

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

CARMINA DESSANA NASCIMENTO PORTO

CLASSIFICAÇÃO DE BIFURCAÇÕES EM IMAGENS DE TOMOGRAFIA
DE COERÊNCIA ÓPTICA INTRAVASCULAR UTILIZANDO REDES
NEURAIS E MÁQUINAS DE VETORES DE SUPORTE

MANAUS
2017

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

CARMINA DESSANA NASCIMENTO PORTO

CLASSIFICAÇÃO DE BIFURCAÇÕES EM IMAGENS DE TOMOGRAFIA
DE COERÊNCIA ÓPTICA INTRAVASCULAR UTILIZANDO REDES
NEURAIS E MÁQUINAS DE VETORES DE SUPORTE

Dissertação apresentada ao Curso de
Mestrado em Engenharia Elétrica, área de
concentração de Controle e Automação de
Sistemas do Programa de Pós-Graduação em
Engenharia Elétrica da Universidade Federal
do Amazonas.

Orientadora: Profa. Dra. Marly Guimarães Fernandes Costa

Co-orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho

MANAUS
2017

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

P853c Porto, Carmina Dessana Nascimento
Classificação de Bifurcações em Imagens de Tomografia de
Coerência Óptica Intravascular Utilizando Redes Neurais e
Máquinas de Vetores de Suporte / Carmina Dessana Nascimento
Porto. 2017
136 f.: il. color; 31 cm.

Orientadora: Marly Guimarães Fernandes Costa
Coorientador: Cícero Ferreira Fernandes Costa Filho
Dissertação (Mestrado em Engenharia Elétrica) - Universidade
Federal do Amazonas.

1. doenças cardiovasculares. 2. tomografia por coerência óptica.
3. redes neurais. 4. máquinas de vetores de suporte. 5. bifurcações.
I. Costa, Marly Guimarães Fernandes II. Universidade Federal do
Amazonas III. Título

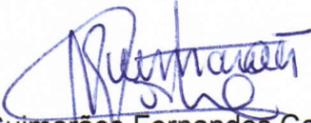
CARMINA DESSANA NASCIMENTO PORTO

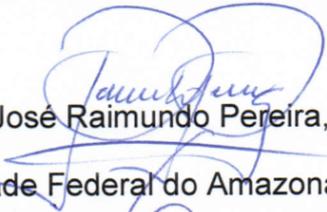
CLASSIFICAÇÃO DE BIFURCAÇÕES EM IMAGENS DE TOMOGRAFIA DE
COERÊNCIA ÓPTICA INTRAVASCULAR UTILIZANDO REDES NEURAIS E
MÁQUINAS DE VETORES DE SUPORTE

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 17 de abril de 2017.

BANCA EXAMINADORA


Profª. Dra. Marly Guimarães Fernandes Costa, Presidente
Universidade Federal do Amazonas- UFAM


Prof. Dr. José Raimundo Pereira, Membro
Universidade Federal do Amazonas- UFAM


Prof. Dr. Marco Antônio Gutierrez, Membro
Universidade de São Paulo- INCOR/HC/USP

AGRADECIMENTOS

Agradeço aos meus orientadores, Profa. Dr^a. Marly Guimarães Fernandes Costa e Prof. Dr. Cícero Ferreira Fernandes Costa Filho, pela oportunidade de participar do seu grupo de pesquisa, por acreditarem em mim desde o primeiro dia, e principalmente, por toda a confiança em mim depositada.

Aos meus pais, sem vocês nenhuma das minhas conquistas seria possível. Nunca conseguirei expressar em palavras a minha gratidão.

Ao meu marido Fábio Porto, por todo seu apoio, pelo incentivo e pelo companheirismo. Obrigada por acreditar em mim em todos os momentos.

Ao meu grande amigo Thales Araújo da Silva, sem seu incentivo eu talvez nem teria começado o mestrado.

À Universidade Federal do Amazonas e em especial ao Centro de Tecnologia Eletrônica e da Informação – CETELI - pela concessão de toda infraestrutura para realização deste trabalho.

À CAPES pela concessão da bolsa de estudos durante os dois anos de trabalho

RESUMO

Estudos em tomografia de coerência óptica intravascular (IV-OCT) tem demonstrado a importância das regiões de bifurcação coronária na análise de imagens médicas intravasculares, uma vez que esta região é mais propensa ao acúmulo de placas, o que pode levar a doença coronariana. Um exame IV-OCT comum adquire centenas de imagens, portanto, uma ferramenta automatizada para classificar as imagens como pertencentes ou não a região de bifurcações pode ser um passo importante para acelerar a análise de imagens IV-OCT e auxiliar métodos automatizados para a quantificação de placas ateroscleróticas. Neste trabalho, avaliamos o desempenho de dois classificadores, SVM e Redes Neurais, na tarefa de classificação de identificação de bifurcações em imagens IV-OCT. O estudo incluiu imagens IV-OCT de 9 pacientes. Para melhorar o desempenho da classificação, treinamos e testamos o SVM com diferentes parâmetros por meio de uma pesquisa de grade e diferentes critérios de parada foram aplicados ao classificador de Rede Neural: erro quadrático médio, parada precoce e regularização. Foram testados diferentes conjuntos de características, utilizando técnicas de seleção de características: PCA, LDA e seleção de características escalares com correlação. Treinamento e teste foram realizados em conjuntos com um máximo de 1460 imagens. Quantificamos nossos resultados em termos de taxa de falsos positivos, taxas de verdadeiro positivo, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e área sob curva ROC. As redes neurais obtiveram a melhor precisão de classificação, 98,83%, superando os resultados encontrados na literatura. Nossos métodos parecem oferecer uma classificação robusta e confiável automatizada de imagens IV-OCT que podem ajudar médicos indicando potenciais imagens a serem analisadas. Métodos para melhorar a generalização das redes neurais aumentaram o desempenho da classificação.

Palavras-chave: doenças cardiovasculares, tomografia por coerência óptica, redes neurais, máquinas de vetores de suporte, bifurcação, classificação, coronária.

ABSTRACT

Studies in intravascular optical coherence tomography (IV-OCT) have demonstrated the importance of coronary bifurcation regions in intravascular medical imaging analysis, as plaques are more likely to accumulate in this region leading to coronary disease. A typical IV-OCT pullback acquires hundreds of frames, thus developing an automated tool to classify the OCT frames as bifurcation or non-bifurcation can be an important step to speed up OCT pullbacks analysis and assist automated methods for atherosclerotic plaque quantification. In this work, we evaluate the performance of two state-of-the-art classifiers, SVM and Neural Networks in the bifurcation classification task. The study included IV-OCT frames from 9 patients. In order to improve classification performance, we trained and tested the SVM with different parameters by means of a grid search and different stop criteria were applied to the Neural Network classifier: mean square error, early stop and regularization. Different sets of features were tested, using feature selection techniques: PCA, LDA and scalar feature selection with correlation. Training and test were performed in sets with a maximum of 1460 OCT frames. We quantified our results in terms of false positive rate, true positive rate, accuracy, specificity, precision, false alarm, f-measure and area under ROC curve. Neural networks obtained the best classification accuracy, 98.83%, overcoming the results found in literature. Our methods appear to offer a robust and reliable automated classification of OCT frames that might assist physicians indicating potential frames to analyze. Methods for improving neural networks generalization have increased the classification performance.

Keywords: Cardiovascular Diseases, Intravascular Optical Coherence Tomography, Neural Networks, Support Vector Machines, Bifurcation, Detection, Coronary.

LISTA DE ILUSTRAÇÕES

Figura 1 - Anatomia do coração humano. Fonte: Adaptado de (Siu, 2012)	18
Figura 2 - Ilustração do estreitamento da artéria por placa aterosclerótica Fonte: Adaptado de (Siu, 2012).....	19
Figura 3 – Aplicação do stent coronário. Fonte: (Webmd, 2016)	20
Figura 4 – Exemplo de imagem IVOCT: A imagem superior corresponde a um corte transversal da artéria coronariana. A imagem inferior é a vista longitudinal da coronária.	22
Figura 5 – Imagens do mesmo corte em OCT e em IVUS. Fonte: (Macedo, 2014)	23
Figura 6 – Diagrama esquemático de um sistema de auxílio ao diagnóstico para interpretação de imagens médicas. Adaptado de (Giger <i>et al.</i> , 2008).	30
Figura 7 - Fluxograma de etapas de classificação de imagens utilizando técnicas de processamento digital de imagens. Fonte: Adaptado de (Gonzalez e Woods, 2009).....	32
Figura 8 – Modelo geométrico de uma bifurcação. Fonte: (Lesage <i>et al.</i> , 2009)	35
Figura 9 –Exemplos de imagem de uma angiografia (a) original (b) segmentada com destaque nos seguintes pontos vaso principal (P), vaso principal distal (DMV) e ramo lateral (SB). Fonte: adaptado de (Cardialysys, 2014).....	36
Figura 10 – Exemplos de imagens IVOCT (a) sem bifurcação e (b) com bifurcação. Foram selecionadas imagens com o lúmen segmentado para melhor identificação das regiões de interesse.....	36
Figura 11 - Fluxograma de etapas típicas de classificação de bifurcações utilizando técnicas de reconhecimento de padrões	40
Figura 12 – Vista axial de imagens IVUS utilizadas no trabalho de Albertiet al (2011) (a) sem bifurcação (b) com bifurcação. Fonte: (Alberti <i>et al.</i> , 2011)	43
Figura 13 – Diagrama em blocos do funcionamento básico do treinamento de uma rede neural. Fonte: Adaptado de (The Mathworks).	53
Figura 14 – Modelo de um neurônio artificial. Fonte: Adaptado de (Haykin, 1994).	53
Figura 15 – Ilustração do fluxo de processamento do algoritmo backpropagation. Fonte: (Matsunaga, 2012)	55
Figura 16 – Exemplo de <i>Overfitting</i> . Fonte: (Hagan <i>et al.</i> , 1996).....	58
Figura 17 – Curvas hipotéticas de erros de treinamento e de validação. Fonte: Adaptado de (Hagan <i>et al.</i> , 1996)	60

Figura 18- Ilustração dos conceitos de hiperplano de separação e margem. Fonte: adaptado de (Theodoridis e Koutroumbas, 2008).....	62
Figura 19 – Exemplos de imagens (a) conjunto 1: com bifurcação normal (b) conjunto 2: bifurcação com sangue residual (c) conjunto 3: bifurcação com placa complexa.	73
Figura 20 - Ilustração do processo de balanceamento dos conjuntos compilados por Macedo <i>et al.</i> (2015).....	76
Figura 21 – Ilustração da técnica K-fold <i>Cross Validation</i> , com k=10.	77
Figura 22- Quatro fases da classificação de dados e os diferentes conjuntos utilizados como entrada para os classificadores implementados.	78
Figura 23 – Diagrama em blocos genérico utilizado em cada fase do trabalho. il representa o número de neurônios na camada de entrada (igual ao número de características, podendo valer 104 e o intervalo de 5 a 20), h_1 e h_2 representam os números de neurônios nas camadas escondidas (podendo valer 5-5, 8-8 e 10-10).	79
Figura 24 – Arquiteturas utilizadas para o treinamento dos classificadores considerando o conjunto total de características como entrada.....	79
Figura 25- Metodologia utilizada para a escolha de <i>kernels</i> para a classificação.....	82
Figura 26 - Ilustração do processo de busca exaustiva.....	83
Figura 27- Acurácia x Números de neurônios nas camadas escondidas com critério de parada erro médio quadrático.....	86
Figura 28 - Acurácia x Números de neurônios nas camadas escondidas com critério de parada antecipada.....	87
Figura 29 - Acurácia x Números de neurônios nas camadas escondidas com critério de parada regularização.....	88
Figura 30- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados. O critério de parada é o erro médio quadrático, as redes têm 10 neurônios nas camadas escondidas.	89
Figura 31- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados com parada antecipada (ES).	90
Figura 32- Acurácia x Número de características selecionadas por LDA para cada conjunto de dados. O critério de parada é a regularização.....	90
Figura 33- Acurácia x Número de características selecionadas por LDA para cada conjunto de dados. O critério de parada é o erro médio quadrático, as redes têm 10 neurônios nas camadas escondidas.	91

