,61738	19	-0,82937	12	1	5	1	12	0,92788
12	0	10	1	24	-7,58E-01	7	1	20	-2	23	-1	20	-1,7206	21	-1	23	-1	1	1
20	-2	6	0	9	0	11	0,83141	14	0,66159	21	-1	1	0,19542	19	-1	6	1	23	-1,2858
8	1,2528	21	-0,6232	19	-9,04E-01	19	-0,854	23	-1	16	-1	2	0,25373	11	1	17	-1	2	1
23	-1	4	2	17	-0,58289	6	0	21	-0,93472	1	0	9	0,23531	15	-0,88067	24	-1	21	-1
13	0,18766	11	1	14	0,7835	18	-1	12	0,57789	24	-0,88301	8	1,1007	3	0	21	-1	14	0,79921
11	1,3949	12	0	23	-1	17	-1,1605	7	0,8063	14	0,71667	24	-0,81846	10	1	18	-1	13	0,80478
19	-1,1016	17	-1	7	0,81805	22	-1	11	1	11	1,0563	2	0,26231	20	-2	11	1	22	-1,407
10	0,48839	22	-0,0845	18	-0,37128	5	1,0796	22	0	3	0,19707	3	0	9	0	12	1	15	-1
1	0,033192	1	0,03346	1	0,18152	1	0,20762	4	1,2764	4	1	4	1	4	1	7	1	7	1
2	0,090416	2	0,0518	2	0,25284	2	0,33963	5	0,64409	5	0,6503	5	0,82708	5	0,71768	8	1	8	1
3	0,077061	3	0,04044	3	0,15252	3	0,27673	6	0,24547	6	0,19345	6	0,19352	6	0,17791	9	1	9	1
15	-0,79817	15	-0,5042	15	-0,64224	15	-0,94449	17	-1	17	-1,2057	17	-1	17	-1	19	-1	19	-1
16	-1,2257	16	-0,8262	16	-0,9741	16	-0,89292	18	-1	18	-0,59703	18	-0,66149	18	-0,73569	20	-1	20	-1
CASO 11		CASO 12		CASO 13		CASO 14		CASO 15		CASO 16		CASO 17		CASO 18		CASO 19		CASO 20	
10	1	13	1	1	0,21352	4	2	7	1,2408	13	0	1	0,31463	4	2	7	1,0624	10	1
11	1	14	0,68369	2	0,25212	5	0,4103	8	1	14	0,69963	2	0,31454	5	0,59498	8	1	11	1
12	1	23	-1	3	0,21944	6	0,11896	9	0,12989	23	-0,786	3	0,18654	6	0,2382	9	0,082786	12	0,94932
21	-0,89618	24	-1	15	-0,86863	17	-1	19	-0,71643	24	-0,7147	15	-0,85903	17	-0,83241	19	-0,70932	21	-0,90413
22	0	16	-1	16	-1	18	-0,32061	20	-2	18	-0,56956	16	-1,0623	18	-0,43615	20	-2	22	-0,51643
3	0	18	-1	4	1,192	19	-0,80895	4	1,3794	19	-0,82945	20	-2	11	1,4085	5	0,25378	18	-0,80327
5	0	11	0,6711	5	0,50254	9	0	5	1	4	1	12	1	2	0,38724	11	1,4919	7	1
14	1	16	-1	23	-0,97891	13	0,21259	3	0,10285	6	0,22133	6	0,19559	1	0,23188	12	0,31065	3	0,34578
1	0	3	0,60757	8	0,10663	3	0	17	-1,0179	2	0,23683	18	-0,5848	22	-0,14086	4	1,538	15	-0,79581
6	0	15	-1	13	0	20	-1,8649	1	0	1	0,22129	5	0,71305	7	0,83265	6	0,074708	17	-1
23	-0,84504	1	0,81812	19	-1	14	0,56352	24	-1	16	-1	8	1,0505	9	0	10	1	16	-1
16	-1,1006	5	1	9	0	8	1,4361	14	0,71692	15	-0,70288	11	1,048	10	0,76146	17	-0,87774	19	-1
4	1,1805	6	1	7	0,90514	24	-1	15	-0,62539	8	1,2607	22	0	20	-1,7003	21	-0,69147	6	0,20272
17	-0,88294	17	-1	20	-1,7461	7	1	18	-0,44016	3	0,22377	21	-1,0902	21	-1	15	-0,56028	9	0,21277
13	0,84858	4	1	24	-0,90074	16	-0,83864	13	0,58399	17	-0,93092	7	1	8	1,3414	3	0,078165	8	1,1424
15	-1	2	0,69026	18	-1	15	-0,50639	16	-0,87158	20	-1,7671	4	1	15	-1	16	-1	20	-1,4811
2	0	22	-1,2657	6	0,21234	1	0,12049	2	0,14011	5	0	10	0,59299	16	-0,93733	2	0,091094	4	1,1299
24	-0,73992	12	1	17	-1,1352	2	0	23	-0,76482	7	0,99749	19	-1	12	0,71149	22	-0,21612	1	0,14483
18	-0,79081	10	1	14	0,64349	23	-0,71832	6	0,095409	9	0,23155	9	0,52777	3	0	1	0,07443	5	1,1255
7	1	7	0,66577	10	0,72324	10	0,52872	10	0,96214	10	1	17	-1,2914	19	-0,92362	18	-0,35142	2	0,32577
8	1	8	0,69181	11	1,1253	11	1	11	1	11	1	13	0	13	0,46375	13	0,30521	13	1,1978
9	0,1204	9	0,82304	12	0,49882	12	0,34431	12	0,52699	12	0,4708	14	0,69006	14	0,88851	14	0,58677	14	0,95152
19	-0,83994	19	-1,3286	21	-1	21	-1	21	-0,84408	21	-1	23	-1,0996	23	-0,75559	23	-0,68389	23	-0,78264
20	-1,8206	20	-1	22	-0,42178	22	-0,25056	22	-0,18634	22	-0,50207	24	-1,0823	24	-0,70716	24	-0,55863	24	-0,7324

Tabela 15: Erros e acertos nos conjuntos de treinamento, validação e teste para cada caso da validação cruzada (configuração H2a).

CONFIGURAÇÃO H2a (saída "PA sistólica")						
CASOS	GRUPOS TREINAMENTO	GRUPOS TESTE	GRUPOS VALIDAÇÃO	RMSE TREINAMENTO	RMSE TESTE	RMSE VALIDAÇÃO
1	345	2	1	29	19	10
2	245	3	1	16	13	16
3	235	4	1	12	20,447	15
4	234	5	1	22	34	11
5	345	1	2	13,887	23	11,467
6	145	3	2	22	16	16
7	135	4	2	14	23	18
8	134	5	2	12	24,501	10,588
9	245	1	3	17	16	14
10	145	2	3	17	12	14
11	125	4	3	17	18	13
12	124	5	3	16	32	16
13	235	1	4	14	19,319	15
14	135	2	4	19	20	20
15	125	3	4	21	16	17
16	123	5	4	15	27	20
17	234	1	5	11	28	22
18	134	2	5	16	18,435	17
19	124	3	5	18	15	16,073
20	123	4	5	18	21	13

Tabela 16: Erros para cada entrada em cada caso da validação cruzada (configuração H2a).

ERROS SISTÓLICA											
CASO 1											
4	31										
5	1										
6	-6										
17	-3										
18	-28										
7	11										
8	40										
9	3										
19	-8										
24	-26,973										
10	24										
22	-27										
11	34										
21	-5										
23	-29										
12	13										
20	-28										
14	50										
13	52										
1	-1,6602										
2	2,3358										
3	13										
15	-15,793										
16	-7										
CASO 2											
7	20										
8	0,44784										
9	9										
19	0										
20	-20										
11	28										
5	11										
10	17										
4	12										
21	-6										
23	-19,232										
18	-20										
13	-3										
24	-25,433										
6	1										
22	-20										
14	7										
17	-8										
12	18										
1	-6										
2	9										
3	20										
15	-17,933										
16	-19										
CASO 3											
10	20										
11	30										
12	23										
21	4										
22	-15,223										
24	-17										
13	-1										
6	5										
7	25										
8	-2										
9	15										
20	-15										
17	3										
19	5										
23	-15										
4	5										
14	2										
18	-15,136										
5	15										
1	-1										
2	15										
3	25										
15	-12,898										
16	-11										
CASO 4											
13	38										
14	41,241										
23	-30										
24	-27										
12	12										
20	-29										
18	-31										
17	-7										
5	3										
7	13										
6	-10										
9	0										
19	-9										
10	23										
22	-32										
8	31										
11	32										
4	33										
21	-8										
1	-11										
2	0										
3	9										
15	-18										
16	-7										
CASO 5											
1	-14										
2	12										
3	24										
15	-33										
16	-24										
11	15										
21	0										
10	-2										
24	-32										
9	11										
23	-12,93										
8	5										
19	0										
20	-18,9										
22	-17										
14	9										
7	19										
12	8										
13	0										
4	4										
5	17,119										
6	6										
17	-9										
18	-14,99										
CASO 6											
7	4										
8	18,596										
9	18										
19	-2										
20	-25										
22	1										
11	20										
21	-19										
16	-17										
13	31										
3	31										
2	16										
23	-9										
14	30,061										
15	-28										
24	-40										
12	7,6715										
1	-2										
10	8										
4	20										
5	19,026										
6	3,0238										
17	-22										
18	-9										
CASO 7											
10	23										
11	36,691										
12	22										
21	-6										
22	-18										
20	-18										
24	-19										
14	13										
23	-18										
15	-10										
9	11										
19	2										
7	22										
2	12										
3	21,575										
1	-5										
2	13										
16	-3										
13	13										
8	3										
4	32										
5	12										
6	2										
17	-7										
18	-18										
CASO 8											
13	-5										
14	34										
23	-18										
24	-30										
12	10,979										
20	-21										
11	-7										
14	13										
8	5										
1	2										
16	-3										
3	21										
9	12,604										
7	7										
22	-17,07										
2	13										
19	-5										
21	-6,927										
15	-16										
4	3										
5	13										
6	3										
17	-4										
18	-19										
CASO 9											
1	-10										
2	10										
3	23										
15	-19										
16	-13										
18	-15										
22	-18										
24	-28,81										
10	10										
6	4										
21	-11										
13	13										
12	17,521										
11	26										
17	-11										
23	-11										
4	21										
14	21										
5	19										
7	21										
8	9										
9	9										
19	2										
20	-19										
CASO 10											
4	6										
5	10										
6	0										
17	-14										
18	-20										
23	-20										
16	-21										
22	-20										
11	22										
10	5										
1	-6										
3	20										
15	-20										
12	19										
14	2										
13	-3,226										
21	-12										
2	9,2988										
24	-28										
7	19										
8	-9										
9	9										
19	0										
20	-20										
CASO 11											
10	11										
11	26										
12	21										
21	-3										
22	-21										
18	-15										
17	-6										
6	7										
2	6										
5	24										
3	24										
23	-6,6548										
24	-22										
15	-24										
16	-17										
1	-24										
14	8,9461										
4	21,142										
13	-1										
7	25										
8	0										
9	5										
19	4,3709										
20	-15										
CASO 12											
13	33										
14	35,693										
23	-20										
24	-37										
10	10										
16	-21										
11	22										
3	20										
22	-20,122										
21	-9										
15	-26										
12	12										
5	10										
4	17										
6	0,76248										
18	-22										
17	-13										
2	7										
8	-9										
9	10										
7	6										
8	26										
9	7										
19	-5										
20	-22										
CASO 13											
1	-11										
2	10										
3	21										
15	-25										
16	-24										
4	4										
17	-17										
6	1										
24	-28,419										
23	-19										
19	0										
18	-19										
20	-20										
13	-1										
5	10,883										
7	20										
14	1										
8	-9										
9	10										
10	-5										
11	17,647										
12	18										
21	-7										
22	-19										
CASO 14											
4	26										
5	18										
6	-8										
17	-16,36										
18	-27										
16	-14										
20	-26										
2	-1										
24	-32										
13	23										
1	-16										
23	-20										
15	-22										
8	15										
19	-4										
7	12										
9	-1										
14	26										
3	14										
10	12										
11	25										
12	16										
21	-18										
22	-27										
CASO 15											
7	7										
8	17										
9	1										
19	-11										
20	-27										
23	-24										
5	6										
1	-9										
2	3										
6	-4										
16	-25										
14	25										
18	-28										
24	-38										
15	-29										
13	19										
3	13										
17	-16										
4	10,047										
10	8										
11	17										
12	6										
21	-16										
22	-27										
CASO 16											
13	27										
14	29,105										
23	-20										
24	-29,107										
1	-6										
9	6										
16	-17										
15	-20										
5	11										
7	16										
18	-22										
20	-21										
3	19										
17	-11										
6	0										
2	8										
8	19										
19	-3										
4	20										
10	18										
11	26										
12	17										
21	-13										
22	-22										
CASO 17											
1	2										
2	16										
3	28										
15	-37										
16	-37,758										
20	-14										
21	-2										
6	9,1703										
17	-11										
11	5										
4	-6										
7	12										
18	-13										
9	15										
12	8										
8	9										
19	0										
5	19										
16	-7										
7	20										
2	9										
13	7,4894										
14	10										
23	-19										
24	-24										
CASO 18											
4	32										
5	11										
6	0										
17	-14										
18	-20										
11	32										
22	-20										
19	0										
10	13										
20	-20										
15	-15										
21	-9										
8	-1										
3	20										
12	20										
1	-10										
9	9										
16	-7										
7	20										
2	9										
13	7,4894										
14	10										
23	-19										
24	-24										
CASO 19											
7	24										
8	7										
9	14										
19	3										
20	-16										
12	21										
4	31										
11	32										
1	-4										
22	-16										
2											

Tabela 17: Erros e acertos nos conjuntos de treinamento, validação e teste para cada caso da validação cruzada (configuração H2b).

CONFIGURAÇÃO H2b (SAÍDA "PA DIASTÓLICA")						
CASOS	GRUPOS TREINAMENTO	GRUPOS TESTE	GRUPOS VALIDAÇÃO	RMSE TREINAMENTO	RMSE TESTE	RMSE VALIDAÇÃO
1	345	2	1	19	26	17
2	245	3	1	31	27	13
3	235	4	1	22	16	13
4	234	5	1	25	17,964	11
5	345	1	2	21	18	16
6	145	3	2	14	37	15
7	135	4	2	24,45	11	15,349
8	134	5	2	24	16	16
9	245	1	3	13	17	33
10	145	2	3	16	28	26
11	125	4	3	12	14,632	31
12	124	5	3	14	15	32
13	235	1	4	23	17	11
14	135	2	4	20	21	9
15	125	3	4	15	31	13
16	123	5	4	24	18	7
17	234	1	5	21	16,68	15
18	134	2	5	20	21	10
19	124	3	5	16	33	8
20	123	4	5	21	17	9

Tabela 18: Erros para cada entrada em cada caso da validação cruzada (configuração H2b).