Figura 34- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados com parada antecipada (ES), o número de neurônios nas camadas escondidas é 5 para o conjunto 1, 8 para o conjunto 2 e 10 para o conjunto 3.	92
Figura 35- Acurácia x Número de características selecionadas por LDA para cada conjunto de dados. O critério de parada é a regularização, as redes têm 10 neurônios nas camadas escondidas.	93
Figura 36- Acurácia x Número de características selecionadas com a técnica seleção escalar de características para cada conjunto de dados. O critério de parada é o erro médio quadrático, as redes têm 10 neurônios nas camadas escondidas.	94
Figura 37- Acurácia x Número de características selecionadas com a técnica de seleção escalar de características para cada conjunto de dados com parada antecipada (ES), o número de neurônios nas camadas escondidas é 5 para o conjunto 1, 8 para o conjunto 2 e 10 para o conjunto 3.....	95
Figura 38- Acurácia x Número de características selecionadas com a técnica seleção escalar de características para cada conjunto de dados, o critério de parada é a regularização.	96
Figura 39- Acurácia de diferentes classificadores SVM utilizando o kernel GRBF obtidos através da busca exaustiva para classificação do Conjunto 1.	97
Figura 40- Acurácia de diferentes classificadores SVM utilizando o kernel GRBF obtidos através da busca exaustiva para classificação do Conjunto 2.	97
Figura 41- Acurácia de diferentes classificadores SVM utilizando o kernel GRBF obtidos através da busca exaustiva para classificação do Conjunto 3.	97
Figura 42 - Acurácia de diferentes classificadores SVM utilizando o kernel P obtidos através da busca exaustiva para classificação do Conjunto 1.	99
Figura 43- Acurácia de diferentes classificadores SVM utilizando o kernel P obtidos através da busca exaustiva para classificação do Conjunto 2.....	99
Figura 44 - Acurácia de diferentes classificadores SVM utilizando o kernel P obtidos através da busca exaustiva para classificação do Conjunto 3.	99
Figura 45- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados utilizando o <i>kernel</i> GRBF.	100
Figura 46- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados utilizando o <i>kernel</i> P.	101
Figura 47- Acurácia x Número de características selecionadas com a técnica LDA para cada conjunto de dados utilizando o <i>kernel</i> GRBF	102

Figura 48- Acurácia x Número de características selecionadas com a técnica LDA para cada conjunto de dados utilizando o <i>kernel</i> P	103
Figura 49- Acurácia x Número de características selecionadas com a técnica de seleção escalar para cada conjunto de dados utilizando o <i>kernel</i> GRBF	104
Figura 50- Acurácia x Número de características selecionadas com a técnica de seleção escalar para cada conjunto de dados utilizando o <i>kernel</i> P	105
Figura 51- Efeitos do balanceamento de dados no desempenho dos classificadores RNA (reg, n=10), são analisadas medidas de acurácia e sensibilidade.	106

LISTA DE TABELAS

Tabela 1 – Decomposição percentual de mortes causadas por doenças cardiovasculares Adaptado de (Tung, 2014).	17
Tabela 2 – Comparação entre um sistema IVOCT (C7-XR, FDOCT) e um sistema IVUS Fonte: Adaptado de (Tung, 2014)	23
Tabela 3 - Número de artigos relacionados a CAD apresentados nos encontros anuais da RSNA entre os anos de 2000 e 2010. Adaptado de (Doi, 2007; Li e Nishikawa, 2015).	31
Tabela 4- Resultados comparativos em termos de número de falsos positivos e falsos negativos do detector Histograma dos Auto vetores (HDA) versus o detector Histograma do Gradiente (HDG) (Kerkeni <i>et al.</i> , 2016)	38
Tabela 5 – Resultados relativos a detecção de bifurcação em imagens IVOCT obtidos por Wang (2014).....	39
Tabela 6–Desempenho do classificador <i>Adaboost</i> desenvolvido por Alberti <i>et al.</i> (2011) em termos de acurácia, sensibilidade, especificidade, precisão e taxa de falsos alarmes.	44
Tabela 7– Especificação dos conjuntos de dados utilizados por Macedo <i>et al.</i> (2015).....	46
Tabela 8 – Número de características selecionadas por técnica e conjunto por (Macedo <i>et al.</i> , 2015)	47
Tabela 9 – Resultados em termos de taxa de falsos positivos, sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, <i>f-measure</i> e AUC obtidos por Macedo <i>et al.</i> , 2015.....	48
Tabela 10 – Principais artigos abordados na revisão bibliográfica	50
Tabela 11 – Percentual de amostras das classes BR e NBR em cada conjunto.	75
Tabela 12 – Passos metodológicos a serem implementados em cada uma das arquiteturas de redes neurais utilizadas	80
Tabela 13- Passos metodológicos a serem implementados em cada dos classificadores obtidos através da busca exaustiva.....	84
Tabela 14 – Melhores resultados obtidos utilizando redes neurais com erro médio quadrático e as 104 características como entrada para cada conjunto.....	86
Tabela 15 - Melhores resultados obtidos utilizando redes neurais com parada antecipada e as 104 características como entrada para cada conjunto.....	87
Tabela 16 - Melhores resultados obtidos utilizando redes neurais com regularização e as 104 características como entrada para cada conjunto.....	88

Tabela 17 - Desempenho dos classificadores RNA com erro médio quadrático e conjuntos reduzidos de características por PCA.	89
Tabela 18- Desempenho dos classificadores RNA com parada antecipada e conjuntos reduzidos de características por PCA.	90
Tabela 19- Desempenho dos classificadores RNA com regularização e conjuntos reduzidos de características por PCA.	91
Tabela 20- Desempenho dos classificadores RNA com erro médio quadrático e conjuntos reduzidos de características por LDA.	92
Tabela 21 - Desempenho dos classificadores RNA com parada antecipada e conjuntos reduzidos de características por LDA.	92
Tabela 22- Desempenho dos classificadores RNA com regularização e conjuntos reduzidos de características por LDA. As redes de todos os conjuntos possuem 10 neurônios nas camadas escondidas.	93
Tabela 23- Desempenho dos classificadores RNA com erro médio quadrático e conjuntos reduzidos de características por seleção escalar. As redes de todos os conjuntos possuem 10 neurônios nas camadas escondidas.	94
Tabela 24- Melhores resultados obtidos utilizando redes neurais com parada antecipada e conjuntos reduzidos de características por seleção escalar.	95
Tabela 25- Melhores resultados obtidos utilizando redes neurais com regularização e conjuntos reduzidos de características por seleção escalar. As redes de todos os conjuntos possuem 10 neurônios nas camadas escondidas.	96
Tabela 26- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> GRBF e o conjunto completo de características.	98
Tabela 27- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> polinomial e o conjunto completo de características.	100
Tabela 28- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> GRBF e diferentes conjuntos de características na entrada do classificador selecionadas por PCA.	101
Tabela 29- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> P e diferentes conjuntos de características na entrada do classificador selecionadas por PCA.	101
Tabela 30- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> GRBF e diferentes conjuntos de características na entrada do classificador selecionadas por LDA.	102

Tabela 31- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> P e diferentes conjuntos de características na entrada do classificador selecionadas por LDA	103
Tabela 32- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> GRBF e diferentes conjuntos de características na entrada do classificador selecionadas por seleção escalar.....	104
Tabela 33- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o <i>kernel</i> P e diferentes conjuntos de características na entrada do classificador selecionadas por seleção escalar.	105
Tabela 34- Comparação entre os melhores resultados obtidos com classificadores RNA utilizando o conjunto total de características.	107
Tabela 35- Comparação entre os melhores resultados obtidos com classificadores RNA e diferentes técnicas de seleção de características.....	108
Tabela 36- Comparação entre os melhores resultados obtidos utilizando o classificador SVM e diferentes <i>kernels</i>	109
Tabela 37- Identificação de bifurcações em imagens médicas: comparações com métodos publicados na literatura em termos de acurácia e sensibilidade.	110
Tabela 38- Resumo dos melhores resultados de classificação obtidos por Macedo <i>et al.</i> (2015)	110

LISTA DE QUADROS

Quadro 1- O algoritmo AdaBoost. Adaptado de (Duarte, 2009).	42
---	----

SUMÁRIO

LISTA DE ILUSTRAÇÕES	8
LISTA DE TABELAS	12
SUMÁRIO	16
INTRODUÇÃO	17
REVISÃO BIBLIOGRÁFICA	26
FUNDAMENTAÇÃO TEÓRICA	50
MATERIAIS E MÉTODOS	71
RESULTADOS	83
DISCUSSÃO E CONCLUSÕES	106
REFERÊNCIAS	110
ANEXO	117
APÊNDICE I	120
APÊNDICE II	125

INTRODUÇÃO

Doenças cardiovasculares (DCVs) são as principais causas de morte em todo o mundo. Mais pessoas morrem anualmente de doenças cardiovasculares do que de qualquer outra causa. Estima-se que 17,5 milhões de pessoas morreram de doenças cardiovasculares em 2012, representando 31% de todas as mortes globais (WHO, 2016). Uma decomposição percentual de mortes devido a DCVs na população americana é mostrada na Tabela 1.

Tabela 1 – Decomposição percentual de mortes causadas por doenças cardiovasculares Adaptado de (Tung, 2014).

Tipo de CDV	Percentual de mortes
Doença arterial coronariana	49,9%
Acidente vascular cerebral	16,5%
Pressão alta	7,5%
Insuficiência cardíaca	7,0%
Doenças arteriais	3,4%
Outras	15,6%

O sistema cardiovascular, formado pelo coração e vasos sanguíneos, é responsável pelo transporte de nutrientes e oxigênio para as diversas partes do corpo. Este sistema é formado pelo coração e por dois circuitos: a circulação pulmonar, que leva o sangue desoxigenado do coração para o pulmão retornado ao coração sangue oxigenado; e a circulação sistêmica, que leva sangue oxigenado do coração para os órgãos, tecidos e células do corpo, levando sangue pobre em oxigênio de volta ao coração (Tung, 2014). A anatomia básica do coração é ilustrada na Figura 1.

O músculo cardíaco precisa de sangue rico em oxigênio para sobreviver. Este sangue é fornecido para o coração pelo seu próprio sistema vascular, chamado também de circulação coronariana.

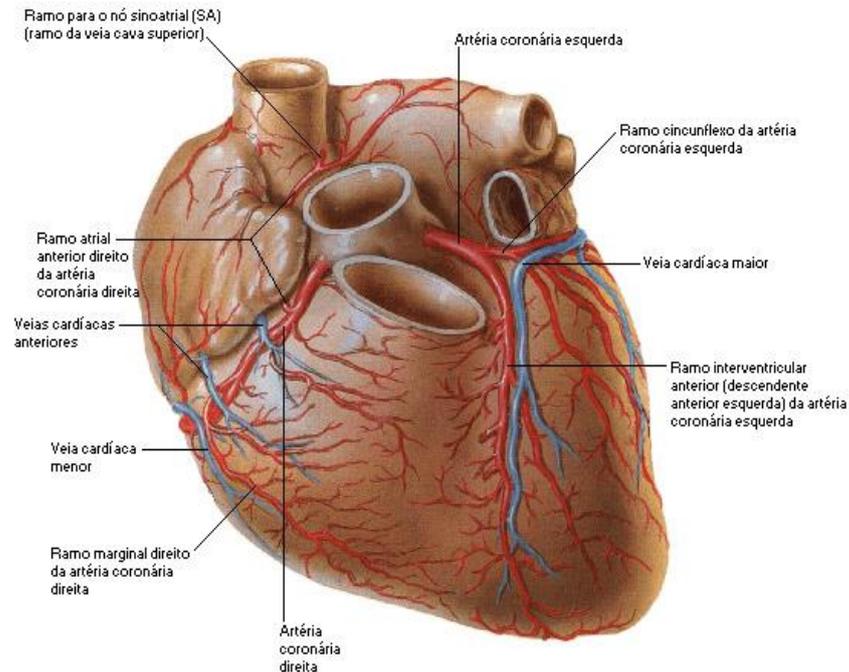


Figura 1 - Anatomia do coração humano. Fonte: Adaptado de (Siu, 2012)

A artéria aorta, principal fornecedora de sangue para o corpo, se ramifica na porção ascendente em duas artérias coronárias, uma à direita e outra à esquerda, que vão irrigar o coração. Estas artérias coronárias fornecem sangue rico em oxigênio para todo o músculo cardíaco. Caso ocorra oclusões das artérias coronárias, o suprimento de oxigênio a determinadas regiões do coração é severamente reduzido ou interrompido, levando ao infarto. As doenças causadas pela oclusão das coronárias são discutidas a seguir (Tung, 2014).

Da Tabela 1, pode-se verificar dentre as doenças cardiovasculares, a maior responsável por mortes é a doença arterial coronariana (DAC), também chamada de aterosclerose. A aterosclerose é caracterizada pelo bloqueio ou interrupção do fornecimento de sangue ao coração devido a um acúmulo de substâncias gordurosas, chamadas de ateromas, nas artérias coronárias. Pode-se observar este processo na Figura 2. O acúmulo de placa provoca o estreitamento da parede das artérias coronárias, restringindo o fornecimento de sangue para o coração podendo causar angina, aterotrombose e infarto agudo do miocárdio (Siu, 2012).

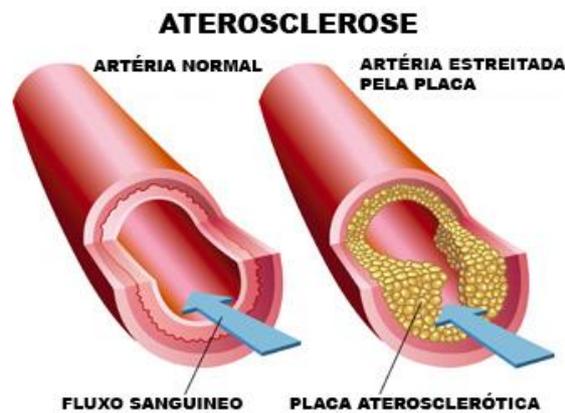


Figura 2 - Ilustração do estreitamento da artéria por placa aterosclerótica Fonte: Adaptado de (Siu, 2012)

As regiões de bifurcação das artérias são caracterizadas por transições na configuração da parede arterial. Estas transições estão associadas a alterações na taxa do fluxo sanguíneo, que desempenha um papel relevante na progressão da aterosclerose (Zarins *et al.*, 1983).

Estas regiões também são predispostas ao desenvolvimento de espessamentos da íntima, incluindo a formação de placas ateroscleróticas. A presença de placas pode reduzir significativamente o fluxo sanguíneo, levando a graves complicações de saúde, como anginas e acidentes vasculares cerebrais. Detectar e caracterizar estas regiões potencialmente vulneráveis pode ajudar a identificar a presença de placas e a compreender a progressão da doença.

A angina, ou angina de peito, refere-se a dor no peito devido ao entupimento das artérias coronárias. Ela ocorre quando há um desequilíbrio entre a demanda e o fornecimento de oxigênio no coração. Este desequilíbrio surge a partir de um aumento da procura, sem um aumento proporcional da oferta de oxigênio, por exemplo, durante a prática de algum exercício físico. Quando os sintomas da doença arterial coronária não mudam ao longo do tempo, o doente é dito ter angina de peito estável (Aha, 2015).

A aterotrombose (Viles-Gonzalez *et al.*, 2004) é um processo patológico evolutivo que começa com a ruptura de uma placa aterosclerótica. A ruptura da placa expõe extenso material pró-coagulante subintimal formando o trombo, que causa a oclusão parcial ou total das artérias,

resultando em angina instável. Quando a oclusão das artérias aumenta o paciente pode sofrer dor no peito, muitas vezes com irradiação para o braço esquerdo associado a náuseas e sudorese. Estes sintomas podem ser referidos a síndrome coronariana aguda (SCA). Se uma parcela significativa dos músculos do coração estiver danificada, o coração pode parar de bater regularmente.

SCA é um termo utilizado para descrever uma série de doenças resultantes da obstrução das artérias coronarianas (Pesaro et al., 2008). Pacientes com oclusão total podem apresentar infarto agudo do miocárdio (IAM) com supra desnivelamento do segmento ST. Obstrução parcial de um vaso pode resultar em um IAM sem supra desnivelamento do segmento ST ou angina instável, que exigirá estabilização clínica precoce; seguida por uma estratificação de risco criteriosa para a definição das estratégias terapêuticas (invasivas ou conservadoras).

Os *stents* coronários são, atualmente, a estratégia de caráter invasivo mais utilizada para tratamento da DAC. Isto é devido tanto aos melhores resultados imediatos (menor incidência de complicações imediatas) quanto tardios (menor chance de recorrência da lesão tratada). O *stent* é uma estrutura minúscula, expansível e em forma de tubo, feita de um metal de aço inoxidável. A estrutura é montada em um balão minúsculo que é aberto dentro de uma artéria coronariana com o intuito de empurrar a placa e restaurar o fluxo sanguíneo (Serruys et al., 2006). A Figura 3 ilustra este processo.

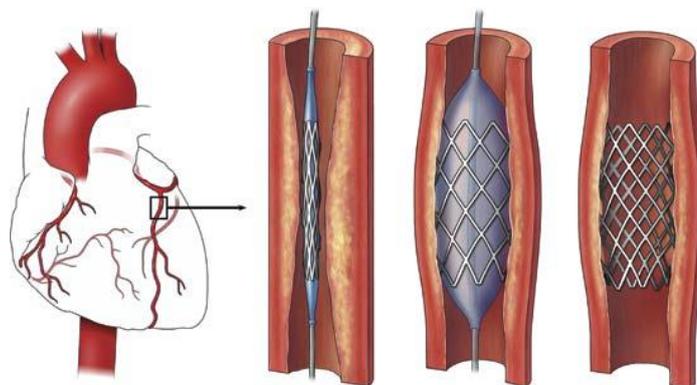


Figura 3 – Aplicação do stent coronário. Fonte: (Webmd, 2016)

No entanto, como o *stent* é um objeto estranho, sua presença pode incitar uma resposta imunitária que tende a contrair ainda mais a região próxima ou interna ao *stent*. O crescimento de tecido cicatricial através do *stent* é conhecido como hiperplasia neointimal e pode fazer com que a artéria seja ocluída novamente, processo conhecido como reestenose. As taxas de reestenose tardia giram em torno de 15 a 25%, permanecendo como um problema clínico de grande importância (Oliveira *et al.*, 2005).

Além de reestenose, a angioplastia e o implante de *stent* também podem causar lesão endotelial e inflamação, tornando tanto o *stent* quanto a artéria, altamente trombogênicos (Gawaz *et al.*, 1996; Caramori *et al.*, 1999).

A fim de prevenir a reestenose e a trombose causada pelos *stents*, modalidades de imagens intravasculares como tomografia por coerência óptica intravascular (*Intravascular Optical Coherence Tomography-IVOCT*) e ultrassom intravascular (*Intravascular Ultrasound - IVUS*) são utilizadas para observar os sintomas clínicos, como o crescimento de placa ou da camada neointima e para monitorar a posição do *stent*.

IVUS e IVOCT são modalidades de imagens médicas invasivas baseadas em tecnologias de ultrassom e infravermelho próximo, respectivamente. Em ambas as modalidades, a aquisição de imagem é realizada pela inserção de um cateter específico dentro da artéria seguido pela execução de um movimento de recuo, o chamado *pullback*. Assim, imagens de seção transversal com informações anatômicas, morfológicas e patológicas das artérias são adquiridas (Rogowska e Brezinski, 2000; Tung *et al.*, 2011; Tsantis *et al.*, 2012).

IVOCT foi proposto como um método de imageamento de alta resolução para a caracterização de placas na artéria coronária (Huang *et al.*, 1991). Trata-se de uma modalidade baseada na interferência entre um campo ótico de baixa coerência que é dividido e posteriormente recombinado após sofrer reflexão na amostra. Esta tecnologia de imageamento foi desenvolvida no início dos anos 90 por pesquisadores do *Massachusetts Institute of*

Technology (MIT) para obter imagens de alta resolução de tecido transparente (tecido do olho) (Marques, 2012).

Para aquisição de imagens OCT a artéria deve estar limpa de sangue antes do processo, uma vez que o feixe de luz infravermelho é atenuado pela presença de glóbulos vermelhos (Costopoulos *et al.*, 2016). A técnica IVOCT proporciona imagens em corte transversal e 3D (2D + tempo) de tecido biológico com resoluções extremamente altas, atingindo até a escala de micron. Várias linhas axiais (*A-line*) são continuamente adquiridas de acordo com a rotação do cateter. Uma revolução completa cria uma seção transversal completa do vaso (*bScan*) (Bezerra *et al.*, 2009). A Figura 4 mostra um exemplo de IVOCT numa artéria coronária.

O imageamento por IVOCT permite uma velocidade de aquisição de cerca de 100 a 160 imagens por segundo a uma velocidade de recuo na faixa de 15mm/s a 25mm/s, resultando em um grande número de imagens por procedimento (Schmitt, 2003). Durante a aquisição de imagens IVOCT em aplicações clínicas, uma única corrida de cateter contém tipicamente centenas de imagens.

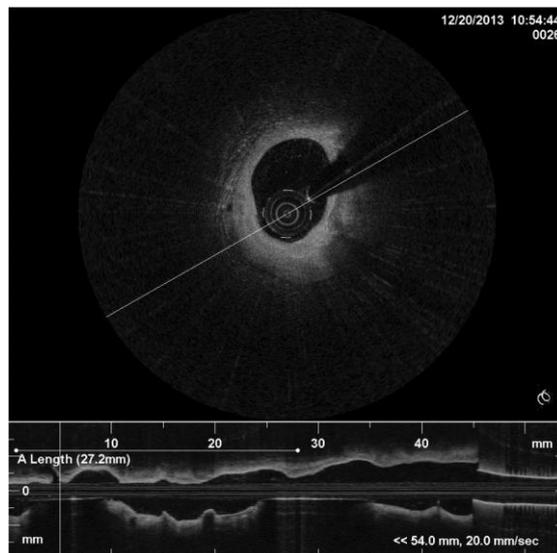


Figura 4 – Exemplo de imagem IVOCT: A imagem superior corresponde a um corte transversal da artéria coronariana. A imagem inferior é a vista longitudinal da coronária.

O mais recente sistema de aquisição de imagens IVOCT é o sistema C7-XR. Imagens adquiridas por este sistema são representadas no domínio da frequência, denominadas imagens FDOCT.

A Figura 5 apresenta exemplos das duas modalidades de imageamento, IVOCT e ultrassom intravascular (IVUS). Na Tabela 2 pode-se verificar uma comparação mais detalhada entre elas.