ERROS DIASTÓLICA																					
CASO 1	CASO 2	CASO 3	CASO 4	CASO 5	CASO 6	CASO 7	CASO 8	CASO 9	CASO 10	CASO 11	CASO 12	CASO 13	CASO 14	CASO 15	CASO 16	CASO 17	CASO 18	CASO 19	CASO 20		
4	24	7	9	10	-13,007	13	27	1	8	7	2	10	-3	13	19,393	14	10	1	4	4	34
5	43	8	44	11	17	14	20	11	17	8	33	11	17	14	10	23	-10	15	-32	17	-5
6	23	9	-33	12	25	24	-4	3	15	9	-70	12	15	24	-2	21	-2	23	-10	16	-15,64
17	-15	19	-5	21	-16	21	0	19	-1	21	-2	24	-21	15	-32	16	-10	15	-32	18	3
18	-13	20	-20	22	-2	24	-10	22	-10	22	-10	16	-10	21	-10	21	-10	21	-10	11	5
20	-20	23	-13	20	-18	17	10	24	-26	13	22	20	-22	13	22	20	-22	15	-31	14	3,3662
23	5	18	-23	13	10	9	-54	19	0,37624	13	11	14	1	11	9	3	10	22	-17	4	21
19	-4,0566	22	-17	6	23	19	15	18	-6	8	40,008	10	1	8	21	12	5	8	38	10	-12
21	-16	11	17	19	15	6	22	10	-18	16	-14	15	-21	2	-19,29	2	-19,29	17	-13	17	-13
11	5	10	3	18	-6	8	40,008	10	-18	16	-14	15	-21	2	-19,29	2	-19,29	17	-13	22	10
13	19,353	24	-12	9	-47,864	10	1	8	21	12	5	8	38	10	-12	17	-13	22	10	24	10
8	24	4	39	8	19,404	12	16	20	-19	1	1	19	8	12	1	1	1	1	1	5	9
7	6	14	24	7	18	20	-20	19	10	10	-11	16	-8	8	30	8	30	24	-25	14	12
12	7,7021	6	40	14	6	7	12	11	4	15	-30,516	7	12	7	12	19	-6	22	-13	11	16
24	-26,413	17	21	23	11	18	-20	12	15	2	-18,326	24	-12	3	1	10	-17,258	1	14,33	15	-19
10	-11	5	86	24	-22	11	19	7	5	21	-6,3099	9	-62	11	9	9	-62	23	-2	12	27
14	4	12	12	5	27	5	28,19	23	5	22	-12	1	8	1	8	1	-3	12	6	3	25
9	-47	21	-2	4	31	4	39	22	-5	3	7	23	-0,5901	7	0	7	0	18	-16	23	23,941
22	-8	13	25	17	-2	22	-17	9	-57	14	10,07	2	-11	20	-33	6	9	6	9	10	-5
1	11	1	12	1	20	1	11	4	21	4	25	4	30	4	30	4	30	7	2	7	19
2	-14	2	-4	2	0	2	-11	5	15,512	5	1,3077	5	7	5	7	5	7	8	23	8	32
3	0	3	-4	3	15	3	2	6	15	6	8	6	10	6	10	6	10	9	-64	9	-40
15	-30	15	-21	15	-13	15	-19	17	-18	17	-10	17	-6	17	-6	17	-6	19	0	19	18
16	-11	16	13	16	3	16	3	18	-6	18	-15	18	-10	18	-10	18	-10	20	-25	20	-3
10	-4,7149	13	19	1	9	4	31	7	13	13	28	1	7	13	28	1	7	7	12	10	-19
11	17,742	14	10	2	-4	5	25	8	27	14	14	2	-10	8	27	5	21,335	8	16	11	21
12	20	23	1	3	3	6	24	9	-58	23	-3	3	9	9	-58	3	9	9	-62,615	12	22
21	-15,391	24	-20	15	-35	17	-8	19	10,508	24	-16,924	15	-32	17	-14	15	-32	19	6,7598	21	-13
22	-9	15	-29	16	-14	18	-9	20	-21,003	17	-8	16	-12	20	-23	16	-12	20	-31	22	-0,54025
1	8	17	-10	9	-41,848	14	11	15	-19,494	6	24	20	-11	6	12	20	-23	9	-52	6	16
6	12	10	-10	19	-6	20	-11	6	12	2	-11	10	-12	2	-11	10	-12	19	18	5	15
5	12,753	5	11	18	-16	13	20	13	20	20	-18	9	-60	20	-18	9	-60	15	-18	10	-12
18	-10	11	10	23	11	3	13	3	13	8	38	21	-12	3	13	3	21	15	-17	19	16
4	28	3	10	24	-27	9	4	4	32	4	30	18	-13	4	30	18	-13	2	-1	18	-15
2	-9	16	-9	2	-5	24	-16	7	3	7	3	6	11	7	3	6	11	7	21	15	-17
3	11,514	12	9	4	23	1	12	16	-2	18	-9	5	10,859	21	-8	5	10,859	21	-8	3	19
15	-15,03	21	-9	7	-1	15	-28	1	9	15	-29	4	28	8	25	4	28	8	25	9	-55
16	-8	22	-9	17	-16	13	21	23	2	3	16,063	17	-12,011	17	-12,011	22	2	12	13	8	19
17	-11	2	-11	8	22	3	10	19	-4	19	-4	8	28	8	28	8	28	1	9	17	-16
13	4	1	7	6	30	24	-19	14	11	5	12	12	4	10	-8	12	4	10	-8	1	10
24	-10,77	4	29	14	3	8	31	5	12	1	6,552	11	8	11	8	11	8	16	-4	16	-1
14	-1	18	-10	23	15	16	-6	18	-8	9	-59	12	24	22	-9,3993	12	24	3	7	4	33
23	2	6	11	13	12	19	-1	17	-11,932	16	-9	7	-1	7	-1	11	18	7	-9	20	-18
7	18,838	7	5	10	-17	10	-9	10	-10	10	-7	19	-3	19	-3	20	-11	4	44	7	15
8	16	8	29	11	3	11	11	11	16	11	11	13	18	13	18	13	18	13	18	13	0
9	-61	9	-60	12	6	12	13	12	17	12	4	14	8	14	8	14	8	14	0	14	7,4392
19	11	19	5	21	-17	21	-9	21	-12	21	-6	23	0	23	0	23	11,319	23	-3,4725	23	8
20	-19	20	-23	22	-0,78616	22	0,16836	22	-7	22	6	24	-22	24	-22	24	-22	24	-13	24	-11

4.2 Divisão dos Dados

Em Redes Neurais Artificiais, normalmente os dados são divididos em três conjuntos: treinamento, validação e testes. O conjunto de treinamento é utilizado para treinar a rede neural, o conjunto de validação é utilizado para parar o treinamento e evitar sobre-treinamento (*overtraining*) e o conjunto de teste é utilizado para calcular o desempenho da rede.

Em relação ao processo para obter a melhor rede, [BISHOP, 1995] afirma que a abordagem mais simples é treinar diversas redes usando o conjunto de treinamento, utilizar o conjunto de validação para comparar o desempenho dessas redes e o conjunto de teste para confirmar os resultados obtidos. Esse método é chamado de *hold out*. Como este método usa apenas uma parcela dos dados (o conjunto de treinamento) para treinar as RNAs e a outra parcela (o conjunto de testes) para medir o desempenho, as métricas obtidas podem não ser tão apuradas quando se utiliza um conjunto de dados não muito grande. A validação cruzada resolve isto ao dividir os dados em k segmentos de n/k elementos. O método de validação cruzada chamado de *k-fold* consiste em treinar a rede utilizando $k-1$ segmentos e testar a rede utilizando o segmento restante. Este processo é repetido para as k possíveis escolhas do conjunto de teste. Se o valor de k é igual a n , o método é chamado de *leave-one-out*. O problema destes dois métodos (*k-fold* e *leave-one-out*) é que mais esforço e tempo computacional são necessários até que o processo termine.

Neste trabalho, o método de validação cruzada *k-fold* é utilizado, mas são utilizados três conjuntos e não apenas dois, ou seja, $k-2$ segmentos são utilizados para treinar a rede, 1 segmento é utilizado para impedir que ocorra sobre-treinamento (*overtraining*) e o segmento restante é utilizado para testar a rede. Com isso, o processo é repetido $k * (k-1)$ e não apenas k vezes. A Figura 12 mostra este processo de divisão em segmentos, da criação das $k*(k-1)$ redes, da escolha da melhor rede e da captação das medidas (média e desvio) com base no conjunto de testes.

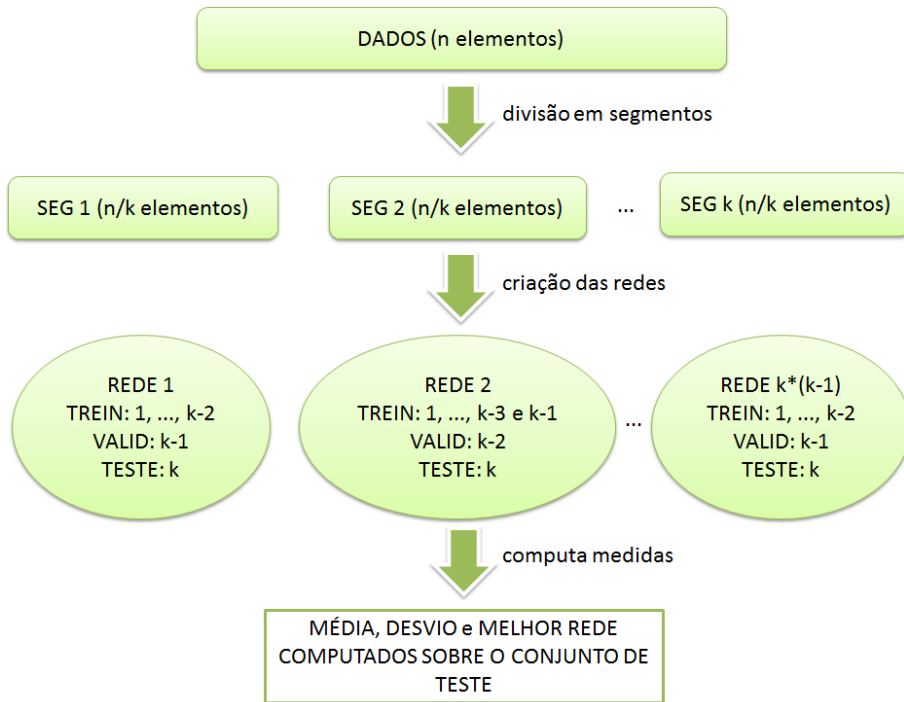


Figura 12: Validação cruzada

Capítulo 5. Experimentos e Resultados

Este capítulo está dividido em duas seções (5.1 e 5.2), as quais são referentes aos experimentos realizados para a tentativa de previsão da hipertensão arterial crônica e da doença periodontal, respectivamente. Como mencionado anteriormente, as Redes Neurais Artificiais são utilizadas com este propósito e os dados são tratados utilizando as técnicas de pré-processamento descritas no Capítulo 4. Inicialmente, foi utilizado um conjunto de dados com 24 elementos, os quais são relativos aos dados de indivíduos do sexo feminino no período pós-parto e, depois, o conjunto diminui devido à aplicação da técnica de detecção de intrusos.

Em relação às RNAs utilizadas em todos os experimentos deste capítulo, elas são do tipo Perceptron Multicamadas com duas camadas (uma intermediária e uma de saída). O algoritmo de treinamento utilizado é o *resilient backpropagation*. O número de neurônios na camada de saída varia de experimento para experimento, uma vez que depende do número de saídas da RNA. As médias e desvios são apresentados para diferentes quantidades de neurônios na camada intermediária em cada experimento e as porcentagens dos dados utilizados nos conjuntos de treinamento, validação e teste são 70%, 15% e 15%, respectivamente. Em todos os experimentos que utilizam a validação cruzada k-fold, o valor de k é 5.

5.1 Hipertensão Arterial Crônica

Esta seção descreve todos os experimentos e apresenta os resultados encontrados nas tentativas de prever a hipertensão arterial crônica e de prever a pressão arterial de indivíduos utilizando redes neurais.

5.1.1 Experimento 1

Este experimento tem como finalidade tentar prever a hipertensão arterial crônica utilizando as técnicas de pré-processamento preenchimento de dados faltantes, normalização *z-score* e seleção de variáveis (algoritmo 2) e a técnica de divisão dos dados chamada *hold out*. Em relação à configuração da rede utilizada, esta consiste de uma rede *feedforward* com treinamento *resilient backpropagation* com duas camadas de neurônios. O número de

neurônios na camada escondida varia (são realizados treinamentos para redes com 1 até 20 neurônios) e o número de neurônios da camada de saída é igual a 1, já que este experimento utiliza a saída “hipertenso” (diz se um indivíduo sofre de hipertensão ou não) para prever a hipertensão arterial crônica. O neurônio da saída é do tipo tangente hiperbólica e, portanto, valores entre 0 e 1 são interpretados como hipertenso e valores entre -1 e 0 são interpretados como saudável. Em relação à seleção de variáveis, o primeiro algoritmo descrito na seção 4.1.5 foi utilizado para selecionar, nas configurações H1, H3 e H5 da seção 4.1.4, as variáveis mais relevantes a serem utilizadas. Na tabela 19, estão os detalhes das redes utilizadas neste experimento. Nela, a coluna “rede” identifica o tipo de rede utilizada, a coluna “divisão” especifica a técnica de divisão de dados utilizada, a coluna “neurônios saída” especifica a quantidade de neurônios na camada de saída por tipo (tgh para tangente hiperbólico e ln para linear), a coluna “neurônios escondidos” especifica a quantidade de neurônios na camada escondida por tipo e a coluna “pré-processamento” especifica os algoritmos de pré-processamento utilizados. Já a tabela 20 apresenta as variáveis de entrada e variáveis de saída para cada uma das 3 redes utilizadas neste experimento.

Tabela 19: Detalhes das redes utilizadas no experimento 1.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward com treinamento backpropagation	Hold out	1 tgh	De 1 a 20	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 2) e Z-score	“hipertenso”

Tabela 20: Variáveis de entrada e saída utilizadas nas redes do experimento 1.

Código	Variáveis de entrada	Variáveis de saída
E1R1	“x13”, “x14”, “x15”	“hipertenso”
E1R2	“IMC” e “peso pós-parto”	“hipertenso”
E1R3	“x13”, “x14”, “x15”, “IMC” e “peso pós-parto”	“hipertenso”

Os resultados do experimento são apresentados nas tabelas 21, 22 e 23. As tabelas são relativas às redes E1R1, E1R2 e E1R3. Nelas, são apresentados a porcentagem de acerto da rede e o desvio padrão para cada número de neurônios da camada escondida. Estes valores foram obtidos sobre o conjunto de testes.

Tabela 21: Resultado do experimento 1 utilizando a rede E1R1 para diferentes números de neurônios.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	54	22,8611
2	53,75	17,8351
3	52,75	17,3566
4	46,5	16,2303
5	55,75	20,1491
6	48,5	25,757
7	50,25	19,8332
8	53,5	16,9442
9	60,75	18,8676
10	53,75	16,6129
11	54,25	17,7167
12	50,75	20,8551
13	50,5	20,1899
14	54,25	20,5372
15	59,75	18,4587
16	48,25	25,4577
17	60	23,6754
18	59	23,9846
19	47,25	26,7284
20	56,5	23,9572

Tabela 22: Resultado do experimento 1 utilizando a rede E1R2 para diferentes números de neurônios.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	92	10,0525
2	89	16,5116
3	89,75	10,5724
4	92	10,0525
5	92	10,0525
6	91	12,0961
7	91	10,2084
8	90	13,7649
9	89	12,0961
10	92	10,0525
11	87,5	13,3278
12	92	10,0525
13	92	10,0525
14	91	10,2084
15	89	15,1831
16	92	10,0525
17	92	10,0525
18	92	10,0525
19	90	12,1395
20	90	15,2177

Tabela 23: Resultado do experimento 1 utilizando a rede E1R3 para diferentes números de neurônios.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	78,5	18,9251
2	80,5	21,0825
3	77,25	18,2436
4	86,5	13,7745
5	82,25	17,7316
6	81,25	14,5887
7	81	17,2901
8	79	16,0263
9	80,5	23,8912
10	84	13,3377
11	82,25	17,7316
12	81,5	18,3604
13	83,25	14,2603
14	80	18,5671
15	74,75	28,1666
16	81,5	19,4733
17	79,75	14,8213
18	81,75	17,1122
19	82	19,695
20	82,75	16,3413

Analisando a tabela 21, relativa à rede E1R1 que utiliza apenas os metabólitos salivares como variáveis de entrada, percebe-se que os resultados médios variam entre 50% e 60,75% de acerto, e o melhor resultado é relativo à quantidade de 9 neurônios com uma taxa média de acerto de 60,75%. Já com relação à tabela 22, relativa à rede E1R2 que não utiliza os metabólitos salivares como variáveis de entrada, os resultados médios quase não variam e ficam em torno de 90% de acerto (resultado muito superior ao da rede E1R1). Uma justificativa para a alta taxa de acerto da rede E1R2 pode ser baseada no fato de que existe uma associação comprovada entre peso e pressão sanguínea, conforme diversos estudos na área ([HUANG *et al.*, 1998], [STAMLER R *et al.*, 1978], [HAVLIK *et al.*, 1983], [BARKER, 2006], etc).

5.1.2 Experimento 2

Este experimento tem como finalidade prever as pressões sistólica e diastólica dos indivíduos utilizando as técnicas de pré-processamento preenchimento de dados faltantes,

normalização *z-score* e seleção de variáveis (algoritmo 2) e a técnica de divisão dos dados chamada *hold out*. Em relação à configuração da rede utilizada, esta consiste de uma rede *feedforward* com treinamento *resilient backpropagation* com duas camadas de neurônios. O número de neurônios na camada escondida varia (são realizados treinamentos para redes com 1 até 20 neurônios) e o número de neurônios da camada de saída é igual a 2, já que este experimento utiliza as saídas “PA sistólica” e “PA diastólica” (pressões sistólica e diastólica dos indivíduos) para prever a pressão arterial. Os neurônios da saída são do tipo linear, já que se deseja obter na saída os valores das pressões (eles estarão normalizados, uma vez que os dados de entrada da rede também estão). Em relação à seleção de variáveis, o primeiro algoritmo descrito na seção 4.1.5 foi utilizado (algoritmo 2) para selecionar, nas configurações H2, H4 e H6 da seção 4.1.4, as variáveis mais relevantes a serem utilizadas. Na tabela 24, estão os detalhes das redes utilizadas neste experimento. Já a tabela 25 apresenta as variáveis de entrada e variáveis de saída para cada uma das 3 redes utilizadas neste experimento.

Tabela 24: Detalhes das redes utilizadas no experimento 2.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward treinamento backpropagation	com Hold out	2 In	De 1 a 20	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 2) e Z score	“PA sistólica” e “PA diastólica”

Tabela 25: Variáveis de entrada e saída utilizadas nas redes do experimento 2.

Código	Variáveis de entrada	Variáveis de saída
E2R1	“x124”, “x131” e “x125”	“PA sistólica” e “PA diastólica”
E2R2	“IMC” e “peso pós-parto”	“PA sistólica” e “PA diastólica”
E2R3	“x124”, “x131”, “x125”, “IMC” e “peso pós-parto”	“PA sistólica” e “PA diastólica”

Os resultados do experimento são apresentados nas tabelas 26, 27 e 28. As tabelas são relativas às redes E2R1, E2R2 e E2R3, respectivamente. Nelas, são apresentados a média do erro absoluto da rede (o módulo da diferença entre a saída prevista pela rede desnormalizada

e a saída real desnormalizada) e o desvio padrão deste erro para cada número de neurônios da camada escondida.

Tabela 26: Resultado do experimento 2 utilizando a rede E2R1 para diferentes números de neurônios.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	26,9185	11,7293	29,2691	11,573
2	25,1208	10,1578	27,3533	8,65
3	25,4358	8,402	27,9318	12,6845
4	22,1916	6,428	29,0851	11,5428
5	26,0225	8,2857	29,3363	9,7504
6	24,5753	5,8205	27,1439	12,0098
7	26,3638	10,5587	27,8852	12,7391
8	23,3433	6,9974	27,8546	9,7812
9	24,7889	4,9561	23,8715	8,7981
10	26,597	10,4846	29,4251	12,4387
11	26,725	9,047	27,9481	10,1756
12	24,5506	10,7333	26,9536	9,5656
13	23,5905	7,4721	29,317	10,1265
14	25,4663	9,9577	27,0071	11,6266
15	24,9828	7,4415	33,7881	16,656
16	23,7235	7,0293	25,7598	9,0136
17	23,1922	7,3831	26,3758	12,301
18	25,5332	6,5815	30,9213	9,603
19	24,6674	7,2388	24,592	10,5168
20	24,5941	7,9493	25,6432	12,0651

Tabela 27: Resultado do experimento 2 utilizando a rede E2R2 para diferentes números de neurônios.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	21,4486	6,5241	27,884	10,9058
2	20,965	5,5072	27,5141	9,6846
3	21,2386	7,3811	25,898	8,6541
4	20,7896	8,3296	31,0499	12,7342
5	21,042	8,732	25,807	10,3884
6	21,2454	8,116	25,3202	12,7806
7	18,6995	6,8605	25,5357	11,0343
8	24,8871	9,1527	31,0448	12,2246
9	21,1019	5,8714	30,3987	12,094
10	20,0375	8,7272	25,767	11,5307
11	23,0957	11,0834	26,6017	10,1622
12	22,5295	8,807	29,2862	11,7319
13	18,8865	7,2113	27,369	9,4708
14	19,5176	8,9413	29,5404	13,1025
15	20,1514	5,4752	23,5097	8,9944
16	19,4852	6,4908	29,0287	10,3113
17	21,3824	7,118	27,9133	9,0228
18	22,2297	8,2673	29,1388	11,8872
19	22,0925	6,8709	26,5548	12,3378
20	23,1462	10,3963	26,7447	10,7935

Tabela 28: Resultado do experimento 2 utilizando a rede E2R3 para diferentes números de neurônios.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	22,0537	10,5505	28,4084	9,7997
2	22,8968	13,0871	28,1079	14,3843
3	21,3073	6,3896	27,0481	8,7504
4	19,3319	7,4927	28,5866	11,8388
5	19,6102	7,7179	26,7718	9,8526
6	22,5567	11,1041	30,071	10,6928
7	20,7012	8,2882	26,9146	8,679
8	20,071	6,3827	29,8711	13,3297
9	18,378	5,6975	26,6245	9,4575
10	19,8271	8,0666	26,2245	11,2641
11	19,908	7,3192	26,9173	9,9165
12	19,522	8,381	30,8757	9,472
13	19,1385	8,1966	27,2619	8,1977
14	17,9452	6,3665	30,1741	13,175
15	20,8558	8,3749	27,9207	10,223
16	18,7763	5,1521	26,5645	11,6398
17	20,3136	8,9852	25,5554	10,6299
18	21,3297	8,9686	27,9325	10,6351
19	20,0831	5,4964	22,9908	11,1121
20	18,1762	5,6286	30,1373	7,4287

A partir da análise das tabelas, é possível notar que, para a rede E2R1, a média do erro variou de 22,19 (4 neurônios) a 26,72 (11 neurônios) para a pressão sistólica e de 23,87 (9 neurônios) a 33,78 (15 neurônios) para a pressão diastólica. Em relação aos resultados da rede E2R2, a erro para a pressão sistólica variou de 18,69 (7 neurônios) a 24,88 (8 neurônios) e para a pressão diastólica variou de 25,32 (6 neurônios) a 31,04 (4 neurônios). Já em relação aos resultados da rede E2R3, o erro variou de 17,94 (14 neurônios) a 22,89 (2 neurônios) para a pressão sistólica e de 25,5 (17 neurônios) a 30,8 (12 neurônios) para a pressão diastólica. Ao analisar os dados e resultados, verificou-se que a pressão sistólica variou de 100 a 170 e o melhor resultado (tabela 28 com 14 neurônios) foi referente ao erro de 17,94, o qual corresponde a 24% da diferença entre o valor mais alto (170) e o valor mais baixo (100). Percebe-se que os resultados foram insatisfatórios para a pressão sistólica. O resultado da pressão diastólica também foi insatisfatório, pois se sabe que, nos dados, ela varia entre 10 e 120 e o menor erro foi 23,87, o qual corresponde a 20% da diferença entre o menor valor e o maior valor, ou seja, 20% de 110.

5.1.3 Experimento 3

Analisando o experimento 1, notou-se que, em geral, o resultado da rede E1R3 foi pior que a rede E1R2. Como as variáveis de entrada da rede E1R3 são formadas pelas variáveis da rede E1R2 mais os metabólitos salivares identificados na seleção de variáveis em relação à saída hipertenso, era esperado que o resultado da rede fosse melhor, visto que a rede deveria ser capaz de dar pouca importância para as variáveis que a atrapalham. Então, a rede deveria ser capaz de identificar que os metabólitos salivares atrapalhavam a rede e, portanto, ela deveria atribuir pouco peso a estas. Pensando nisso, o experimento 3 foi desenvolvido. Nele, os pesos são inicializados com valores baixos (0,01), diferindo do experimento 1, onde os pesos foram inicializados utilizando o algoritmo Nguyen-Widrow, o qual utiliza de uma certa aleatoriedade para inicializar os pesos. Partindo do princípio que as variáveis metabólicas estavam atrapalhando o desempenho, a melhor rede seria aquela em que as sinapses associadas a tais variáveis apresentassem valor nulos ou muito próximos de zero. Deste modo, esta inicialização em 0,01 ajudaria a rede a dar uma menor importância às variáveis metabólicas e, conseqüentemente, ocorreria uma melhora nos resultados. Os resultados do experimento 3 são apresentados nas tabelas 29, 30 e 31.

Tabela 29: Resultado do experimento 3 utilizando a rede E1R1 para diferentes números de neurônios.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	54,25	13,5019
2	56	6,8056
3	54,25	11,8405
4	52	11,0501
5	53,25	12,1693
6	55	7,6089
7	53,5	14,8767
8	58,25	6,7424
9	55,25	11,4104
10	55,25	11,4104
11	56,75	16,88
12	55	12,8759
13	54,25	11,8405
14	57,5	11,528
15	53,25	12,1693
16	58,5	8,5993
17	55,75	17,2653
18	55	7,6089
19	54,5	14,5909
20	58,75	12,965

Tabela 30: Resultado do experimento 3 utilizando a rede E1R2 para diferentes números de neurônios.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	92	10,0525
2	85	21,3985
3	88	18,8065
4	90	15,2177
5	92	10,0525
6	90	15,2177
7	92	10,0525
8	91	12,0961
9	89	16,5116
10	91	12,0961
11	91	12,0961
12	88	18,8065
13	89	16,5116
14	91	12,0961
15	90	15,2177
16	90	15,2177
17	90	13,7649
18	90	13,7649
19	91	12,0961
20	89	16,5116

Tabela 31: Resultado do experimento 3 utilizando a rede E1R3 para diferentes números de neurônios.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	79	18,4676
2	84,75	19,4311
3	83,25	18,1568
4	82,5	18,8135
5	84,75	17,1276
6	85,5	16,295
7	84,5	17,2367
8	83,5	18,0715
9	83,5	15,5682
10	84,5	15,9687
11	84,5	14,5909
12	85,75	17,4171
13	85,5	16,295
14	86	18,4676
15	87,5	12,4076
16	84,5	18,4177
17	85,75	17,4171
18	85,75	16,1632
19	86,5	15,2264
20	84,5	17,2367

A partir da análise da tabela 29 (relativa à rede E1R1), é possível verificar que os resultados variam pouco com o aumento do número de neurônios. O melhor resultado é relativo a 8 neurônios com uma taxa média de acerto de 58,25%. Em relação à tabela 30 (relativa à rede E1R2), o resultado varia entre 85% e 92% de acerto. A melhor taxa média de acerto é de 92% e é relativa a 1 neurônio. A partir da análise da tabela 31 (relativa à rede E1R3), é possível verificar que os resultados variam entre 79% e 87,5% e a melhor taxa média de acerto é referente a 15 neurônios. Esperava-se que o resultado referente à rede E1R3 (tabela 31) apresentasse resultado igual ou superior ao resultado da rede E1R1 (tabela 29), o que não aconteceu, apesar de os resultados terem sido um pouco melhores quando comparados ao experimento 1.