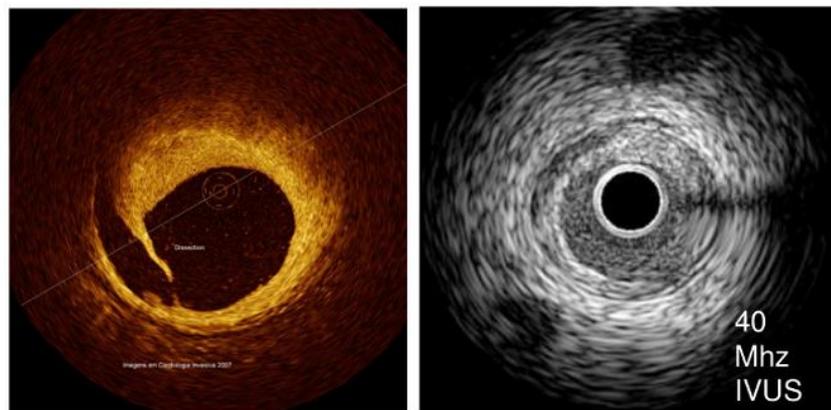


Figura 5 – Imagens do mesmo corte em OCT e em IVUS. Fonte: (Macedo, 2014)

**Tabela 2 – Comparação entre um sistema IVOCT (C7-XR, FDOCT) e um sistema IVUS
Fonte: Adaptado de (Tung, 2014)**

Característica	C7-XR	IVUS
Resolução espacial	10 μ m-20 μ m	100 μ m - 200 μ m
Largura do feixe	20 – 40mm	200 -300mm
Taxa de aquisição de imagens	100 imagens/s	30 imagens/s
Velocidade do pullback	20mm/s	0.5-1mm/s
Diâmetro de alcance máximo	10mm	15mm
Penetração no tecido	1,0mm – 2,0 mm	10mm
Linhas por imagem	500	256
Resolução lateral	19 μ m	225 μ m
Sonda de fio-guia	Sim	Sim
Remoção do sangue das artérias	Sim	Não

Uma comparação sistemática e padronizada entre modalidades de imagens invasivas e não invasivas para a caracterização da placa coronária, desenvolvida por Maurovich-Horvat *et al.* (2012), demonstraram que a tomografia computadorizada coronariana e o IVUS estão

fracamente associados com a composição da placa e classificação da lesão de acordo com achados histopatológicos, enquanto IVOCT foi fortemente associada.

Estudos *in vitro* têm demonstrado que, em face da resolução espacial das imagens IVOCT (que giram em torno de 10 μm a 20 μm), as mesmas podem ser utilizadas para analisar os componentes e a espessura da placa ou da camada neointima (Costopoulos et al., 2016).

As propriedades ópticas inerentes aos constituintes típicos de cada placa fornecem um contraste suficiente para a diferenciação de lipídios, cálcio e tecido fibroso. Quando utilizado para avaliar segmentos arteriais com *stent*, a IVOCT pode diferenciar com precisão a escora do *stent* e o tecido vascular em torno dele. Portanto, a IVOCT pode ser uma ferramenta valiosa para avaliar a cicatrização adequada de segmentos coronarianos com *stent* e orientar a terapia farmacológica para prevenir a trombose tardia causada por *stent* (Costopoulos et al., 2016).

Ferramentas para identificação de bifurcações em imagens IVOCT poderiam então, além de acelerar o processo de análise destas imagens, auxiliar na automatização de processos de análise de *stent* (Wanget al., 2014) e de análise de placa (Elbasiony e Levkowitz, 2011).

O domínio de interesse deste trabalho é a identificação de bifurcações em imagens clínicas IVOCT. Alguns trabalhos publicados na literatura já desenvolveram métodos para identificação de bifurcações em imagens vasculares, dentre os quais citam-se os trabalhos de Zhou *et al.* (2007); Alberti *et al.* (2011); Mengliu e Hamarneh, (2014) e Kerkeni *et al.*,(2016). Sendo a modalidade de imageamento IVOCT relativamente recente, poucos são trabalhos que fazem uso dela, a saber: Wang *et al.* (2014) e Macedo *et al.* (2015).

O trabalho proposto nesta dissertação concentra-se nas etapas finais da tarefa de reconhecimento de padrões. Trata-se, portanto, da avaliação da redução de dimensionalidade de dados (PCA, LDA e Seleção Escalar de Características) em associação com classificadores, buscando com tais experimentos melhorar a taxa de reconhecimento obtida em Macedo *et al.*, 2015.

As técnicas de classificação utilizadas são máquinas de vetores de suporte e redes neurais artificiais. No que diz respeito ao uso das redes neurais, serão avaliados três critérios de parada: o erro médio quadrático, parada antecipada e regularização.

1.1 OBJETIVOS

1.1.1 Objetivo geral

O objetivo geral proposto nesta pesquisa é avaliar diferentes metodologias de classificação na tarefa de identificar regiões de bifurcação em imagens IVOCT.

1.1.2 Objetivos específicos

Os objetivos específicos consistem em:

- Avaliar se o uso de redes neurais, com diferentes critérios de parada: erro médio quadrático, parada antecipada e regularização, na tarefa de identificação de bifurcações podem contribuir para a obtenção de melhores taxas de identificação de bifurcação em imagens IVOCT.
- Avaliar se o uso de Máquinas de Vetores de Suporte, utilizando *grid-search* para a otimização dos parâmetros inerentes ao classificador, podem contribuir para a obtenção de melhores taxas de identificação de bifurcação em imagens IVOCT.
- Avaliar se o uso de métodos de redução de características como: Análise de componentes principais, Análise do Discriminante Linear e Seleção Escalar de Características, em associação aos classificadores SVM e Redes Neurais podem contribuir para obtenção de melhores taxas de identificação de bifurcação em imagens IVOCT.

1.1 ORGANIZAÇÃO

Esta dissertação está dividida nos seguintes capítulos e seções:

- Capítulo 1: Introdução;
- Capítulo 2: Revisão Bibliográfica;
- Capítulo 3: Fundamentação Teórica;
- Capítulo 4: Materiais e Métodos;
- Capítulo 5: Resultados;
- Capítulo 6: Discussões e Conclusões;
- Referências
- Anexo
- Apêndices

O capítulo um trata a caracterização do problema, o contexto histórico da área de estudo, a delimitação do trabalho e descreve os objetivos gerais e específicos desta dissertação. No capítulo dois são apresentados trabalhos relacionados a identificação de bifurcações arteriais em imagens médicas.

O capítulo três aborda a fundamentação teórica necessária para o desenvolvimento do trabalho, como redes neurais artificiais e máquinas de vetores de suporte para a classificação dos dados e métodos para seleção de características.

No capítulo quatro são apresentados os materiais utilizados no desenvolvimento da pesquisa. O banco de dados é descrito com detalhes e também são apresentados os procedimentos realizados no desenvolvimento desta pesquisa.

No capítulo cinco são apresentados os resultados dos testes preliminares realizados bem como os comentários acerca dos mesmos. No capítulo seis, as conclusões do trabalho são apresentadas. Em anexo encontram-se uma breve descrição das métricas de desempenho utilizadas para avaliar os classificadores.

No apêndice I, encontra-se uma cópia do artigo publicado intitulado “Identificação de regiões de bifurcação em imagens de tomografia intravascular por coerência ótica utilizando máquinas de vetores de suporte” de Carmina Nascimento, Marly Guimarães Costa, Maysa Macedo, Marco Gutierrez e Cícero Costa Filho. Apresentado no XXV Congresso Brasileiro de Engenharia Biomédica, realizado em Foz do Iguaçu, Paraná, de 17 a 20 de outubro de 2016

No apêndice II, encontra-se uma cópia do artigo intitulado “*Classification of bifurcations regions in IVOCT images using support vector machine and artificial neural network models*” de Carmina Nascimento, Cícero Costa Filho, Maysa Macedo, Marco Gutierrez e Marly Guimarães Costa, publicado em “*Proceedings of the Medical Imaging 2017: Computer-Aided Diagnosis conference*”, Orlando, Flórida, Estados Unidos.

REVISÃO BIBLIOGRÁFICA

A detecção de bifurcações vasculares é muito importante em aplicações clínicas, tais como o diagnóstico de estenose do vaso sanguíneo e no planejamento cirúrgico, onde os pontos de ramificação são utilizados como pontos de referência. A identificação de bifurcações também tem auxiliado na reconstrução 3D dos vasos.

Dentre os trabalhos publicados na literatura que desenvolveram métodos para identificação de bifurcação em imagens vasculares, nota-se o uso de duas abordagens distintas: técnicas de processamento digital de imagens (Baboiu e Hamarneh, 2012; Wang *et al.*, 2014 e Kerkeni *et al.*, 2016) e técnicas de reconhecimento de padrões (Zhou *et al.*, 2007; Alberti *et al.*, 2011 e Macedo *et al.*, 2015).

2.1 PESQUISA BIBLIOGRÁFICA

A pesquisa bibliográfica buscou identificar de que forma são detectadas as bifurcações arteriais em imagens vasculares, focando nos algoritmos utilizados para esta tarefa. Portanto, não são detalhadas as técnicas de segmentação utilizadas pelos autores.

As seguintes bases de dados literárias foram utilizadas: *IEEE Explorer*, *Web of Science* e *PubMed*. Os resultados obtidos na busca foram filtrados baseados na sua relevância e compatibilidade com as palavras-chave selecionadas. A partir da análise destes artigos, foram identificadas as principais técnicas de identificação de bifurcações em imagens vasculares, as quais serão apresentadas na seção 2.3. Antes porem apresenta-se uma breve perspectiva histórica do uso de ferramentas computacionais para auxílio ao diagnóstico.

2.2 DIAGNÓSTICO AUXILIADO POR COMPUTADOR NA RADIOLOGIA

Datam de 1950 as primeiras discussões sobre o uso de ferramentas computacionais para a análise de anormalidades em imagens médicas (Giger *et al.*, 2008). Naquele momento, a expectativa era de que os computadores seriam capazes de substituir o radiologista na detecção de anormalidades, uma vez que computadores e máquinas podem desempenhar algumas tarefas de forma mais eficiente que o ser humano (Doi, 2007).

Embora resultados animadores tenham sido reportados, as tentativas iniciais não foram bem-sucedidas. Os computadores não tinham um alto poder de processamento, ainda não existiam técnicas avançadas de processamento de imagens e o acesso a imagens digitais era muito restrito. Não era possível então, atingir o objetivo inicial de desenvolver um *software* automático para detecção ou diagnóstico através de imagens digitais que pudesse substituir o papel do especialista (Doi, 2007; Giger *et al.*, 2008).

Na década de 1980 uma nova abordagem foi proposta, o diagnóstico auxiliado por computador (CAD). Esta abordagem assumia que a saída de um *software* pode ser utilizada pelos médicos, mas jamais substituí-los, diferenciando-se da proposta inicial que visava dar total autonomia ao *software* (Doi, 2007; Giger *et al.*, 2008).

Neste contexto, CAD pode ser definido como um diagnóstico feito por um radiologista, levando em consideração a saída de uma análise computacional das imagens, no seu processo de decisão. A decisão médica final é feita pelo radiologista, não pelo computador. Assim, o uso do CAD não tem a intenção de substituir o radiologista, mas auxiliar na interpretação da imagem e na tomada de decisão do especialista. A Figura 6 esquematiza este conceito.

O objetivo do CAD é aprimorar a qualidade do diagnóstico mediante o uso da resposta do computador como referência. Esta resposta pode ser muito útil, uma vez que o diagnóstico do radiologista é baseado em uma avaliação subjetiva e está sujeito a variações intra e interpessoais. Fatores como baixa qualidade da imagem, sobreposição de estruturas e fadiga

visual podem influenciar negativamente a tomada de decisão do radiologista (Azevedo-Marques, 2001).

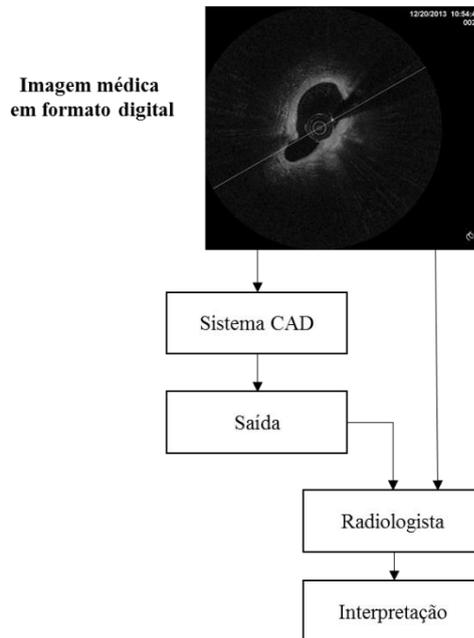


Figura 6 – Diagrama esquemático de um sistema de auxílio ao diagnóstico para interpretação de imagens médicas. Adaptado de (Giger *et al.*, 2008).