5.1.4 Experimento 4

Este experimento é igual ao experimento 2, diferindo apenas na inicialização dos pesos, os quais são inicializados semelhantemente ao experimento 3, ou seja, eles são fixados com um valor baixo (0,01). As tabelas 32, 33 e 34 são referentes aos resultados obtidos para as redes E2R1, E2R2 e E2R3, respectivamente.

Tabela 32: Resultado do experimento 4 utilizando a rede E2R1 para diferentes números de neurônios.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	21,0302	6,1297	21,3509	8,8486
2	21,3769	6,029	19,8201	8,5131
3	20,5843	6,1063	20,64	8,6743
4	20,9417	6,1341	20,1491	8,3822
5	21,4558	6,9273	20,616	8,5693
6	21,0836	5,9182	20,261	8,6891
7	21,6075	6,4011	21,5355	8,4085
8	21,2668	5,854	21,7264	9,0497
9	20,5743	6,8605	21,0943	8,9711
10	21,6406	5,9939	20,0182	9,4503
11	21,6381	6,823	20,1527	8,4666
12	20,5386	5,824	20,1122	8,5839
13	21,5976	6,244	21,0809	7,9609
14	21,1765	6,3713	20,2871	8,114
15	21,6396	6,8419	21,2831	8,6751
16	20,926	6,3206	21,5083	9,7586
17	21,1877	6,1254	20,1858	9,0224
18	21,715	6,4883	21,7143	9,2403
19	21,9244	6,6931	20,5584	8,7523
20	21,066	6,1813	21,1229	9,4425

Tabela 33: Resultado do experimento 4 utilizando a rede E2R2 para diferentes números de neurônios.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	20,7656	8,1832	22,1718	9,3371
2	19,4785	8,5446	21,9599	9,0298
3	19,2624	7,1787	22,1128	9,149
4	19,8929	7,6044	21,6958	9,4796
5	20,1267	7,4	21,931	9,4781
6	19,8775	7,6535	21,5848	9,5042
7	19,8225	8,5095	22,3048	9,5468
8	20,5702	8,0403	21,5442	9,1999
9	20,09	7,2219	22,5168	9,1261
10	19,8175	7,9171	22,4909	8,8152
11	20,1996	8,0873	22,1835	9,043
12	19,9464	7,8406	21,8941	9,6268
13	19,734	7,3385	22,0448	9,4711
14	20,0642	6,9478	21,7643	9,7516
15	19,4435	7,7294	21,4722	9,8314
16	19,7481	7,5908	21,808	9,4931
17	20,4404	6,5121	21,5218	9,9338
18	19,5049	7,3476	21,1643	9,4801
19	19,9276	7,789	21,8287	9,4202
20	19,5799	8,034	22,0675	9,1999

Tabela 34: Resultado do experimento 4 utilizando a rede E2R3 para diferentes números de neurônios.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	18,5548	7,4462	20,9108	8,9832
2	18,0055	5,8756	19,6876	10,3672
3	17,8822	5,937	20,9215	9,1685
4	17,8035	5,9404	21,9989	8,8783
5	17,4196	5,9315	20,5561	9,8964
6	17,7025	6,7972	20,8461	9,6112
7	19,1357	7,672	21,2606	8,9366
8	18,8187	6,015	22,4005	9,7247
9	19,5161	6,8322	20,638	8,5188
10	18,8197	6,1903	20,4902	8,7771
11	18,2616	7,145	20,9738	10,4015
12	18,1456	6,7192	20,6281	8,4295
13	17,4546	6,8305	19,7312	8,6095
14	18,2366	7,9356	21,0256	9,2601
15	18,8766	5,1381	21,7394	9,293
16	18,7479	5,8479	22,38	9,4287
17	17,4722	7,5293	20,8082	9,3475
18	16,3661	6,3028	19,6055	9,8492
19	17,5277	5,0242	20,4671	8,8375
20	16,4551	4,6167	21,4703	10,6565

Da análise dos resultados, percebe-se que eles variaram muito pouco com a variação do número de neurônios. Para a rede E2R1, o erro da pressão sistólica variou de 20,53 a 21,92 e da pressão diastólica variou de 19,82 a 21,72. Para a rede E2R2, a variação do erro foi de 19,26 a 20,76 para a sistólica e de 21,16 a 22,51 para a diastólica. Para a rede E2R3, o erro da pressão sistólica variou de 17,41 a 19,13 e da diastólica de 19,60 a 21,99. Percebe-se que, diferentemente do experimento 3, a rede em que as variáveis de entrada são compostas pelas variáveis metabólicas mais as variáveis não metabólicas (E2R3) apresentou resultados superiores aos resultados das duas outras redes. Portanto, a rede conseguiu atribuir um peso menor às variáveis metabólicas, as quais, por algum motivo, estavam atrapalhando o desempenho. Nota-se que, apesar da melhora, os erros ainda são grandes demais e que, portanto, são considerados insatisfatórios.

5.1.5 Experimento 5

Como no experimento 1, este experimento tem como finalidade tentar prever a hipertensão arterial crônica utilizando as técnicas de pré-processamento preenchimento de dados faltantes, normalização *z-score* e seleção de variáveis. A técnica de divisão dos dados utilizada é o método de validação cruzada chamado *k-fold*. Este experimento se diferencia do

experimento 1 pois, na seleção de variáveis, o segundo algoritmo descrito na seção 4.1.5 (algoritmo 3) foi utilizado para a configuração H1 da seção 4.1.4. A tabela 11 da seção 4.1.5 contém os resultados da seleção, a tabela 35 desta seção apresenta os detalhes das redes utilizadas neste experimento e, na tabela 36, estão as variáveis de entrada e variáveis de saída da rede utilizada neste experimento.

Tabela 35: Detalhes das redes utilizadas no experimento 5.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward com treinamento backpropagation	Validação cruzada k-fold	1 tgh	De 1 a 20	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 3) e Z-score	“hipertenso”

Tabela 36: Variáveis de entrada e saída utilizadas nas redes do experimento 5.

Código	Variáveis de entrada	Variáveis de saída
E5R1	“x13”, “x14”, “x15”, “x16” e “x25”	“hipertenso”

Os resultados do experimento podem ser visualizados na tabela 37, a qual é referente à rede E5R1.

Tabela 37: Resultados para a rede E5R1 do experimento 5.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	65,25	19,7001
2	55,5	12,6595
3	66,5	21,5272
4	58,75	19,4581
5	68,75	18,4872
6	62,75	18,1713
7	65,5	19,5946
8	54	22,1003
9	62,5	20,995
10	58,5	14,1514
11	62	20,1573
12	60	15,2177
13	64,25	20,3441
14	62,25	20,55
15	67,25	19,0204
16	67,25	19,0204
17	65,75	18,0113
18	59,75	13,8103
19	56,75	23,1855
20	59,25	23,2987

A partir da análise da tabela 37 (rede E5R1), é possível verificar que a taxa de acerto média variou de 55,5% a 68,75% e a quantidade de neurônios do melhor resultado foi 5. Se comparado com o experimento 1 (rede E1R1), é possível verificar que os resultados foram superiores. Isso se deve ao fato de o algoritmo 3 da seção 4.1.5 ser menos rígido do que o algoritmo 2, ou seja, ele permite selecionar um maior número de variáveis. Do resultado do experimento, percebe-se que as variáveis não selecionadas pelo algoritmo 2 e selecionadas pelo algoritmo 3 (“x16” e “x25”) foram importantes para a melhora do desempenho da rede neural.

5.1.6 Experimento 6

A fim de melhorar os resultados do experimento 2, este utiliza as mesmas configurações que ele. Difere do experimento 2 pois o método de divisão de dados é a validação cruzada *k-fold* e o método de seleção de variáveis é o algoritmo 3 da seção 4.1.5. As variáveis selecionadas com o algoritmo 3 foram apresentadas nas tabelas 11 e 12 da seção 4.1.5. A tabela 38 apresenta os detalhes das redes utilizadas neste experimento e a tabela 39 apresenta as variáveis de entrada e variáveis de saída para cada uma das três redes utilizadas neste.

Tabela 38: Detalhes das redes utilizadas no experimento 6.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward com treinamento backpropagation	Validação cruzada k-fold	2 ln	De 1 a 20	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 3) e Z-score	"PA sistólica" e "PA diastólica"

Tabela 39: Variáveis de entrada e saída utilizadas nas redes do experimento 6.

Código	Variáveis de entrada	Variáveis de saída
E6R1	"x124", "x131", "x125", "x132", "x133", "x123", "x88", "x134", "x97", "x89", "x126", "x98", "x90", "x99", "x52", "x127", "x38", "x34", "x100", "x55", "x114", "x103", "x88", "x104", "x91", "x102", "x101", "x93" e "x94"	"PA sistólica" e "PA diastólica"
E6R2	"IMC" e "peso pós-parto"	"PA sistólica" e "PA diastólica"
E6R3	"x124", "x131", "x125", "x132", "x133", "x123", "x88", "x134", "x97", "x89", "x126", "x98", "x90", "x99", "x52", "x127", "x38", "x34", "x100", "x55", "x114", "x103", "x88", "x104", "x103", "x88", "x104", "x91", "x102", "x101", "x93", "x94", "IMC" e "peso pós-parto"	"PA sistólica" e "PA diastólica"

As tabelas 40, 41 e 42 contêm os resultados das redes E6R1, E6R2 e E6R3, respectivamente. Elas são semelhantes aos outros experimentos que tiveram como tentativa prever as pressões dos indivíduos, ou seja, apresentam as médias e os desvios dos erros para as pressões sistólica e diastólica para cada quantidade de neurônios na camada intermediária.

Tabela 40: Resultados para a rede E6R1 do experimento 6.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	21,0302	6,1297	21,3509	8,8486
2	21,3769	6,029	19,8201	8,5131
3	20,5843	6,1063	20,64	8,6743
4	20,9417	6,1341	20,1491	8,3822
5	21,4558	6,9273	20,616	8,5693
6	21,0836	5,9182	20,261	8,6891
7	21,6075	6,4011	21,5355	8,4085
8	21,2668	5,854	21,7264	9,0497
9	20,5743	6,8605	21,0943	8,9711
10	21,6406	5,9939	20,0182	9,4503
11	21,6381	6,823	20,1527	8,4666
12	20,5386	5,824	20,1122	8,5839
13	21,5976	6,244	21,0809	7,9609
14	21,1765	6,3713	20,2871	8,114
15	21,6396	6,8419	21,2831	8,6751
16	20,926	6,3206	21,5083	9,7586
17	21,1877	6,1254	20,1858	9,0224
18	21,715	6,4883	21,7143	9,2403
19	21,9244	6,6931	20,5584	8,7523
20	21,066	6,1813	21,1229	9,4425

Tabela 41: Resultados para a rede E6R2 do experimento 6.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	20,7656	8,1832	22,1718	9,3371
2	19,4785	8,5446	21,9599	9,0298
3	19,2624	7,1787	22,1128	9,149
4	19,8929	7,6044	21,6958	9,4796
5	20,1267	7,4	21,931	9,4781
6	19,8775	7,6535	21,5848	9,5042
7	19,8225	8,5095	22,3048	9,5468
8	20,5702	8,0403	21,5442	9,1999
9	20,09	7,2219	22,5168	9,1261
10	19,8175	7,9171	22,4909	8,8152
11	20,1996	8,0873	22,1835	9,043
12	19,9464	7,8406	21,8941	9,6268
13	19,734	7,3385	22,0448	9,4711
14	20,0642	6,9478	21,7643	9,7516
15	19,4435	7,7294	21,4722	9,8314
16	19,7481	7,5908	21,808	9,4931
17	20,4404	6,5121	21,5218	9,9338
18	19,5049	7,3476	21,1643	9,4801
19	19,9276	7,789	21,8287	9,4202
20	19,5799	8,034	22,0675	9,1999

Tabela 42: Resultados para a rede E6R3 do experimento 6.

Neurônios	Média do erro (Sistólica)	Desvio do erro (Sistólica)	Média do erro (Diastólica)	Desvio do erro (Diastólica)
1	18,5548	7,4462	20,9108	8,9832
2	18,0055	5,8756	19,6876	10,3672
3	17,8822	5,937	20,9215	9,1685
4	17,8035	5,9404	21,9989	8,8783
5	17,4196	5,9315	20,5561	9,8964
6	17,7025	6,7972	20,8461	9,6112
7	19,1357	7,672	21,2606	8,9366
8	18,8187	6,015	22,4005	9,7247
9	19,5161	6,8322	20,638	8,5188
10	18,8197	6,1903	20,4902	8,7771
11	18,2616	7,145	20,9738	10,4015
12	18,1456	6,7192	20,6281	8,4295
13	17,4546	6,8305	19,7312	8,6095
14	18,2366	7,9356	21,0256	9,2601
15	18,8766	5,1381	21,7394	9,293
16	18,7479	5,8479	22,38	9,4287
17	17,4722	7,5293	20,8082	9,3475
18	16,3661	6,3028	19,6055	9,8492
19	17,5277	5,0242	20,4671	8,8375
20	16,4551	4,6167	21,4703	10,6565

A partir da análise da tabela 40, é possível ver que o resultado da rede E6R1 varia muito pouco e o melhor resultado é de um erro de 20,5 para a pressão sistólica (3 neurônios) e de 19,8 para a pressão diastólica (2 neurônios). Os resultados da rede E6R2 também variam pouco e o melhor erro foi de 19,2 para a pressão sistólica (3 neurônios) e de 21,5 para a pressão diastólica (8 neurônios). Já com relação aos resultados da rede E6R3, eles variaram mais e o melhor resultado para a pressão sistólica foi de um erro de 16,36 para a pressão sistólica (18 neurônios) e de 19,6 para a pressão diastólica (18 neurônios). Os resultados foram melhores do que os resultados do experimento 2. Como no experimento 5, a utilização do algoritmo 3 da seção 4.1.5 para seleção de variáveis foi responsável pelo aumento do desempenho das redes. A melhora do desempenho está relacionada não só à redução do erro, mas como também à redução do desvio padrão, conforme se pode constatar através da comparação das tabelas deste experimento com as tabelas do experimento 2. Quando os resultados deste experimento são comparados com os resultados do experimento 4, verifica-se que os resultados são semelhantes.