Comumente, os sistemas e técnicas CAD são classificados em detecção auxiliada por computador (CADe) e diagnóstico auxiliado por computador (CADx). CADe são relacionados a procedimentos onde o radiologista utiliza as respostas do *software* para localizar as regiões de interesse. Desta forma, a caracterização, diagnóstico e a avaliação do paciente são tarefas exclusivas do médico. O CADe pode ser visto como uma ferramenta de detecção ou identificação. O CADx estende a análise computacional para que a saída do *software* seja uma caracterização de alguma região ou lesão previamente identificada, podendo ser visto como uma ferramenta de auxílio ao diagnóstico diferencial (Giger *et al.*, 2008).

Em geral os sistemas CAD se utilizam de técnicas provenientes de duas áreas do conhecimento: visão computacional, que envolve o processamento da imagem para realce, segmentação e extração de atributos e reconhecimento de padrões, que inclui métodos para a seleção de atributos, estatística e classificação (Azevedo-Marques, 2001).

A ideia do CAD pode ser aplicada a todas as modalidades de obtenção de imagens médicas, uma vez que sua base conceitual é muito ampla. O número de artigos relacionados a CAD apresentados nos encontros anuais da Sociedade de Radiologia da América do Norte (RSNA) entre os anos de 2000 e 2010 é listado na Tabela 3. A grande parte destes trabalhos está relacionada, basicamente, a três regiões: tórax (618 trabalhos), mamas (433 trabalhos) e cólon (213 trabalhos).

Tabela 3 - Número de artigos relacionados a CAD apresentados nos encontros anuais da RSNA entre os anos de 2000 e 2010. Adaptado de (Doi, 2007; Li e Nishikawa, 2015).

Região	Ano											Total geral
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	
Tórax	22	37	53	94	70	48	62	72	73	45	42	618
Mamas	23	28	32	37	48	49	47	39	51	42	37	433
Cólon	4	10	21	17	15	30	25	32	24	14	21	213
Cérebro	-	4	2	10	9	15	12	13	20	3	11	99
Fígado	3	-	5	9	9	9	8	8	22	8	15	96
Esqueleto	2	7	7	9	8	5	7	11	6	4	3	69
Vascular, etc.*	5	-	12	15	2	7	6	17	31	16	27	138
Total anual	59	86	132	191	161	163	167	192	227	132	156	1666

*Cardíaco, próstata, pediátrico e dentário.

RSNA: Sociedade de Radiologia da América do Norte

A partir da Tabela 3 podemos notar que outras regiões, como por exemplo o cérebro, o fígado, o esqueleto e o sistema vascular, ainda que com poucos trabalhos desenvolvidos, também são objetos de interesse nas pesquisas.

2.3 IDENTIFICAÇÃO DE BIFURCAÇÕES VASCULARES EM IMAGENS MÉDICAS

A identificação automática de bifurcações vasculares em imagens médicas implica em utilizar a saída de um programa de computador para auxílio na detecção de locais de interesse em um determinado conjunto de imagens, caracterizada assim como CADe. O uso de tais sistemas tem sido benéfico em exames em que muitas imagens são adquiridas e a maior parte dos casos é caracterizada como sendo “normal” (Giger *et al.*, 2008).

Muitos algoritmos para detecção de bifurcação foram propostos nos últimos anos. A maioria dos métodos propostos na literatura para a detecção automática de bifurcações são baseados em uma análise topológica realizada após a segmentação das imagens.

As características assimétricas das bifurcações e o uso de diferentes modalidades de imagens médicas, tornam difícil a tarefa de fixar parâmetros para a maioria dos algoritmos. Neste sentido, o uso de métodos baseados em aprendizagem de máquina tem aumentado.

Ao analisar os artigos selecionados nesta revisão bibliográfica nota-se duas principais abordagens para a identificação de bifurcações vasculares em imagens médicas, as baseadas em técnicas aprendizado de máquina e as baseadas em técnicas de visão computacional. Ambas as abordagens compartilham etapas fundamentais para o sucesso da tarefa de classificação, as atividades relativas ao processamento digital de imagens.

De acordo com Gonzalez e Woods (2009), o processamento digital de imagens envolve processos cujas entradas e saídas são imagens e, além disso, envolve processos de extração de atributos de imagens até o reconhecimento de objetos individuais. As etapas compreendidas no processamento digital de imagens podem ser sumarizadas como mostrado na Figura 7.

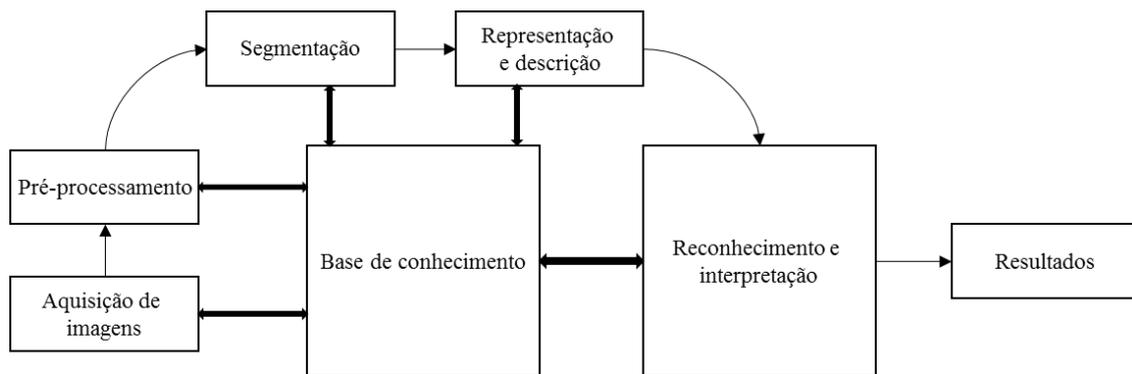


Figura 7 - Fluxograma de etapas de classificação de imagens utilizando técnicas de processamento digital de imagens. Fonte: Adaptado de (Gonzalez e Woods, 2009)

A aquisição de imagens é o primeiro processo da Figura 7. Esta etapa consiste em obter uma representação da informação visual, que deve ser a mais fiel possível ao elemento físico e ao mesmo tempo, ser processável por um computador. (Gonzalez e Woods, 2009).

No que diz respeito a aquisição de imagens vasculares, os métodos amplamente utilizados são: raios-X, imagem por ressonância magnética (MRI) o ultrassom (US), a tomografia por emissão de pósitrons (PET), a tomografia computadorizada (CT) e a tomografia por coerência óptica (OCT) (Upputuri *et al.*, 2015).

De acordo com Gonzalez e Woods (2009), a etapa de pré-processamento envolve tarefas como filtragem e realce de imagens, sendo definido como o processo de manipular uma imagem de forma que o resultado seja mais adequado do que o original para uma aplicação específica. Não existe uma teoria geral para o pré-processamento de imagens, as tarefas realizadas nesta etapa são orientadas de acordo com o problema em questão. Quando uma imagem é processada para a interpretação visual, o observador é o juiz final do bom funcionamento de um método em particular.

Os procedimentos de segmentação dividem uma imagem em suas partes ou objetos constituintes. Em geral, a segmentação automática é uma das tarefas mais difíceis do processamento digital de imagens. Um bom procedimento de segmentação aumenta as chances de sucesso na resolução de problemas que requerem que objetos sejam individualmente identificados (Gonzalez e Woods, 2009).

A base de conhecimento, que se relaciona com quase todas as etapas ilustradas na Figura 7, é um dos pontos fundamentais do processamento digital de imagens. O conhecimento sobre o domínio do problema está codificado em um sistema de processamento digital de imagens na forma de uma base (banco de dados) de conhecimento.

Este conhecimento pode ser tão simples quanto o detalhamento de regiões de uma imagem na qual se sabe onde a informação de interesse pode estar localizada, limitando a busca que precisa ser conduzida na procura daquela informação. A base pode ser também bastante complexa, como por exemplo, uma lista inter-relacionada de todos os principais defeitos possíveis em um problema de inspeção. Além de orientar a operação de cada módulo de

processamento, a base de conhecimento também controla a interação entre os módulos (Gonzalez e Woods, 2009).

A representação e descrição geralmente partem do resultado de uma etapa de segmentação. É necessário converter os dados provenientes da segmentação de forma adequada para o processamento computacional. Deve-se escolher a forma com a qual os dados devem ser representados, fronteiras ou regiões completas.

A partir da Figura 7, identifica-se que a próxima etapa do processo é chamada de reconhecimento e interpretação. É neste ponto em que geralmente ocorre a divisão entre abordagens baseadas em visão computacional e em abordagens baseadas em reconhecimento de padrões.

Enquanto as abordagens baseadas em processamento digital de imagens continuam aplicando algoritmos para extrair o máximo de informações possíveis e atribuir rótulos a objetos na imagem, as abordagens baseadas em reconhecimento de padrões aplicam classificadores baseados em aprendizagem de máquina.

2.4 ABORDAGENS BASEADAS EM TÉCNICAS DE VISÃO COMPUTACIONAL

O reconhecimento é o processo que atribui um rótulo a um objeto tomando como base os seus descritores, computados em regiões ao redor de pontos de interesse. Pontos de interesse, na maioria das vezes são vértices de contornos ou regiões homogêneas na imagem. Eles costumam ser identificados por meio de detectores de pontos e normalmente carregam as seguintes informações: uma coordenada 2D na imagem, uma orientação e uma escala.

Dentre os algoritmos utilizados para identificação de bifurcações em imagens médicas, destacam-se aqueles baseados em modelos. Num sentido amplo, modelos encapsulam informações sobre a estrutura de um objeto em análise. A principal distinção feita em termos de modelos neste caso é entre aparência (detalhes relacionados a iluminação) e informações

geométricas. Normalmente, em tarefas de detecção são utilizados os dois tipos de modelos para compor a base do conhecimento (Lesage *et al.*, 2009).

Os modelos de aparência expressam conhecimento inicial a respeito das propriedades de iluminação das estruturas vasculares. A aparência dos vasos sanguíneos em imagens médicas é extremamente dependente da modalidade de imagem utilizada. Em imagens como angiotomografia computadorizada e angioressonância, os vasos são estruturas mais brilhantes que as estruturas encontradas ao seu redor. Esta informação caracteriza um modelo de aparência (Lesage *et al.*, 2009).

Outra característica chave dos vasos sanguíneos é a sua forma. O primeiro modelo geométrico de bifurcação, desenvolvido por Murray (1926), incluía relações entre ângulos de ramificação e larguras dos vasos baseado em condições fisiológicas ótimas. Este modelo foi testado e aprimorado ao longo dos anos. Atualmente o modelo geométrico mais aceito de uma bifurcação, uma adaptação do modelo original, é mostrado na Figura 8.

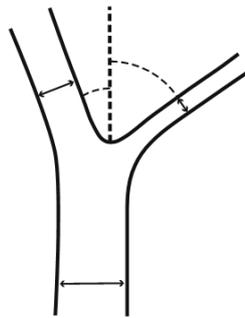


Figura 8 – Modelo geométrico de uma bifurcação. Fonte: (Lesage *et al.*, 2009)

Naturalmente, o modelo que representa uma bifurcação também depende da modalidade de imagem em questão. Em modalidades de imagens que captam o exterior do vaso sanguíneo, como é o exemplo da angiografia (Figura 9), uma técnica estabelecida para o imageamento da coronária, o modelo da Figura 8 pode ser considerado.