5.1.7 Experimento 7

Este experimento e o próximo têm como objetivo retirar os possíveis intrusos do conjunto de dados a fim de verificar se tal retirada contribui para um aumento no desempenho das redes. A detecção de intrusos foi especificada anteriormente (seção 4.1.7) e, conforme visto, seis elementos do conjunto de dados foram selecionados como possíveis intrusos. Como, inicialmente, a quantidade de elementos no conjunto era de vinte e quatro (um número já considerado pequeno), a retirada destes seis possíveis intrusos reduziu o tamanho do conjunto para dezoito elementos, o que tornou esse conjunto ainda menor. As redes utilizadas neste experimento são idênticas às redes utilizadas no experimento 5 e se diferenciam apenas no que diz respeito à detecção de intrusos, a qual não era realizada e agora é. Como no experimento 5, o intuito deste experimento é prever se um indivíduo sofre ou não de hipertensão arterial crônica. A tabela 43 apresenta detalhes das redes utilizadas neste experimento e a tabela 44 especifica as variáveis de entrada e saída de cada uma das três diferentes redes utilizadas neste experimento.

Tabela 43: Detalhes das redes utilizadas no experimento 7.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward com treinamento backpropagation	Validação cruzada k-fold	1 tgh	De 1 a 10	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 3) e Z-score	“hipertenso”

Tabela 44: Variáveis de entrada e saída utilizadas nas redes do experimento 6.

Código	Variáveis de entrada	Variáveis de saída
E7R1	“x13”, “x14”, “x15”, “x16” e “x25”	“hipertenso”
E7R2	“IMC”, “peso pós-parto”, “idade”, “parto” e “sangramento á sondagem”.	“hipertenso”
E7R3	“x13”, “x14”, “x15”, “x16”, “x25”, “IMC”, “peso pós-parto”, “idade”, “parto” e “sangramento á sondagem”.	“hipertenso”

As tabelas 45, 46 e 47 apresentam as médias e desvios da taxa de acerto da rede com diferentes números de neurônios na camada intermediária.

Tabela 45: Resultados para a rede E7R1 do experimento 7.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	67,917	29,528
2	67,917	27,741
3	66,25	24,85
4	62,917	24,85
5	63,75	23,301
6	68,333	24,572
7	66,667	23,1
8	65,833	27,293
9	60,833	25,521
10	67,917	25,258

Tabela 46: Resultados para a rede E7R2 do experimento 7.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	88,333	13,355
2	87,083	15,643
3	87,5	19,022
4	88,333	13,355
5	87,083	13,376
6	88,75	15,12
7	90	12,566
8	88,333	13,355
9	79,583	25,287
10	88,75	15,12

Tabela 47: Resultados para a rede E7R3 do experimento 7.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	87,083	14,925
2	88,333	18,417
3	86,25	19,734
4	86,667	16,086
5	88,333	14,907
6	87,083	14,925
7	84,167	15,029
8	87,083	15,643
9	85,833	18,357
10	80,417	18,59

Ao analisar as tabelas, é possível verificar que, para a rede E7R1, a taxa média de acerto variou entre 60,8% e 68,3%, sendo que a taxa de 68,3% foi obtida quando se usou 6 neurônios na camada intermediária. Em relação à rede E7R2, a variação foi entre 79,5% e

90%, sendo que a taxa de 90% foi obtida quando se usou 7 neurônios na camada escondida. Já a partir da análise da tabela 47, é possível verificar que a taxa média de acerto variou entre 80,4% e 88,3% e que a melhor taxa de acerto foi obtida com 5 neurônios na camada oculta. Analisando os resultados da rede E7R1, a qual utiliza apenas as variáveis metabólicas como variáveis de entrada, é possível verificar que o resultado foi ruim, uma vez que o desvio foi alto e que, portanto, a retirada dos intrusos não contribuiu para o aumento do desempenho desta rede. Em relação aos resultados das duas outras redes (E7R2 e E7R3), não foi possível comparar com o experimento 5, já que nele apenas uma rede foi utilizada (E5R1). Mas, é possível compará-los com os resultados dos experimentos 3 (tabela 30, rede E3R2 e tabela 31, rede E3R3) e 1 (tabela 22, rede E1R2 e tabela 23, rede E1R3). Em relação a estes experimentos, a rede E7R3 apresentou melhores resultados, o que indica uma pequena melhora no desempenho devido à remoção dos intrusos. Já a rede E7R2 também apresentou desempenho inferior às redes do experimento 1 (E1R2) e 3 (E3R2). Como os resultados da rede E7R1 e E7R2 não melhoraram em relação aos resultados das redes dos experimentos anteriores (E5R1, E3R1, E1R1, E3R2 e E1R2) e o resultado da rede E7R3 não apresentou melhora significativa em relação às redes E3R3 e E1R3, conclui-se que a remoção dos seis intrusos não contribuiu para a melhora do desempenho das redes e que, portanto, indica que alguns ou todos os elementos selecionados como intrusos não o são.

5.1.8 Experimento 8

Semelhantemente ao experimento 7, este utiliza a técnica de eliminação de intrusos para tentar melhorar os resultados dos experimentos 2, 4 e 6, os quais tentam prever as pressões sistólica e diastólica de indivíduos. Os seis possíveis intrusos encontrados na seção 4.1.7 foram removidos e o experimento 6 foi executado novamente, dando origem a este experimento. Como no experimento 6, o objetivo deste experimento é prever as pressões utilizando três configurações de variáveis de entrada distintas, conforme especificado na tabela 49. Detalhes das redes utilizadas neste experimento podem ser visualizados na tabela 48.

Tabela 48: Detalhes das redes utilizadas no experimento 8.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward com treinamento backpropagation	Validação cruzada k-fold	2 ln	De 1 a 20	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 3) e Z-score	“PA sistólica” e “PA diastólica”

Tabela 49: Variáveis de entrada e saída utilizadas nas redes do experimento 8.

Código	Variáveis de entrada	Variáveis de saída
E8R1	“x124”, “x131”, “x125”, “x132”, “x133”, “x123”, “x88”, “x134”, “x97”, “x89”, “x126”, “x98”, “x90”, “x99”, “x52”, “x127”, “x38”, “x34”, “x100”, “x55”, “x114”, “x103”, “x88”, “x104”, “x91”, “x102”, “x101”, “x93” e “x94”	“PA sistólica” e “PA diastólica”
E8R2	“IMC” e “peso pós-parto”	“PA sistólica” e “PA diastólica”
E8R3	“x124”, “x131”, “x125”, “x132”, “x133”, “x123”, “x88”, “x134”, “x97”, “x89”, “x126”, “x98”, “x90”, “x99”, “x52”, “x127”, “x38”, “x34”, “x100”, “x55”, “x114”, “x103”, “x88”, “x104”, “x91”, “x102”, “x101”, “x93”, “x94”, “IMC” e “peso pós-parto”	“PA sistólica” e “PA diastólica”

As tabelas 50, 51 e 52 são referentes aos resultados das redes E8R1, E8R2 e E8R3, respectivamente. Nelas, como nos experimentos anteriores que envolviam as pressões sistólica e diastólica, são apresentados as médias e desvios dos erros obtidos na tentativa de prevê-las para diferentes números de neurônios na camada intermediária.

Tabela 50: Resultados para a rede E8R1 do experimento 8.

Neurônios	Média Erro (Sistólica)	Desvio Erro (Sistólica)	Média Erro (Diastólica)	Desvio Erro (Diastólica)
1	20,018	8,114	18,964	7,0048
2	22,581	8,723	18,775	5,7608
3	25,878	12,01	23,163	8,8568
4	22,077	7,8102	23,473	11,955
5	22,535	11,628	26,682	12,598
6	24,237	9,3114	22,813	9,2774
7	24,081	11,542	27,221	12,239
8	27,854	14,41	26,639	13,13
9	26,729	7,815	28,552	15,67
10	28,946	15,103	28,171	12,537

Tabela 51: Resultados para a rede E8R2 do experimento 8.

Neurônios	Média Erro (Sistólica)	Desvio Erro (Diastólica)	Média Erro (Diastólica)	Desvio Erro (Diastólica)
1	20,15	7,2467	18,588	8,418
2	19,356	8,0501	18,744	8,0297
3	21,037	8,1499	19,321	7,0599
4	20,293	7,9914	20,725	8,7438
5	21,568	9,1725	22,494	10,588
6	20,8	8,1212	21,106	10,515
7	23,37	9,4231	23,707	11,583
8	23,556	9,613	22,848	11,454
9	20,115	5,9253	23,657	11,17
10	25,727	12,81	24,277	10,544

Tabela 52: Resultados para a rede E8R3 do experimento 8.

Neurônios	Média Erro (Sistólica)	Desvio Erro (Diastólica)	Média Erro (Diastólica)	Desvio Erro (Diastólica)
1	17,135	5,1839	24,304	15,089
2	19,816	7,9255	20,832	8,7448
3	20,72	7,4968	19,527	8,9731
4	21,211	11,427	20,381	3,9555
5	23,066	9,4851	22,246	9,1055
6	20,399	7,1389	25,013	10,539
7	21,182	8,2738	28,06	12,431
8	23,895	10,681	25,36	11,304
9	27,834	15,876	28,62	11,695
10	24,474	8,7217	29,976	16,873

A partir da análise das tabelas, é possível verificar que, para a rede E8R1, a média dos erros variou de 20 (1 neurônio) a 28,9 (10 neurônios) para a pressão sistólica e variou de 18,7 (2 neurônios) a 27,2 (7 neurônios) para a pressão diastólica. Para a rede E8R2, os erros médios variaram entre 19,3 (2 neurônios) e 25,7 (10 neurônios) para a pressão sistólica e entre 18,5 (1 neurônio) e 24,2 (10 neurônios) para a pressão diastólica. Os erros médios da rede E8R3 variaram de 17,1 (1 neurônio) a 27,8 (9 neurônios) para a pressão sistólica e de 19,5 (3 neurônios) a 29,9 (10 neurônios) para a pressão diastólica. A partir da análise destes resultados, é possível verificar que, para uma quantidade de neurônios baixa (1 ou 2 neurônios), os resultados deste experimento foram ligeiramente superiores ao resultado do experimento 6. Já para quantidade de neurônios alta, os resultados foram piores. Se for levada em conta apenas a média e o desvio do melhor caso, o que é completamente normal, já que o

número de neurônios na camada escondida é escolhido de acordo com o melhor resultado, pode-se afirmar que este experimento apresentou resultados ligeiramente superiores (vantagem pouco expressiva) ao experimento 6 e que, portanto, a eliminação dos intrusos foi responsável por uma pequena melhora no desempenho das redes. Como no experimento anterior o desempenho não melhorou com a remoção dos intrusos e neste experimento constatou-se que houve uma melhora insignificante, conclui-se que nem todos os elementos selecionados como intrusos realmente o eram.

5.1.9 Experimento 9

Este experimento é idêntico ao experimento 5, mas a quantidade de dados utilizados para tentar realizar a previsão da hipertensão arterial crônica é maior. No experimento 5, vinte e quatro indivíduos eram utilizados para treinar, testar e validar a rede neural. Neste, trinta indivíduos são utilizados. Esta diferença se deve pois, ao decorrer deste trabalho, mais dados foram obtidos. Como a quantidade de dados era pequena, a inclusão dos dados referentes a estes novos seis indivíduos aumentou a amostra em vinte e cinco por cento, o que torna o estudo em questão mais confiável. Optou-se por repetir o experimento 5 e não o experimento 7, pois nota-se que os resultados obtidos no experimento 7 foram inferiores aos resultados obtidos no experimento 5 (resultado da tentativa de remover os intrusos). As configurações das redes utilizadas neste experimento podem ser visualizadas na tabela 53 e as variáveis de entrada e saída foram especificadas na tabela 54 para cada uma destas redes.

Tabela 53: Detalhes das redes utilizadas no experimento 9.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward com treinamento backpropagation	Validação cruzada k-fold	2 ln	De 1 a 20	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 3) e Z-score	“PA sistólica” e “PA diastólica”

Tabela 54: Variáveis de entrada e saída utilizadas nas redes do experimento 8.

Código	Variáveis de entrada	Variáveis de saída
E9R1	“x13”, “x14”, “x15”, “x16” e “x25”	“hipertenso”
E9R2	“IMC”, “peso pós-parto”, “idade”, “parto” e “sangramento á sondagem”.	“hipertenso”
E9R3	“x13”, “x14”, “x15”, “x16”, “x25”, “IMC”, “peso pós-parto”, “idade”, “parto” e “sangramento á sondagem”.	“hipertenso”

As tabelas 55, 56 e 57 apresentam os resultados obtidos neste experimento para as redes E9R1, E9R2 e E9R3, respectivamente. Nelas, são mostradas as taxas médias de acerto e desvios obtidos para diferentes quantidades de neurônios.

Tabela 55: Resultados para a rede E9R1 do experimento 9.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	56.667	21.709
2	61.333	22.242
3	60.667	14.368
4	66	17.538
5	64.667	17.953
6	62	16.897
7	64	20.611
8	62	18.458
9	58	19.191
10	60.667	18.557

Tabela 56: Resultados para a rede E9R2 do experimento 9.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	73.333	16.884
2	75.333	20.126
3	80.667	13.374
4	78	17.695
5	77.333	15.522
6	76	17.734
7	78	18.458
8	76.667	13.979
9	76	15.222
10	74	15.888

Tabela 57: Resultados para a rede E9R3 do experimento 9.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	71.333	20.126
2	72	20.069
3	76.667	17.486
4	70.667	18.742
5	70.667	20.833
6	76	17.734
7	72	19.369
8	72.667	16.174
9	75.333	21.453
10	72	17.889

A partir da análise das tabelas, é possível verificar que as taxas médias de acerto da rede E9R1 variam de 56,6% (1 neurônio) a 66% (4 neurônios). Para a rede E9R2, as taxas variam de 73,3% (1 neurônio) a 80,6 (3 neurônios) e, para a rede E9R3, elas variam de 70,66% a 76%. Da análise destes resultados, é possível ver que a inclusão de novos dados não alterou muito o resultado quando utilizados apenas os metabólitos salivares. Enquanto o melhor resultado da rede E5R1 foi uma taxa de acerto média de 68,75% com um desvio de 18,48, o melhor resultado da rede equivalente deste experimento (E9R1) foi uma taxa média de 66% com um desvio de 17,538. Já se compararmos as redes E9R2 e E9R3 com as redes E7R2 e E7R3, nota-se que ocorre uma piora expressiva, apesar dos resultados das redes que envolvem as variáveis não metabólicas continuarem sendo superiores aos das redes que envolvem as variáveis metabólicas.