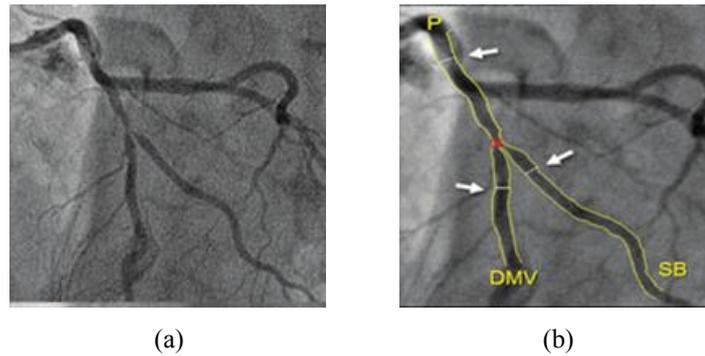


Figura 9 – Exemplos de imagem de uma angiografia (a) original (b) segmentada com destaque nos seguintes pontos vaso principal (P), vaso principal distal (DMV) e ramo lateral (SB). Fonte: adaptado de (Cardialysis, 2014).

Já em imagens que captam o interior do lúmen, como por exemplo o IVOCT, a bifurcação assemelha-se a uma cavidade no tecido do vaso sanguíneo. A Figura 10 mostra um comparativo entre uma imagem IVOCT sem bifurcação e com bifurcação. Este padrão de cavidade também é encontrado em imagens do tipo IVUS.

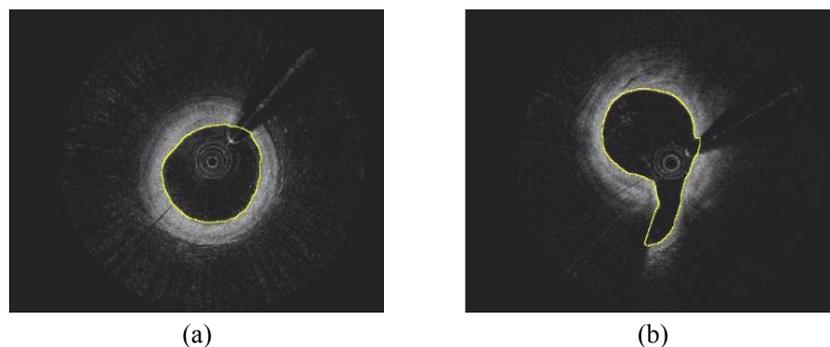


Figura 10 – Exemplos de imagens IVOCT (a) sem bifurcação e (b) com bifurcação. Foram selecionadas imagens com o lúmen segmentado para melhor identificação das regiões de interesse.

Em Baboiu e Hamarneh (2012), os autores apresentam um modelo analítico para a identificação de bifurcações, e o combinam com a análise de auto vetores para criar um filtro capaz de identificar bifurcações, chamado de filtro *bifurcation-Ness*. O método proposto é baseado no comportamento das bifurcações no espaço-escala.

O algoritmo de Baboiu e Hamarneh (2012) foi testado em imagens médicas e em imagens sintéticas. As imagens sintéticas são do tipo 2D e foram geradas com diferentes espessuras de vasos sanguíneos, com e sem o acréscimo de ruído gaussiano e de ruído do tipo “sal e pimenta”. Já as imagens médicas utilizadas no trabalho, são provenientes de um conjunto de 5 diferentes angiografias de retina com centenas de bifurcações cada uma.

No que diz respeito as imagens sintéticas, o algoritmo foi capaz de classificar corretamente 95% das bifurcações quando o desvio padrão foi de 0,4 e a densidade do ruído “sal e pimenta” foi 35%. De acordo com os autores, este foi o maior nível de ruído aplicado. Quando testado no conjunto de imagens médicas, o algoritmo desenvolvido foi capaz de detectar 403 bifurcações, dentre 421 existentes, resultando em uma acurácia de 95,7%. Ainda de acordo com os autores, em todas as imagens utilizadas para fins de teste, a acurácia se manteve entre 95% e 97%.

Embora simples, o algoritmo desenvolvido apresentou bons resultados tanto em imagens 2D quanto 3D. No entanto, o algoritmo não é capaz de diferenciar bifurcações de cruzamentos de vasos.

Em Mengliu e Hamarneh (2014), uma nova característica, denominada Histograma dos Auto vetores (HDA) é proposta. O pensamento essencial por trás do HDA é que a bifurcação pode ser detectada graças a observação da orientação local do vaso sanguíneo, fornecida pelo auto vetor correspondente ao menor autovalor contido na matriz hessiana. Dado um pixel, o histograma HDA para todos os elementos da sua vizinhança é calculado. Pixels cujos histogramas das vizinhanças contém três picos, são considerados os mais prováveis a serem pontos de bifurcação.

Utilizando o trabalho de Mengliu e Hamarneh (2014) como base, Kerkeni *et al.* (2016) propuseram um novo detector baseado em uma escala específica do histograma dos auto vetores, onde a vizinhança de cada pixel é tomada a partir de uma escala ótima, levando a uma classificação mais precisa.

O método foi testado em imagens clínicas reais, obtidas a partir de cinco diferentes pacientes. Os angiogramas foram adquiridos durante exames de rotina realizados no departamento de cardiologia do Hospital Universitário Fattouma Bourguiba, Monastir, Tunísia.

De acordo com os autores, as imagens utilizadas apresentam diferentes níveis de dificuldade de classificação, devido a iluminação e ruído.

Os autores compararam o detector HDA desenvolvido com o detector HDG, conforme consta na Tabela 4.

Tabela 4- Resultados comparativos em termos de número de falsos positivos e falsos negativos do detector Histograma dos Auto vetores (HDA) versus o detector Histograma do Gradiente (HDG) (Kerkeni *et al.*, 2016)

Imagem	HDG		HDA	
	FP	FN	FP	FN
Imagem 1	6	5	4	3
Imagem 2	7	3	3	2
Imagem 3	23	5	8	5
Imagem 4	10	5	4	4
Imagem 5	8	5	3	2
Média	10,8	4,6	4,4	3,2

FP: Falsos positivos e FN: Falsos negativos

Uma das vantagens relatadas por Kerkeni *et al.* (2016) é a diminuição das taxas de falsos negativos quando o algoritmo é comparado com o HDG. Porém os resultados poderiam ser melhor analisados se os autores expusessem a tabela de confusão completa, indicando a quantidade de bifurcações existentes no seu banco de dados para que fosse possível comparar os algoritmos utilizando outras métricas, como acurácia, sensibilidade e especificidade.

Além da matriz hessiana, outras abordagens baseadas em modelos levam em consideração as características geométricas das bifurcações, como é o caso de Wang *et al.* (2014), onde os autores partem do princípio de que uma bifurcação pode ser identificada computando a distância do centro do lúmen ao contorno mais afastado da camada íntima. Conforme pode ser visualizado na Figura 10, uma bifurcação em imagens IVOCT tem uma tendência a distorcer o formato elíptico do contorno do lúmen.

Em Wang *et al.* (2014), os autores propõem um método completamente automático para detecção de bifurcações baseado em segmentações de todos os componentes típicos em imagens IVOCT. O trabalho é extremamente dependente da fase de pré-processamento da imagem, que

compreende as seguintes etapas: detecção do cateter, detecção do fio guia, detecção do artefato de proteção e detecção do contorno do lúmen.

O algoritmo desenvolvido por Wang *et al.* (2014) foi avaliado em imagens IVOCT de 25 diferentes *pullbacks*. Todas as imagens foram adquiridas com um sistema OCT C7-XR (St. Jude Medical, Westford, MA, USA). Destes *pullbacks*, 23 eram compostos por 271 imagens e 2 por 541 imagens.

Dois especialistas avaliaram as imagens e de maneira independente, identificaram imagens com bifurcações. O primeiro especialista avaliou todas as imagens dos 25 *pullbacks* e identificou um total de 82 bifurcações. O segundo especialista avaliou um subconjunto de 5 *pullbacks*. Os resultados obtidos em termos de detecção de bifurcação são apresentados pelos autores através das taxas de verdadeiros positivos e de falsos positivos. A Tabela 5 mostra esses resultados.

Tabela 5 – Resultados relativos a detecção de bifurcação em imagens IVOCT obtidos por Wang (2014)

Grupos	Detecção de bifurcações (%)	
	Sensibilidade	Taxa de falsos positivos
Conjunto total	94,0	4,9
Subconjunto com 5 <i>pullbacks</i>	93,3	6,7

Ressalta-se que o desempenho deste método é fortemente depende da eficácia da segmentação, na etapa de pré-processamento e que, os autores conseguiram, através de operações em imagens relativamente simples, alcançar bons resultados.

2.5 ABORDAGENS BASEADAS EM RECONHECIMENTO DE PADRÕES

Técnicas de reconhecimento de padrões tratam da classificação ou reconhecimento automático de objetos. Na tarefa de identificação de bifurcações vasculares, estas técnicas buscam diferenciar imagens vasculares como pertencendo a região de bifurcação ou não.

O reconhecimento de padrões existentes em uma imagem é uma habilidade fundamental na análise de imagens. Pode-se definir um padrão para o caso de análise de imagens como sendo

uma descrição quantitativa ou estrutural de um objeto ou alguma entidade de interesse em uma imagem. Em geral, um padrão é formado por um ou mais descritores. Em outras palavras, um padrão é um arranjo de descritores, também chamados de características (Gonzalez e Woods, 2009).

As abordagens que utilizam técnicas de reconhecimento de padrões costumam seguir uma rotina definida de procedimentos, que envolvem: pré-processamento das imagens, segmentação, extração de características e finalmente a classificação dos dados. Estas etapas podem ser observadas na Figura 11.

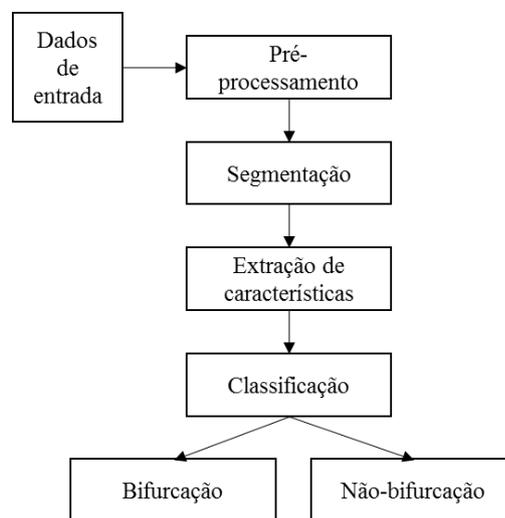


Figura 11 - Fluxograma de etapas típicas de classificação de bifurcações utilizando técnicas de reconhecimento de padrões

Observando a Figura 11, podemos notar que as primeiras etapas de uma classificação baseada em técnicas de reconhecimento de padrões são aquelas descritas na seção anterior, relativas ao processamento digital de imagens.

O pré-processamento normalmente compreende atividades como suavização da imagem e remoção de ruídos. Na segmentação, a área de interesse é evidenciada. Esta é uma etapa de grande importância, uma vez que o conjunto de características utilizado como entrada para o classificador é extraído da imagem segmentada, influenciando diretamente no desempenho do mesmo.

Para que um classificador apresente um desempenho aceitável é preciso que as características utilizadas sejam extraídas levando em consideração a base de conhecimentos relacionada ao problema que precisa ser solucionado. Normalmente, as características são extraídas a partir dos descritores da imagem.

O conjunto de características extraídas de cada imagem, forma um vetor de características. Os vetores de padrões podem ser gerados de diversas formas, a depender do tipo de imagem utilizado e do tipo de problema a ser solucionado. No entanto, pode-se destacar que a seleção de uma medida ou medidas nas quais se baseiam os componentes do vetor de padrões possui uma influência profunda no desempenho de um sistema de análise de imagens baseado nessa abordagem (Theodoridis e Koutroumbas, 2008).

Diversos tipos de características podem ser extraídos a partir de um conjunto de imagens. A literatura mostra que imagens IVUS comumente são avaliadas utilizando características de textura (Brathwaite *et al.*, 2002; Alberti *et al.*, 2011). De imagens OCT e imagens vasculares 3D, costuma-se extrair características geométricas (Mengliu e Hamarneh, 2014; Macedo *et al.*, 2015). Além destas, também são utilizados algoritmos para extração de características baseados em abordagens espaço-escala, como é o caso em Zhou *et al.* (2007). Estes algoritmos costumam ser utilizados para resolver problemas de detecção de bordas, de objetos de interesse e de junções em imagens (Lindeberg, 1999).

Após a extração das características a próxima etapa é o treinamento e teste do classificador. Dentre os artigos selecionados nesta revisão bibliográfica os classificadores mais utilizados são: *AdaBoost*, *Random Forest* e máquinas de vetores de suporte (SVM).

Boosting é uma abordagem de aprendizado de máquina baseada na ideia de criar uma regra de classificação altamente precisa a partir de uma combinação de regras relativamente fracas e imprecisas. O classificador resultante apresenta um desempenho melhor do que o de uma regra simples e é menos custoso de construir do que uma regra complexa (Schapire e

Freund, 2012). O algoritmo *AdaBoost* de Freund e Schapire (1995) foi o primeiro algoritmo prático de *boosting*, e continua sendo um dos mais utilizados e estudados, com aplicações em vários campos.

O *AdaBoost* chama um algoritmo-base em várias iterações, t , onde $t \in [1, \dots, T]$. Em cada iteração, a distribuição de pesos do conjunto de treinamento é atualizada para utilização pelo algoritmo-base. A atualização é feita de forma a aumentar os pesos dos exemplos incorretamente classificados, em confronto com os pesos dos exemplos corretamente classificados (Duarte, 2009). O Quadro 1 mostra um pseudocódigo para o algoritmo *AdaBoost*.

Quadro 1- O algoritmo *AdaBoost*. Adaptado de (Duarte, 2009).

Algoritmo <i>AdaBoost</i>	
01:	Entrada: conjunto de exemplos: $T_r = \{(x_i, y_i) \mid (i = 1, \dots, n)\}$ onde $x_i \in X$ e $y_i \in \{-1, +1\}$
02:	para $i = 1$ até n faça
03:	$D_1(i) = 1/n$ %inicialização da distribuição inicial
04:	fim para
05:	para $t = 1$ até T faça
06:	$h_t = Ab(T_r, D_t)$ %treinamento do algoritmo-base utilizando a distribuição D_t e obtenha o classificador $h_t: X \rightarrow \{-1, +1\}$
07:	$\epsilon_t = \sum_{\forall i \mid y_i \neq h_t(x_i)} D_t(i)$ %cálculo da taxa do erro ponderado do classificador
08:	$\alpha_t = \frac{1}{2} \ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$ %cálculo do poder de voto do classificador
09:	$Z_t = \sum_{i=1}^n D_t(i) e^{-\alpha_t y_i h_t(x_i)}$ %cálculo da constante de normalização
10:	para $i = 1$ até n faça
11:	$D_{t+1}(i) = D_t(i) e^{-\alpha_t y_i h_t(x_i)} / Z_t$ %atualização da distribuição de exemplos
12:	fim para
13:	fim para
14:	Saída: o classificador final
15:	$H(x) = \text{senal}(\sum_{t=1}^T \alpha_t h_t(x))$

Em Zhou *et al.* (2007), os autores propõem um método para detecção automática de bifurcações em CT torácicas utilizando *AdaBoost* em associação a filtros especialmente projetados para a extração de características. As entradas do classificador são oriundas da segmentação de imagens de CT torácica normalizadas, totalizando de 303 amostras.

O classificador foi treinado com 100 amostras positivas (com bifurcação) e 100 amostras negativas (sem bifurcação). Para o teste foram utilizados 50 exemplos negativos e 50 exemplos positivos.

Para fins de teste, os autores optaram pela técnica estatística denominada *bootstrap*, comumente utilizada quando o número de amostras analisadas é considerado pequeno. O erro médio quadrático, de acordo com os autores variou entre 3,16% e 3,63%, num intervalo de confiança de 95%.

O primeiro trabalho de detecção de bifurcações em imagens intravasculares foi desenvolvido em 2011, por Alberti *et al.* Os autores utilizaram o classificador *AdaBoost* para identificar bifurcações em imagens do tipo IVUS.

No trabalho, são analisadas imagens adquiridas de *pullbacks* das artérias coronárias de 10 diferentes pacientes. Cada sequência contém uma média de 3000 imagens, das quais 24 correspondem a bifurcações. Dois especialistas determinaram o padrão ouro, que foi utilizado no treinamento e teste do classificador.

As características utilizadas para entrada do classificador levaram em consideração o modelo das artérias e bifurcações neste tipo de imagem. Visualmente, a região do lúmen apresenta um perfil elíptico que tende a ser um pouco mais acentuado nas imagens contendo bifurcações, o que caracteriza uma excentricidade. Na Figura 12 este comportamento pode ser melhor observado.

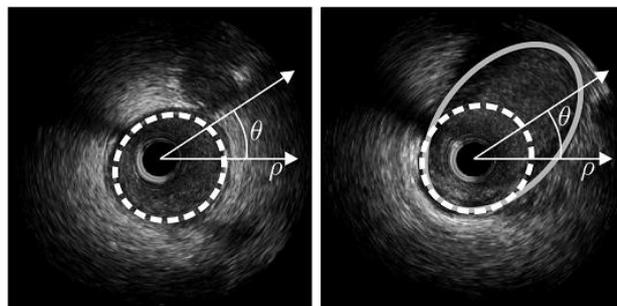


Figura 12 – Vista axial de imagens IVUS utilizadas no trabalho de Alberti *et al.* (2011) (a) sem bifurcação (b) com bifurcação. Fonte: (Alberti *et al.*, 2011)

O método desenvolvido pelos autores analisa as propriedades radiais da textura do vaso sanguíneo para detectar as seções angulares correspondentes a bifurcações e desta forma explorar as excentricidades relativas a esta modalidade de imageamento, conforme citado anteriormente.

Através de uma função de transformação, as imagens são mapeadas de coordenadas cartesianas para coordenadas polares e um conjunto de descritores de textura é definido. Para extrair informações relativas a variação radial, os autores determinam características estatísticas como desvio padrão, média e valores máximos da distância do centro à borda do vaso sanguíneo. Estas operações resultam na extração de 166 características.

Na tarefa de classificação, as características são apresentadas ao classificador *AdaBoost* que utiliza um algoritmo de árvore de decisão como regra simples. Para treinamento e teste, os autores optam pelo uso da técnica de validação cruzada chamada *leave one out* aplicada as 10 sequências de *pullbacks*.

O desempenho do classificador é avaliado em termos de acurácia, sensibilidade, especificidade, precisão e taxa de alarmes falsos (falsos positivos). Os resultados obtidos pelos autores são detalhados na Tabela 6.

Tabela 6–Desempenho do classificador *Adaboost* desenvolvido por Alberti *et al.* (2011) em termos de acurácia, sensibilidade, especificidade, precisão e taxa de falsos alarmes.

Acurácia	Sensibilidade	Especificidade	Precisão	FAR
94±4,5%	75,09±13,7%	93,51±4,71%	92,56±3,8%	6,49±4,71%

Os resultados apresentados por Alberti *et al.* (2011) são animadores quando consideramos que em uma tarefa de detecção como esta, a sensibilidade é um dos parâmetros de avaliação do classificador mais importantes, uma vez que ela expressa o número de imagens pertencentes a regiões de bifurcações que foram corretamente classificadas.

Mengliu e Hamarneh (2014) somaram esforços para desenvolver novas características, com o intuito de identificar bifurcações em imagens vasculares obtidas através de tomografia computadorizada 3D.

Os autores desenvolveram três novas características geométricas: distribuição de Von Mises-Fisher, histograma dos auto vetores e concorrência dos auto vetores. Estas novas características são utilizadas em conjunto com outras já propostas na literatura para identificação de bifurcações vasculares em imagens 3D, são elas: interseção esférica (baseada em modelo), taxa do fecho convexo, análise dos componentes principais e escala.

O conjunto de características geométricas formado foi utilizado como entrada para um classificador *Random Forest*, com o objetivo de mensurar a contribuição de cada uma das características desenvolvidas ao desempenho do classificador. O classificador gerado foi utilizado em conjuntos de imagens reais oriundas de tomografias computadorizadas torácicas e imagens 3D sintéticas.

Neste trabalho não são apresentados resultados quantitativos, mas os autores afirmam que, em termos de taxa de erro, seu trabalho apresenta um melhor desempenho que aquele desenvolvido por Zhou *et al.* (2007).

O primeiro trabalho a utilizar classificadores baseados em reconhecimento de padrões para identificar bifurcações em imagens IVOCT foi o de Macedo *et al.* (2015). Os autores propuseram um método automático para a identificação de bifurcações, compreendendo as etapas de segmentação do lúmen, e classificação das imagens como pertencentes ou não a região de bifurcação.

Para o desenvolvimento de tal tarefa, os autores analisaram imagens de *pullbacks* de nove diferentes pacientes. Os autores então dividiram as imagens em três conjuntos, levando em consideração as dificuldades de classificação presentes nas amostras.

A quantidade de imagens em cada conjunto e sua respectiva dificuldade de classificação estão detalhadas na Tabela 7. A classe de imagens com bifurcações é chamada de (BR) e a classe de imagens sem bifurcações de (NBR).

Tabela 7– Especificação dos conjuntos de dados utilizados por Macedo *et al.*(2015)

Conjunto	Problemas relativos a IV-OCT	#NBR¹	#BR²	#Total
Conjunto 1	Imagens sem dificuldade de classificação	1184	171	1355
Conjunto 2	Lesões complexas, dissecções	1256	177	1433
Conjunto 3	Lesões complexas, dissecções e presença de sangue no lúmen	1282	178	1460

¹ NBR –imagens em região sem bifurcação

² BR –imagens em região de bifurcação

Comparando as Figuras 11 e 13, nota-se uma semelhança entre o modelo de bifurcação em imagens IVUS e IVOCT. Macedo *et al.* (2015) partem da hipótese de que o corte transversal do vaso sanguíneo contém informações importantes que podem ser usadas na caracterização e detecção das bifurcações.

Alguns trabalhos encontrados na literatura extraem estas informações a partir de uma análise de textura, como em Alberti *et al.* (2011) ou a partir do contorno do lúmen, como em Wang *et al.* (2014). O método proposto por Macedo *et al.* (2015) difere destes por utilizar descritores de forma do contorno do lúmen segmentado.

Os descritores utilizados para extrair características geométricas do lúmen são: distância do centroide, circularidade, energia de curvatura, curvatura (máxima, mínima e média), variância radial, relação entre os eixos, variância da assinatura, variância setorial, diferença de área entre as imagens, área triangular média, entre outros, compondo um conjunto de 104 diferentes características.

O grande número de dimensões do conjunto de características aumenta a complexidade dos classificadores e, por conseguinte, o custo computacional. Para diminuir estes efeitos, costuma-se usar técnicas de redução de dimensionalidade, que tem por objetivo representar um conjunto de dados alta dimensão em outro espaço, de dimensão menor, procurando manter as características do conjunto.

Macedo *et al.* (2015) utilizaram duas técnicas para a tarefa de seleção de características, regressão progressiva ortogonal dos quadrados mínimos (FROLS) e múltiplos (MFROLS). As técnicas foram aplicadas em cada um dos conjuntos descritos na Tabela 7, resultando em um diferente número de características selecionadas em cada conjunto. Estes resultados podem ser observados na Tabela 8.

Tabela 8 – Número de características selecionadas por técnica e conjunto por (Macedo *et al.*, 2015)

Conjunto	Técnica	Número de características selecionadas
1	FROLS	12
1	MFROLS	5
2	FROLS	21
2	MFROLS	7
3	FROLS	28
3	MFROLS	9

Macedo *et al.* (2015) utilizaram diferentes classificadores, quais sejam: *bifurcation estimator* (BE), máquinas de vetores de suporte (SVM), *random forest* (RF) e *AdaBoost* (AB) para fins de comparação de desempenho. Para fins de treinamento e teste do classificador, os autores utilizaram a técnica de validação cruzada *cross-fold* com 10 pastas.