5.1.10 Experimento 10

Este experimento é idêntico ao experimento 6, mas, como no experimento 9, a quantidade de dados utilizados para realizar a previsão das pressões arteriais é maior. Optou-se por repetir o experimento 6 e não o experimento 8, pois nota-se que os resultados obtidos no experimento 8 foram inferiores aos resultados obtidos no experimento 6 (piora causada pela tentativa de remover os intrusos). As configurações das redes utilizadas neste experimento podem ser visualizadas na tabela 58 e as variáveis de entrada e saída foram especificadas na tabela 59 para cada uma destas redes.

Tabela 58: Detalhes das redes utilizadas no experimento 10.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento	Variáveis de saída
Feedforward com treinamento backpropagation	Validação cruzada k-fold	2 ln	De 1 a 20	Preenchimento de dados faltantes, Seleção de variáveis (algoritmo 3) e Z-score	“PA sistólica” e “PA diastólica”

Tabela 59: Variáveis de entrada e saída utilizadas nas redes do experimento 10.

Código	Variáveis de entrada	Variáveis de saída
E8R1	“x124”, “x131”, “x125”, “x132”, “x133”, “x123”, “x88”, “x134”, “x97”, “x89”, “x126”, “x98”, “x90”, “x99”, “x52”, “x127”, “x38”, “x34”, “x100”, “x55”, “x114”, “x103”, “x88”, “x104”, “x91”, “x102”, “x101”, “x93” e “x94”	“PA sistólica” e “PA diastólica”
E8R2	“IMC” e “peso pós-parto”	“PA sistólica” e “PA diastólica”
E8R3	“x124”, “x131”, “x125”, “x132”, “x133”, “x123”, “x88”, “x134”, “x97”, “x89”, “x126”, “x98”, “x90”, “x99”, “x52”, “x127”, “x38”, “x34”, “x100”, “x55”, “x114”, “x103”, “x88”, “x104”, “x91”, “x102”, “x101”, “x93”, “x94”, “IMC” e “peso pós-parto”	“PA sistólica” e “PA diastólica”

As tabelas 60, 61 e 62 apresentam os resultados obtidos para as redes E10R1, E10R2 e E10R3, respectivamente. Elas mostram a média e desvio dos erros para as pressões sistólica e diastólica para diferentes números de neurônios na camada intermediária.

Tabela 60: Resultados para a rede E10R1 do experimento 10.

Neurônios	Média Erro (Sistólica)	Desvio Erro (Sistólica)	Média Erro (Diastólica)	Desvio Erro (Diastólica)
1	19.3353	5.1156	21.6842	8.4783
2	19.7547	6.151	22.5722	7.8426
3	18.7446	4.8029	22.5884	7.0797
4	18.8611	5.3825	25.0568	8.1107
5	23.0706	8.6227	26.2399	10.3825
6	22.3151	8.0135	25.3603	9.4845
7	21.6635	6.0608	27.2052	9.0957
8	23.4766	7.6241	26.1461	9.1975
9	20.8464	6.5267	25.8966	11.1491
10	23.2636	6.9852	30.357	12.0479

Tabela 61: Resultados para a rede E10R2 do experimento 10.

Neurônios	Média Erro (Sistólica)	Desvio Erro (Sistólica)	Média Erro (Diastólica)	Desvio Erro (Diastólica)
1	20.5344	5.7507	23.0844	8.2248
2	20.5716	6.0212	21.7962	8.4613
3	20.7827	7.4071	22.2212	7.9494
4	20.4184	5.6548	23.1236	8.7484
5	19.9819	6.4276	24.1784	9.6239
6	20.3552	6.3843	26.8595	9.0058
7	21.3739	7.4296	25.8908	8.9958
8	20.7636	8.4991	26.5921	8.9941
9	23.5677	7.826	27.8789	13.9561
10	22.6144	7.9969	25.5347	8.3869

Tabela 62: Resultados para a rede E10R3 do experimento 10.

Neurônios	Média Erro (Sistólica)	Desvio Erro (Sistólica)	Média Erro (Diastólica)	Desvio Erro (Diastólica)
1	18.7985	6.0391	20.8014	9.7144
2	18.6588	5.7739	21.9879	7.8546
3	19.1215	4.9761	23.6983	9.5272
4	18.0714	5.8759	23.7157	9.097
5	20.0276	5.4136	24.6192	8.5222
6	19.5329	5.4583	23.8613	8.3676
7	20.1429	5.9358	28.1498	11.8557
8	18.0081	4.0332	22.6577	8.2371
9	21.5455	6.1307	26.9823	8.8531
10	23.8393	7.9631	32.8054	15.7111

Analisando os resultados, é possível verificar que os resultados da rede E10R1 variam de 18,74 (3 neurônios) a 23,47 (8 neurônios) para a pressão sistólica e de 21,68 (1 neurônio) a 30,35 (10 neurônios) para a pressão diastólica. Os erros da rede E10R2 variam de 19,98 (5 neurônios) a 23,56 (9 neurônios) para a pressão sistólica e de 21,79 (2 neurônios) a 27,87 (9 neurônios) para a pressão diastólica. Já os erros da rede E10R3 variam de 18 (8 neurônios) a 23,83 (10 neurônios) para a pressão sistólica e de 20,8 (1 neurônio) a 32,08 (10 neurônios) para a pressão diastólica. A partir da análise dos resultados, é possível verificar que o único resultado que melhorou ao aumentar a quantidade de dados foi o referente ao uso apenas dos metabólitos salivares para prever a pressão sistólica. O melhor resultado para a pressão sistólica foi de um erro médio de 20,5 e desvio de 5,8 no experimento 6 contra um erro médio de 18,7 e desvio de 4,8 neste experimento. Todos os outros resultados deste experimento foram inferiores quando comparados aos resultados do experimento 6.

5.1.11 Conclusão

A partir da análise dos experimentos com o emprego de RNAs, verificou-se que, quando utilizados apenas os metabólitos salivares como variáveis de entrada, uma taxa média de 68,75% de acerto e um desvio de 18,48% foi o melhor resultado obtido para a tentativa de prever a hipertensão arterial crônica (experimento 5). Quando utilizados apenas o IMC e o peso pós-parto para a tentativa de previsão, uma taxa de 92% de acerto foi a melhor taxa encontrada (experimentos 1 e 3). Já com relação às tentativas de previsão utilizando como variáveis de entrada os metabólitos salivares mais o IMC e o peso pós-parto, o melhor resultado obtido foi de 88,33% com desvio de 14,9% (experimento 7). Na tentativa de prever a pressão sistólica, o melhor resultado foi um erro médio de 17,13 e um desvio de 5,18

(experimento 8). Já na tentativa de prever a pressão diastólica, o melhor resultado foi um erro médio de 18,77 e com desvio de 5,76 (experimento 8). Na maioria dos casos, a utilização do algoritmo 3 (seleção de variáveis) foi responsável pela melhora dos resultados. Além disso, a eliminação dos possíveis intrusos causou uma piora dos resultados de alguns experimentos e os que foram beneficiados tiveram melhoras pouco significativas, indicando que as entradas detectadas ou parte delas não eram intrusos. Com relação à inclusão dos novos dados, os resultados dos experimentos que fizeram uso deles foram muito semelhantes aos resultados dos outros experimentos.

5.2 Doença Periodontal

Esta seção descreve todos os experimentos e apresenta os resultados encontrados nas tentativas de classificar indivíduos com relação à doença periodontal.

5.2.1 Experimento 1

Este experimento utiliza as variáveis metabólicas selecionadas através da técnica de seleção de características descritas na seção 4.1.5 deste trabalho para classificar um indivíduo quanto à doença periodontal. Tais variáveis são referentes à configuração DP1 e estão especificadas na tabela 12 da seção 4.1.5 e também na tabela 54 desta seção. Neste experimento, o preenchimento de dados faltantes, a normalização *z-score* e a seleção de características foram realizados (especificados na seção 4.1). Além disso, a detecção de intrusos não foi realizada (por não ter melhorado o desempenho das redes) e a validação cruzada *k-fold* foi utilizada a fim de dividir o conjunto de dados em treinamento, validação e teste, escolher a melhor rede e computar os resultados (média e desvio) para diferentes números de neurônios na camada escondida da rede. Como a variável “diagnóstico periodontal” foi representada utilizando a notação maximamente esparsa, foram utilizados três neurônios do tipo tangente hiperbólica na camada da saída, as quais indicam se o indivíduo é saudável ou sofre de gengivite ou sofre de periodontite. O número de neurônios na camada escondida varia de 1 a 10 dado que a melhor forma de descobrir tal número é analisando os resultados e escolhendo aquele que fornece os melhores resultados. Quanto à distribuição dos dados, em um total de 21 indivíduos, 15 são diagnosticados com periodontite, 3 com gengivite e 3 saudáveis. Foram utilizados apenas 21 indivíduos e não 24 como nos experimentos da seção 5.1, pois 3 não estavam rotulados. Os detalhes da rede

utilizada são resumidos na tabela 63 e as variáveis de entrada e saída são apresentadas na tabela 64.

Tabela 63: Detalhes das redes utilizadas no experimento 1.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento
Feedforward com treinamento backpropagation	Validação cruzada k-fold	3 tgh	De 1 a 10	Preenchimento de dados faltantes, seleção de variáveis (algoritmo 3) e normalização Z-score

Tabela 64: Variáveis de entrada e saída utilizadas na rede do experimento 1.

Variáveis de entrada	Variáveis de saída
“x128”, “x54”, “x71”, “x130”, “x129”, “x72”, “x86”, “x80”, “x53” e “x127”	“doença periodontal” (notação maximamente esparsa)

A tabela 65 apresenta os resultados obtidos na execução do experimento. Tal tabela apresenta a média da porcentagem de acerto e o desvio para diferentes números de neurônios na camada escondida (1 a 10).

Tabela 65: Resultados do experimento 1.

Neuronios	Média Porcentagem	Desvio Porcentagem
1	61	33,3674
2	81,75	13,8101
4	85,125	10,0024
5	86,25	6,9948
6	86,875	5,4399
7	87,75	3,9648
8	86,25	8,6127
9	87,5	5,33
10	87,125	5,4631

Através da análise dos resultados, verifica-se que a taxa de acerto varia de 61% (1 neurônio) a 87,75% (7 neurônios). Considerando o melhor caso (7 neurônios), um resultado médio de 87,75% de acerto com um desvio de 3,96 foi obtido, o qual é um resultado

considerado bom, uma vez que a média é alta e o desvio baixo. Comparando com os resultados da rede E7R1 do experimento 7 da seção 5.1.1.7 (tentativa de prever a hipertensão arterial crônica), verifica-se que os resultados deste experimento foram bem melhores, o que indica que a doença periodontal é mais facilmente diagnosticada a partir dos metabólitos salivares do que a hipertensão arterial crônica.

5.2.2 Experimento 2

Este experimento tem como finalidade a utilização dos índices de diagnóstico da doença periodontal para classificar os indivíduos quanto à doença periodontal (configuração DP3 da seção 4.1.5). É esperado que a taxa de acerto da rede seja alta, visto que a classificação correta dos indivíduos quanto à doença periodontal é feita utilizando estes índices. Como as próprias variáveis utilizadas para a classificação real dos indivíduos são utilizadas como variáveis de entrada, então a rede utilizada neste experimento apenas mapeará as regras utilizadas na classificação real dos indivíduos. Como no experimento 1, o preenchimento de dados faltantes, a normalização *z-score*, a seleção de características e a validação cruzada *k-fold* foram utilizados. A variação do número de neurônios na camada escondida e a distribuição dos dados também são idênticas ao experimento 1. Os detalhes da rede utilizada são resumidos na tabela 66 e as variáveis de entrada e saída são apresentadas na tabela 67.

Tabela 66: Detalhes das redes utilizadas no experimento 2.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento
Feedforward com treinamento backpropagation	Validação cruzada k-fold	3 tgh	De 1 a 10	Preenchimento de dados faltantes, seleção de variáveis (algoritmo 3) e normalização Z-score

Tabela 67: Variáveis de entrada e saída utilizadas na rede do experimento 2.

Variáveis de entrada	Variáveis de saída
“Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”.	“doença periodontal” (notação maximamente esparsa)

A tabela 68 apresenta a porcentagem média de acerto e o desvio obtido na execução do experimento 2 para diferentes números de neurônios na camada escondida.

Tabela 68: Resultados do experimento 2.

Neuronios	Média Porcentagem	Desvio Porcentagem
1	68,5	25,3436
2	86,375	13,0673
3	88,25	13,626
6	95	7,1067
7	94,625	7,591
8	94,625	7,1631
9	95,625	6,4976
10	96,125	7,2594

Como esperado, o resultado foi muito bom, uma vez que a melhor taxa média foi de 96,12% com 10 neurônios. Comparado com o experimento 1, onde apenas os metabólitos salivares foram utilizados, vê-se que existe uma diferença significativa entre os resultados. Esta diferença poderia ter sido menor se mais dados fossem utilizados para treinar a rede do experimento 1 e se os números de elementos de cada classificação fossem parecidos, uma vez que o pequeno número de indivíduos que sofrem de gengivite e que são saudáveis provavelmente não possibilita que a rede generalize muito bem os dados.

5.2.3 Experimento 3

Este experimento é parecido com os experimentos 1 e 2, diferindo apenas nas variáveis de entrada utilizadas na rede neural. Tais variáveis são formadas pela união das variáveis metabólicas do experimento 1 com os índices de diagnóstico da doença periodontal utilizados no experimento 2. O objetivo deste experimento é verificar se a união dos índices de diagnóstico com os metabólitos salivares implica em resultados superiores ou pelo menos parecidos com os resultados do experimento 2. A tabela 69 resume os detalhes da rede utilizada neste experimento e a tabela 70 apresenta as variáveis de entrada e saída utilizadas.

Tabela 69: Detalhes das redes utilizadas no experimento 3.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento
Feedforward com treinamento backpropagation	Validação cruzada k-fold	3 tgh	De 1 a 10	Preenchimento de dados faltantes, seleção de variáveis (algoritmo 3) e normalização Z-score

Tabela 70: Variáveis de entrada e saída utilizadas na rede do experimento 3.

Variáveis de entrada	Variáveis de saída
x128, x54, x71, x130, x129, x72, x86, x80, x53, x127, "Profundidade de bolsa", "Nível de perda de inserção", "Placa", "PBS rasa", "PBS média", "NCI raso" e "NCI médio".	"doença periodontal" (notação maximamente esparsa)

Os resultados deste experimento podem ser visualizados na tabela 71, a qual apresenta a porcentagem média de acerto da rede e o desvio para diferentes quantidades de neurônios (1 a 10).

Tabela 71: Resultados do experimento 3.

Neuronios	Média Porcentagem	Desvio Porcentagem
1	62	32,0747
2	80,5	16,6021
3	83,75	14,8115
6	88,5	6,5713
7	87,125	8,6045
8	87,75	7,5336
9	85,625	7,6159
10	86,5	7,0353

A partir da análise da tabela, é possível verificar que a taxa média de acerto varia entre 62% (1 neurônio) e 88,5% (6 neurônios). Os resultados foram próximos aos resultados do experimento 1 e não chegaram próximo aos do experimento 2. Já era esperado que o resultado não fosse tão bom quanto o resultado do experimento 2, uma vez que nele a rede realiza apenas o mapeamento das regras utilizadas para a classificação real dos indivíduos quanto à doença periodontal. Entretanto, era esperado que o resultado deste fosse um pouco superior ao do experimento 1, uma vez que os índices de diagnóstico ajudariam a aumentar o desempenho da rede, mas não foi isto o que aconteceu.

5.2.4 Experimento 4

Este experimento é a repetição dos experimentos 1, 2 e 3, mas com uma quantidade maior de dados. Como mencionado nos experimentos 9 e 10 da hipertensão arterial crônica, ao decorrer deste trabalho mais dados foram adquiridos, permitindo assim repetir os experimentos com uma quantidade maior de dados. Anteriormente, foram utilizados dados referentes a 21 indivíduos, sendo que 15 foram diagnosticados com periodontite, 3 com gengivite e 3 com saúde periodontal. Como mencionado anteriormente, na literatura, são apresentadas diversas formas de realizar o diagnóstico da doença periodontal a partir do exame periodontal. Nos dados utilizados anteriormente não foi documentada a forma como o diagnóstico foi realizado, ou seja, as regras utilizadas sobre os dados coletados durante o exame periodontal (variáveis de “a” a “l” da seção 3.1.1.2) para classificar os indivíduos com saúde periodontal, gengivite e periodontite. Por essa razão, quando mais dados foram obtidos, um novo diagnóstico teve que ser realizado, resultando em alguns diagnósticos diferentes do anterior. Após esse novo diagnóstico, dos 30 indivíduos, 12 foram diagnosticados com saúde periodontal, 11 com gengivite e 7 com periodontite. As configurações compartilhadas entre as redes utilizadas neste experimento podem ser vistas na tabela 72 e as entradas e saídas utilizadas em cada rede deste experimento podem ser visualizadas na tabela 73.

Tabela 72: Detalhes das redes utilizadas no experimento 4.

Rede	Divisão	Neurônios saída	Neurônios escondidos	Pré-processamento
Feedforward com treinamento backpropagation	Validação cruzada k-fold	3 tgh	De 1 a 10	Preenchimento de dados faltantes, seleção de variáveis (algoritmo 3) e normalização Z-score

Tabela 73: Variáveis de entrada e saída utilizadas na rede do experimento 4.

Código	Variáveis de entrada	Variáveis de saída
E4R1	“x128”, “x54”, “x71”, “x130”, “x129”, “x72”, “x86”, “x80”, “x53” e “x127”	“doença periodontal” (notação maximamente esparsa)
E4R2	“Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”.	“doença periodontal” (notação maximamente esparsa)
E4R3	x128, x54, x71, x130, x129, x72, x86, x80, x53, x127, “Profundidade de bolsa”, “Nível de perda de inserção”, “Placa”, “PBS rasa”, “PBS média”, “NCI raso” e “NCI médio”.	“doença periodontal” (notação maximamente esparsa)

As tabelas 74, 75 e 76 são referentes aos resultados das redes E4R1, E4R2 e E4R3, respectivamente. Elas apresentam as taxas médias de acerto e os desvios para diferentes números de neurônios na camada intermediária.

Tabela 74: Resultados da rede E4R1 do experimento 4.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	51.6	22.2347
2	69	21.9043
3	78.8	18.3831
4	80.4	17.2867
5	79.6	15.303
6	81.4	14
7	82.2	14.1835
8	83.8	12.2911
9	81	13.4465
10	82.6	14.115

Tabela 75: Resultados da rede E4R2 do experimento 4.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	44.8	17.0845
2	57.8	17.4992
3	60	15.57
4	62.8	15.0474
5	61.2	16.7742
6	62.4	16.1508
7	64.2	14.297
8	64.2	14.8515
9	64.6	11.6706
10	64.2	14.0115

Tabela 76: Resultados da rede E4R3 do experimento 4.

Neurônios	Média Porcentagem	Desvio Porcentagem
1	56.8	18.7961
2	79.2	20.4831
3	95	11.5032
4	95.6	10.8544
5	97.8	6.9019
6	99.2	4.8576
7	99.4	3.4289
8	99.6	2.8141
9	100	0
10	100	0

O resultado obtido com a rede E4R1 (tabela 74) foi muito inferior ao resultado obtido no experimento 1, ao qual possui as mesmas configurações e variáveis de entrada e saída. O melhor resultado do experimento 1 foi um acerto médio de 87,75% com um desvio de 3,96% e o melhor resultado da rede E4R1 deste experimento foi um acerto médio de 83,8% com um desvio de 12,29%, evidenciando assim a piora dos resultados após a inclusão dos novos dados. Comparando a tabela 75, a qual é referente ao resultado da rede E4R2, com a tabela 68 (experimento 2), a qual utiliza as mesmas variáveis de entrada e saída e possui a mesma configuração da rede E4R2, é possível ver que a rede E4R2 deste experimento apresentou resultados inferiores, uma vez que esta apresentou melhor resultado referente a um acerto médio de 64,6% e desvio de 11,67% e o melhor resultado do experimento 2 (tabela 68) foi um acerto médio de 96,12% e desvio de 7,26%. Esta piora expressiva nos resultados (tanto em relação à rede E4R1, quanto em relação à rede E4R2) pode ser justificada pela regra utilizada para o diagnóstico da doença periodontal, uma vez que o diagnóstico feito inicialmente não foi documentado e que, quando novos dados foram obtidos, foi necessário realizá-lo novamente (baseado no exame periodontal). A regra utilizada após a inclusão dos novos dados utiliza o PBS, NCI e o SS obtidos sobre cada sítio dos dentes para realizar o diagnóstico. Como as variáveis PBS, NCI e SS (descritas na seção 3.1.2.2) utilizadas nas redes do experimento 1, 3 e 4 são médias obtidas sobre os sítios, parte da informação é perdida, fazendo com que a rede não consiga mapear adequadamente a regra de diagnóstico. Diferente das outras redes, a E4R3 apresentou resultados muito bons e muito superiores à rede do experimento 3 (tabela 71). Enquanto a rede E4R3 apresentou no melhor caso um acerto médio de 100% com um desvio de 0%, a rede do experimento 3 apresentou média de 88,5% com um desvio de 6,57%, evidenciando uma grande melhora após a inclusão dos novos dados. Essa melhora no resultado da rede E4R3 pode ser justificada pela união dos metabólitos salivares com o exame periodontal.

5.2.5 Conclusão

A partir da análise dos experimentos com RNAs é possível verificar que, antes de adicionar os novos dados, a taxa média de acerto tanto da rede que utilizou como variáveis de entrada apenas os metabólitos salivares (E1R1) como da rede que utilizou os dados do exame periodontal (E2R1) e da rede que utilizou os metabólitos mais os dados do exame periodontal (E3R1) foi alta (acima de 85% de acerto). Quando novos dados foram adicionados, um novo diagnóstico da doença periodontal teve que ser realizado, já que não haviam sido

documentados os critérios utilizados para o diagnóstico feito anteriormente. Com isso, houve uma piora nos resultados referentes às redes E1R1 (no experimento 4 chamada de E4R1) e E2R1 (no experimento 4 chamada de E4R2) e uma grande melhora na rede E3R1 (no experimento 4 chamada de E4R3). A piora dos resultados da rede E2R1 pode ser explicada pela mudança nos critérios utilizados para realizar o diagnóstico periodontal, uma vez que tal critério utilizou informações coletadas em cima de cada sítio dos dentes analisados (PBS, NCI e SS) e a rede utilizou apenas as médias destas informações. Já a piora da rede E1R1 pode ser justificada pela maior complexidade do critério de diagnóstico utilizado após a inclusão dos novos dados. A união dos metabólitos salivares com os dados do exame periodontal na rede E4R3 permitiu com que a rede encontrasse nos metabólitos salivares as informações faltantes nas médias das variáveis utilizadas para realizar o diagnóstico periodontal. Com isso, houve uma grande melhora nos resultados chegando aos 100% de acerto. É importante ressaltar que as redes mais importantes são a rede do experimento 1 (E1R1) e a primeira rede do experimento 4 (E4R1), as quais possuem configurações equivalentes. Isso se deve ao fato de as outras redes utilizarem as variáveis obtidas no exame periodontal nas variáveis de entrada e que, se as temos disponíveis, seria muito mais lógico realizar o diagnóstico utilizando os critérios de diagnóstico do que uma rede neural.

Capítulo 6. Conclusão e Trabalhos Futuros

Analisando os resultados obtidos nos experimentos com RNAs descritos no Capítulo 5, é possível constatar que as redes neurais artificiais quando usadas para analisar os dados dos metabólitos salivares foram capazes de fornecer porcentagem de acerto alta para tentativa de prever a doença periodontal, uma vez que o experimento que fez uso apenas destes metabólitos para tentar prever a doença (experimento 2, Seção 5.2.2) forneceu bons resultados (média de acerto de cerca de 88% e desvio de 4%). Estes resultados podem ser considerados bons, pois foi possível verificar que existe uma grande correlação entre os metabólitos salivares e a doença periodontal, o que permite a criação de outros experimentos utilizando estes metabólitos a fim de obter resultados melhores e, se estes forem muito altos, poderão vir a permitir o desenvolvimento de aparelhos que façam o diagnóstico da doença a partir da saliva. Apesar de o resultado ter sido alto, este não foi satisfatório para o diagnóstico, uma vez que 88% de acerto ainda permite que uma grande quantidade de indivíduos seja classificada incorretamente quanto à doença periodontal.

Em relação à hipertensão arterial crônica, os resultados das tentativas de previsão foram inferiores aos resultados da doença periodontal (68,75% de acerto quando apenas considerados os metabólitos salivares e 18,48% de desvio). Com aplicação de técnicas de pré-processamento extras (seleção de características) foi possível aumentar a taxa de acerto da tentativa de previsão da hipertensão arterial crônica, quando utilizados apenas os metabólitos salivares, de 60,75% com desvio de 18,86% (experimento 1) para 68,75% com desvio de 18,48% (experimento 5). Apesar do aumento, os resultados foram insatisfatórios quando considerados apenas os metabólitos salivares, uma vez que a média do melhor experimento foi apenas razoável e o desvio foi alto. Quanto à utilização apenas do IMC e do peso para a tentativa de previsão, o resultado do experimento 1 foi alto (92% de acerto médio com 10,05% de desvio). Este resultado evidencia o que vem sendo apresentado em diversos trabalhos relacionados à hipertensão arterial, que é a relação existente entre o peso e a pressão arterial ([HUANG *et al.*, 1998], [STAMLER R *et al.*, 1978], [HAVLIK *et al.*, 1983], [BARKER, 2006]). Quando os metabólitos salivares foram utilizados em conjunto com o IMC e o peso, o melhor resultado (tabela 47, 5 neurônios) foi de 88,33% com desvio de 14,9%. Como conclusão dos resultados das tentativas de previsão da hipertensão arterial crônica através do emprego de RNAs, percebe-se que os experimentos realizados com os metabólitos parecem

revelar certa associação entre tais metabólitos e a hipertensão arterial crônica, mas é necessário realizar outros experimentos a fim de confirmá-la.

Em relação à tentativa de prever as pressões sistólica e diastólica, os resultados obtidos foram muito ruins, uma vez que o melhor resultado obtido para a pressão sistólica foi no experimento 8, quando utilizados os metabólitos salivares, o IMC e o peso pós-parto (erro médio de 17,13 e desvio 5,18) e para pressão diastólica foi também no experimento 8 com um erro médio de 18,77 e com um desvio de 5,76 quando utilizados apenas os metabólitos salivares. Como conclusão, não foi possível realizar a previsão das pressões arteriais devido aos erros encontrados serem grandes.

Como trabalhos futuros, poder-se-ia repetir os experimentos realizados neste trabalho com uma quantidade maior de dados, uma vez que a quantidade utilizada neste foi muito pequena e tendo um maior número de dados, certamente as RNAs generalizariam melhor os dados e, conseqüentemente, o desempenho seria melhor. Além disso, outras técnicas de classificação podem ser utilizadas a fim de tentar obter resultados superiores aos aqui obtidos. Também, poder-se-ia verificar os intrusos de outra forma a fim de ver se os seis selecionados eram realmente intrusos ou se apenas parte deles eram.

Bibliografia

[ALMEIDA, 2013] Almeida, Priscila. METABOLOMA SALIVAR E SANGUÍNEO DE CRIANÇAS COM INSUFICIÊNCIA RENAL CRÔNICA POR MEIO DA RESSONÂNCIA MAGNÉTICA NUCLEAR. Dissertação de mestrado. Rio de Janeiro: Universidade Federal do Rio de Janeiro. 2012.

[BARKER, 2006] BARKER, D.J.P., 2006, "Birth Weight and Hypertension". In: *Hypertension*. v. 48, pp. 357–358.

[BAUMBACH & HEISTAD, 1989] BAUMBACH, G.L., HEISTAD, D.D., 1989, "Remodeling of cerebral arterioles in chronic hypertension". In: *Hypertension*. v. 13, pp. 968–972.

[BAUTISTA *et al.*, 2004] BAUTISTA, L.E., VERA, L.M., ARENAS, I.A., *et al.*, 2004, "Independent association between inflammatory markers (C-reactive protein, interleukin-6, and TNF- α) and essential hypertension". In: *Journal of Human Hypertension*. v. 19, pp. 149–154.

[BECK *et al.*, 1996] BECK, J., GARCIA, R., HEISS, G., *et al.*, 1996, "Periodontal Disease and Cardiovascular Disease". In: *Journal of Periodontology*. v. 67, pp. 1123–1137.

[BISHOP, 1995] BISHOP, C.M., 1995, *Neural networks for pattern recognition*. S.I., Oxford university press. Acessado em: 7 Março 2014.

[CHOMYSZYN-GAJEWSKA, 2010] CHOMYSZYN-GAJEWSKA, M., 2010, "[Evaluation of chosen salivary periodontal disease markers]". In: *Przegląd lekarski*. v. 67, pp. 213–216.

[DALTON & DESHMANE, 1991] DALTON, J., DESHMANE, A., 1991, "Artificial neural networks". In: *IEEE Potentials*. v. 10, pp. 33–36.

[FAUSETT, 2006] FAUSETT, 2006, *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. S.I., Pearson Education.

[FIDALGO *et al.*, 2013] FIDALGO, T.K., FREITAS-FERNANDES, L.B., ANGELI, R., *et al.*, 2013, "Salivary metabolite signatures of children with and without dental caries lesions". In: *Metabolomics*. v. 9, pp. 657–666.

- [FIDALGO et al., 2014] FIDALGO, T.K., FREITAS-FERNANDES, L.B., ALMEIDA, F.C., et al., 2014, "Longitudinal evaluation of salivary profile from children with dental caries before and after treatment". In: *Metabolomics*. pp. 1–11.
- [FYFE, 2005] FYFE, C., 2005. "Artificial Neural Networks". In: GABRYS, Bogdan, LEIVISKÄ, Kauko & STRACKELJAN, Jens (eds.), *Do Smart Adaptive Systems Exist?* S.I.: Springer Berlin Heidelberg. Studies in Fuzziness and Soft Computing, 173. pp. 57–79.
- [GIANNOBILE et al., 2009] GIANNOBILE, W.V., BEIKLER, T., KINNEY, J.S., et al., 2009, "Saliva as a diagnostic tool for periodontal disease: current state and future directions". In: *Periodontology 2000*. v. 50, pp. 52–64.
- [GOTOH et al, 2005] GOTOH, S., WATANABE, Y., FUJIBAYASHI, T., 2005, "Validity of stimulated whole saliva collection as a sialometric evaluation for diagnosing Sjogren's syndrome. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, St. Louis, v. 99, n. 3, p. 299-302.
- [GRAU et al., 2004] GRAU, A.J., BECHER, H., ZIEGLER, C.M., et al., 2004, "Periodontal Disease as a Risk Factor for Ischemic Stroke". In: *Stroke*. v. 35, pp. 496–501.
- [GRIGOR'EV et al, 2002] GRIGOR'EV, I. V., ULANOVA, E. A., LADIK, B. B., 2002, "Characteristics of salivary protein spectrum in patients with depressive disorders (in search for biological markers of depression)". *Klin Lab Diagn*, Moskva, v. 1, n. 1, p. 15-18.
- [HART et al., 1980] HART, M.N., HEISTAD, D.D., BRODY, M.J., 1980, "Effect of chronic hypertension and sympathetic denervation on wall/lumen ratio of cerebral vessels.". In: *Hypertension*. v. 2, pp. 419–423.
- [HAVLIK et al., 1983] HAVLIK, R.J., HUBERT, H.B., FABSITZ, R.R., et al., 1983, "Weight and Hypertension". In: *Annals of Internal Medicine*. v. 98, pp. 855–859.
- [HAYKIN, 1994] HAYKIN, S., 1994, *Neural Networks: A Comprehensive Foundation*. 1st. Upper Saddle River, NJ, USA, Prentice Hall PTR.
- [HOPFIELD, 1988] HOPFIELD, J.J., 1988, "Artificial neural networks". In: *IEEE Circuits and Devices Magazine*. v. 4, pp. 3–10.
- [HU & HWANG, 2001] HU, Y.H., HWANG, J.-N., 2001, *Handbook of Neural Network Signal Processing*. S.I., CRC Press.

- [HUANG *et al.*, 1998] HUANG, Z., WILLETT, W.C., MANSON, J.E., et al., 1998, "Body Weight, Weight Change, and Risk for Hypertension in Women". In: *Annals of Internal Medicine*. v. 128, pp. 81–88.
- [HUMPHREY *et al.*, 2008] HUMPHREY, L.L., FU, R., BUCKLEY, D.I., et al., 2008, "Periodontal Disease and Coronary Heart Disease Incidence: A Systematic Review and Meta-analysis". In: *Journal of General Internal Medicine*. v. 23, pp. 2079–2086.
- [JAIN & ZONGKER, 1997] JAIN, A., ZONGKER, D., 1997, "Feature selection: evaluation, application, and small sample performance". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. v. 19, pp. 153–158.
- [JIMENEZ *et al.*, 2013] JIMENEZ, M., GIOVANNUCCI, E., KRALL KAYE, E., et al., 2013, "Predicted vitamin D status and incidence of tooth loss and periodontitis". In: *Public Health Nutrition*. v. FirstView, pp. 1–9.
- [JOU *et al.*, 2010] JOU, Y.-J., LIN, C.-D., LAI, C.-H., et al., 2010, "Proteomic identification of salivary transferrin as a biomarker for early detection of oral cancer". In: *Analytica chimica acta*. v. 681, pp. 41–48.
- [KAUFMAN & LAMSTER, 2000] KAUFMAN, E., LAMSTER, I.B., 2000, "Analysis of saliva for periodontal diagnosis". In: *Journal of Clinical Periodontology*. v. 27, pp. 453–465.
- [KIRSCHBAUM *et al.*, 1992] KIRSCHBAUM, C., READ, G. F., HELLHAMMER, D. H., 1992, "Assessment of hormones and drugs in saliva in biobehavioral research. Kirkland: Hogefe and Huber Publishers".
- [KOSHY *et al.*, 2005] KOSHY, G., KAWASHIMA, Y., KIJI, M., et al., 2005, "Effects of single-visit full-mouth ultrasonic debridement versus quadrant-wise ultrasonic debridement". In: *Journal of clinical periodontology*. v. 32, pp. 734–743.
- [LÖE, 1993] LÖE, H., 1993, "Periodontal Disease: The sixth complication of diabetes mellitus". In: *Diabetes Care*. v. 16, pp. 329–334.
- [MATTA *et al.*, 2010] MATTA, A., RALHAN, R., DESOUZA, L.V., et al., 2010, "Mass spectrometry-based clinical proteomics: Head-and-neck cancer biomarkers and drug-targets discovery". In: *Mass spectrometry reviews*. v. 29, pp. 945–961.
- [MBULAITEYE *at al.*, 2006] MBULAITEYE, S. M., WALTERS, M., ENGELS, E. A., BAKAKI, P. M., NDUGWA, C. M., OWOR, A. M. et al, 2006, "High levels of Epstein-Barr

virus DNA in saliva and peripheral blood from Ugandan mother-child pairs". *J Infect Dis*, Chicago, v. 193, n. 3, p. 422-426.

[MOURA *et al.*, 2008] MOURA, S.A.B. DE, MEDEIROS, A.M.C. DE, COSTA, F.R.H. DA, et al., 2008, "Valor Diagnóstico da Saliva em Doenças Orais e Sistêmicas: Uma Revisão de Literatura". In: *Pesquisa Brasileira em Odontopediatria e Clínica Integrada*. v. 7, pp. 187–194.

[MOURA *et al.*, 2004] MOURA, S. A. B., 2004, "Análises clínica, sialométrica e sialoquímica em indivíduos portadores da síndrome do ardor bucal". 2004. 136f. Tese (Doutorado em Estomatologia), Universidade Federal da Paraíba/Universidade Federal da Bahia, João Pessoa.

[MOYNIHAN & PETERSEN, 2004] MOYNIHAN, P., PETERSEN, P.E., 2004, "Diet, nutrition and the prevention of dental diseases". In: *Public Health Nutrition*. v. 7, pp. 201–226.

[OZMERIC, 2004] OZMERIC, N., 2004, "Advances in periodontal disease markers". In: *Clinica Chimica Acta*. v. 343, pp. 1–16.

[PEREIRA, 2012] Pereira, Luciana. Análise dos metabólitos salivares de mães com hipertensão arterial crônica e seus lactantes por meio de ressonância magnética nuclear. Dissertação de mestrado. Rio de Janeiro: Universidade Federal do Rio de Janeiro. 2012.

[PETERSEN *et al.*, 2005] PETERSEN, P.E., BOURGEOIS, D., OGAWA, H., et al., 2005, "The global burden of oral diseases and risks to oral health". In: *Bulletin of the World Health Organization*. v. 83, pp. 661–669.

[PFAFFE *et al.*, 2011] PFAFFE, T., COOPER-WHITE, J., BEYERLEIN, P., et al., 2011, "Diagnostic potential of saliva: current state and future applications". In: *Clinical chemistry*. v. 57, pp. 675–687.

[PIHLSTROM *et al.*, 2005] PIHLSTROM, B.L., MICHALOWICZ, B.S., JOHNSON, N.W., 2005, "Periodontal diseases". In: *The Lancet*. v. 366, pp. 1809–1820.

[POMMERANZENBAUM, 2014] POMMERANZENBAUM, I.R., 2014, Redes Neurais Artificiais na predição das principais séries do índice ibovespa e suas aplicações em sistemas automatizados de negociação. Dissertação de M.Sc., Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

[RAJARAMAN & ULLMAN, 2012] RAJARAMAN, A., ULLMAN, J.D., 2012, *Mining of massive datasets*. S.l., Cambridge University Press. Acessado em: 6 Abril 2014.

[RESHMA *et al.*, 2013] RESHMA, A., ARUNACHALAM, R., PILLAI, J., *et al.*, 2013, "Chromogranin A: Novel biomarker between periodontal disease and psychosocial stress". In: *Journal of Indian Society of Periodontology*. v. 17, pp. 214.

[RINGNÉR, 2008] RINGNÉR, M., 2008, "What is principal component analysis?". In: *Nature biotechnology*. v. 26, pp. 303–304.

[RMN INFOPÉDIA] ressonância magnética nuclear. In Infopédia [Em linha]. Disponível em: <[http://www.infopedia.pt/\\$ressonancia-magnetica-nuclear](http://www.infopedia.pt/$ressonancia-magnetica-nuclear)>. Acesso em: 18 fev. 2014.

[SASS *et al.*, 2002] SASS, Nelson; SANTOS, Jorge F. K.; ATALLAH, Alvaro N.; CAMANO, Luiz. Hipertensão Arterial Crônica Leve e Moderada na Gravidez: Práticas Terapêuticas Baseadas em Evidências. 2002.

[SCANNAPIECO *et al.*, 2003] SCANNAPIECO, F.A., BUSH, R.B., PAJU, S., 2003, "Associations Between Periodontal Disease and Risk for Atherosclerosis, Cardiovascular Disease, and Stroke. A Systematic Review". In: *Annals of Periodontology*. v. 8, pp. 38–53.

[SCULLEY & LANGLEY-EVANS, 2003] SCULLEY, D.V., LANGLEY-EVANS, S.C., 2003, "Periodontal disease is associated with lower antioxidant capacity in whole saliva and evidence of increased protein oxidation". In: *Clinical science (London, England: 1979)*. v. 105, pp. 167–172.

[SCULLEY & LANGLEY-EVANS, 2002] SCULLEY, D.V., LANGLEY-EVANS, S.C., 2002, "Salivary antioxidants and periodontal disease status". In: *Proceedings of the Nutrition Society*. v. 61, pp. 137–143.

[SIBAI, 2012] SIBAI, B.M., 2012. Disponível em: <http://journals.lww.com/greenjournal/Fulltext/2002/08000/Chronic_Hypertension_in_Pregnancy.29.aspx>. Acessado em: 4 Fevereiro 2014.

[SIBAI *et al.*, 1998] SIBAI, B.M., LINDHEIMER, M., HAUTH, J., *et al.*, 1998, "Risk Factors for Preeclampsia, Abruptio Placentae, and Adverse Neonatal Outcomes among Women with Chronic Hypertension". In: *New England Journal of Medicine*. v. 339, pp. 667–671.

- [SILVA-BOGHOSSIAN et al, 2011] SILVA-BOGHOSSIAN, C. M., SOUTO, R. M., LUIZ, R. R., COLOMBO, A. P., 2011, "Association of red complex, *A. actinomycetemcomitans* and non-oral bacteria with periodontal diseases". *Arch Oral Biol.*, v. 56, n. 9, p. 899-906.
- [SON et al., 2012] SON, A., PERA, C., UEDA, P., et al., 2012, "Clinical effects of supragingival plaque control on uncontrolled type 2 diabetes mellitus subjects with chronic periodontitis". In: *Brazilian Journal of Oral Sciences.* v. 11, pp. 47–51.
- [STAMLER R et al., 1978] STAMLER R, STAMLER J, RIEDLINGER WF, et al., 1978, "Weight and blood pressure: Findings in hypertension screening of 1 million americans". In: *JAMA.* v. 240, pp. 1607–1610.
- [STRECKFUS et al, 2000] STRECKFUS, C., BIGLER, L., DELLINGER, T., DAI, X., KINGMAN, A., TRIGPEN, J. T., 2000, "The presence of soluble cerbB-2 concentrations in the saliva and serum among women with breast carcinoma: a preliminary study. *ClinCancer Res, Philadelphia,* v. 6, n. 6, p. 2363-2370.
- [SUGIMOTO et al., 2010] SUGIMOTO, M., WONG, D.T., HIRAYAMA, A., et al., 2010, "Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles". In: *Metabolomics.* v. 6, pp. 78–95.
- [TOTAN et al., 2006] TOTAN, A., GREABU, M., TOTAN, C., et al., 2006, "Salivary aspartate aminotransferase, alanine aminotransferase and alkaline phosphatase: possible markers in periodontal diseases?". In: *Clinical Chemistry And Laboratory Medicine: CCLM / FESCC.* v. 44, pp. 612–615.
- [VLACHOS et al., 2002] VLACHOS, M., DOMENICONI, C., GUNOPULOS, D., et al., 2002. "Non-linear dimensionality reduction techniques for classification and visualization". In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.* S.l.: ACM. 2002. pp. 645–651.
- [WILTON et al., 1989] WILTON, J.M.A., CURTIS, M.A., GILLETT, L.R., et al., 1989, "Detection of high-risk groups and Individuals for periodontal diseases: laboratory markers from analysis of saliva". In: *Journal of Clinical Periodontology.* v. 16, pp. 475–483.
- [YEH et al., 2011] YEH, C.-J., CHUANG, W.-Y., CHAO, Y.-K., et al., 2011, "High expression of phosphorylated 4E-binding protein 1 is an adverse prognostic factor in esophageal squamous cell carcinoma". In: *Virchows Archiv.* v. 458, pp. 171–178.