Cada conjunto foi treinado e testado com cada um dos classificadores implementados, o número de características de entrada foi variado utilizando o conjunto total de características e aquelas obtidas através dos métodos FROLS e MFROLS.

Os resultados apresentados por Macedo *et al.* (2015) são apresentados em termos das seguintes métricas: taxa de falsos positivos, sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* (F1) e área sob a curva ROC. Os resultados encontram-se na Tabela 9. Todos os valores apresentados estão em termos percentuais.

Tabela 9 – Resultados em termos de taxa de falsos positivos, sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e AUC obtidos por Macedo *et al.*, 2015.

Classificador	FPR	S	ACC	E	P	FA	F1	AUC
Conjunto 1								
BE+FROLS	11,08±6,37	92,40±5,21	91,96±4,76	88,92±6,37	98,29±0,97	1,60±0,93	95,19±2,99	93,03±4,25
BE+MFROLS	15,16±6,15	90,03±5,40	89,37±5,03	84,84±6,15	97,60±1,01	2,20±0,91	93,60±3,21	89,90±4,37
SVM	21,67±9,27	93,33±2,62	91,44±2,85	78,33±9,27	96,77±1,33	3,12±1,32	95,00±1,70	94,95±2,71
SVM+FROLS	21,05±8,83	94,09±2,68	92,18±2,49	78,95±8,83	96,89±1,24	3,04±1,27	95,45±1,50	94,90±2,73
SVM+MFROLS	28,66±8,53	96,12±1,33	92,99±2,03	71,34±8,53	95,87±1,21	4,14±1,23	95,99±1,17	94,38±3,40
RF	33,86±12,98	98,23±0,93	94,17±1,60	66,14±12,98	95,29±1,75	4,90±1,89	96,72±0,88	95,81±2,94
RF+FROLS	32,12±12,63	97,38±1,57	93,66±1,97	67,88±12,63	95,48±1,69	4,64±1,82	96,41±1,11	95,19±4,28
RF+MFROLS	32,06±16,78	97,98±1,27	94,10±2,24	67,94±16,78	95,52±2,27	4,64±2,45	96,67±1,24	95,44±3,45
AB	27,48±13,87	98,90±0,89	95,58±2,08	75,52±13,87	96,17±1,86	3,97±1,98	97,51±1,15	96,41±3,18
AB+FROLS	38,46±13,78	98,99±0,67	94,25±1,85	61,54±13,78	94,70±1,86	5,57±2,07	96,79±1,01	94,96±4,13
AB+MFROLS	39,05±14,03	98,99±0,67	94,17±1,67	60,95±14,03	94,63±1,87	5,66±2,10	96,75±0,90	95,11±3,98
Conjunto 2								
BE+FROLS	11,93±5,09	90,67±4,26	90,36±4,07	88,07±5,09	98,17±0,78	1,67±0,70	94,24±2,56	91,55±4,05
BE+MFROLS	18,63±7,98	89,73±5,30	88,69±4,26	81,37±7,98	97,19±1,11	2,63±1,12	93,22±2,76	86,60±4,16
SVM	25,46±8,98	91,48±2,02	89,39±1,66	74,54±8,98	96,25±1,25	3,58±1,26	93,79±1,02	92,50±2,90
SVM+FROLS	29,28±14,35	92,84±2,41	90,10±3,07	70,72±14,35	95,76±2,08	4,14±2,08	94,26±1,81	92,95±3,17
SVM+MFROLS	51,21±14,35	98,41±1,19	92,25±2,23	48,79±14,35	93,17±1,92	7,25±2,17	95,71±1,21	92,96±2,56
RF	45,26±11,86	97,29±0,67	92,04±1,52	54,74±11,86	93,87±1,49	6,37±1,64	95,55±0,83	94,17±2,61
RF+FROLS	46,34±13,94	97,29±0,94	91,90±1,99	53,66±13,94	93,74±1,78	6,53±1,95	95,47±1,09	93,98±2,81
RF+MFROLS	44,08±11,96	96,41±1,43	91,42±2,14	55,92±11,96	93,96±1,55	6,21±1,67	95,17±1,21	93,22±2,92
AB	41,27±11,36	98,25±0,98	93,37±1,91	58,73±11,36	94,43±1,48	5,81±1,61	96,30±1,05	94,73±2,86
AB+FROLS	43,01±11,56	98,89±1,08	93,72±1,89	56,99±11,56	94,25±1,46	6,05±1,60	96,51±1,05	94,71±3,14
AB+MFROLS	52,55±14,51	99,13±0,87	92,74±1,81	47,45±14,51	93,08±1,82	7,41±2,08	96,00±0,98	92,60±3,18
Conjunto 3								
BE+FROLS	12,88±5,20	90,09±5,39	89,72±4,78	87,12±5,20	98,05±0,78	1,79±0,74	93,82±3,10	90,82±4,18
BE+MFROLS	31,57±9,26	84,56±3,80	82,60±3,80	68,43±9,26	95,08±1,40	4,37±1,23	89,48±2,50	78,17±6,17
SVM	26,31±12,42	90,64±2,74	88,56±2,68	73,69±12,42	96,14±1,77	3,67±1,77	93,28±1,63	91,96±4,16
SVM+FROLS	30,39±14,78	92,67±1,58	89,86±2,52	69,61±14,78	95,68±2,03	4,21±2,05	94,14±1,44	91,55±3,56
SVM+MFROLS	31,96±15,16	92,20±1,97	89,25±1,62	68,04±15,16	95,46±2,04	4,45±2,15	93,77±0,94	92,82±3,92
RF	48,33±12,10	97,11±1,73	91,57±1,81	51,67±12,10	93,56±1,52	6,71±1,70	95,29±1,02	93,62±3,43
RF+FROLS	47,78±13,22	97,19±1,78	91,71±1,82	52,22±13,22	93,65±1,67	6,63±1,85	95,37±1,03	93,66±2,96
RF+MFROLS	52,88±16,18	98,05±0,66	91,85±2,14	47,12±16,18	93,08±2,02	7,34±2,25	95,49±1,15	93,11±4,01
AB	39,93±11,41	98,05±1,62	93,42±0,94	60,07±11,41	94,69±1,45	5,54±1,58	96,32±0,53	94,36±4,19
AB+FROLS	43,33±11,44	97,66±1,47	92,67±1,10	56,67±11,44	94,24±1,40	6,01±1,57	95,90±0,61	94,61±3,66
AB+MFROLS	50,69±12,61	97,35±1,99	91,50±1,60	49,31±12,61	93,31±1,49	7,02±1,69	95,26±0,92	92,66±4,43

BE- estimador de bifurcação; FROLS - regressão progressiva ortogonal dos quadrados mínimos; MFROLS - múltiplos FROLS; SVM – Máquinas de Vetores de Suporte; RF–Random Forest; AB–*Adaboost*; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; AUC- área sob a curva ROC.

A partir da Tabela 9, nota-se que os autores conseguiram alcançar bons valores de sensibilidade, uma das métricas mais importantes para a tarefa de classificação. Porém o

desempenho alcançado em termos de especificidade é relativamente baixo, em alguns casos chegando a menos de 50%. O mesmo ocorre com a taxa de falsos positivos, o menor valor alcançado pelos autores foi de $11,08 \pm 6,37\%$.

Os resultados de Macedo *et al.* (2015) mostram que ainda são necessários esforços no sentido de melhorar o desempenho de classificação, de forma a evitar que um especialista receba imagens sem a presença de bifurcação marcadas como imagens com bifurcação, por exemplo. Um classificador com um desempenho melhor em termos de falsos positivos poderia implicar em resultados melhores nas etapas de análise de *stent* e análise de placa.

Um resumo dos principais trabalhos levantados nesta pesquisa encontra-se na Tabela 10.

Tabela 10 – Principais artigos abordados na revisão bibliográfica

Ano	Autores	Título	Tipo de Imagem	Abordagem	Características	Classificação	Medida de desempenho	Resultados
2007	Zhou <i>et al.</i>	<i>Vascular structure segmentation and bifurcation detection</i>	CTtorácica	Reconhecimento de Padrões	Características de aparência	AdaBoost	Taxa média de erro	Taxa média de erro: 3,16~ 3,63%. Confiança: 95%.
2011	Alberti <i>et al.</i>	<i>Automatic branching detection in coronary IVUS sequences</i>	IVUS	Reconhecimento de padrões	Textura	AdaBoost	Acurácia, sensibilidade, especificidade, precisão e taxa de falsos alarmes.	ACC ¹ = 94±4,5% Sens ² = 75,09± 13,7% Espec ³ = 93,51±4,71% Precisão = 92,56±3,8% FA ⁴ = 6,49±4,71%
2012	Baboiu e Hamarneh (2012).	<i>Vascular Bifurcation detection in Scale-Space</i>	Angiografia de retina e 2D sintéticas	Processamento digital de imagens	-	Análise a partir da Matriz Hessiana	Acurácia e Sensibilidade	ACC ¹ = 95% ~97% Sens ² = 95%±0,4
2014	Mengliu e Hamarneh	<i>Bifurcation Detection in 3D vascular images using novel features and random forest</i>	CT torácica e imagens 3D sintéticas	Reconhecimento de padrões	Geométricas	Random Forest	Não apresenta	Não apresenta resultados quantitativos
2014	Wang <i>et al.</i>	<i>Fully automated side branch detection in intravascular optical coherence tomography pullback runs</i>	IVOCT	Processamento digital de imagens	-	Medidas de distância entre o centro do lúmen e as bordas	Sensibilidade e taxa de falsos positivos.	Sens ² = 94% FPR ⁵ = 4,9%
2015	Macedo <i>et al.</i>	<i>A bifurcation identifier for IV-OCT using orthogonal least squares</i>	IVOCT	Reconhecimento de padrões	Geométricas	SVM, AdaBoost, random forest e orthogonal least squares	Acurácia, sensibilidade, especificidade, precisão, taxa de falsos alarmes, <i>f-measure</i> .	ACC ¹ = 95,58±2,08% Sens ² = 98,89±0,67%

		<i>and supervised machine learning</i>						
2016	Kerkeni <i>et al.</i>	<i>Automatic bifurcation detection in coronary x-ray angiographies</i>	Raios-X	Processamento digital de imagens	-	Análise de histograma	Média de falsos positivos e de falsos negativos.	FP ⁶ = 4,4 FN ⁷ =3,2

¹ACC – acurácia; ²Sens – sensibilidade; ³Spec– especificidade; ⁴FA – taxa de falos alarmes; ⁵FP – falsos positivos; ⁶FP – falsos positivos; ⁷FN – falos negativos

FUNDAMENTAÇÃO TEÓRICA

3.1 REDES NEURAIS ARTIFICIAIS

Redes neurais artificiais (RNAs) são sistemas de computação adaptativos que tem o seu funcionamento inspirado nas características de processamento de informação encontradas nos neurônios reais e nas características de suas interconexões (Haykin, 1994).

Embora RNAs sejam abstrações das contrapartes biológicas, a ideia das RNAs não é replicar o funcionamento dos sistemas biológicos, mas fazer uso do que é conhecido sobre a funcionalidade das redes biológicas humanas para a solução de problemas complexos (Basheer e Hajmeer, 2000).

A atratividade das RNAs vem das características de processamento de informações notáveis do sistema biológico como não-linearidade, alto paralelismo, robustez, tolerância a falhas, aprendizagem, capacidade de lidar com informação imprecisa e confusa, e sua capacidade de generalizar (Jain *et al.*, 1996). Modelos artificiais com estas características são muito desejáveis porque:

- A não-linearidade permite um melhor ajuste dos dados;
- Tolerância a ruídos resulta em uma generalização precisa na presença de dados incertos e erros de medição;
- Paralelismo implica em processamento rápido e tolerância a falhas de hardware;
- A adaptabilidade permite que o sistema seja capaz de atualizar (modificar) a sua estrutura interna em resposta a mudanças no ambiente;
- A generalização permite a aplicação do modelo para dados desconhecidos.

As redes neurais são ajustadas, ou treinadas, de modo que uma determinada entrada conduza a uma saída desejada. A Figura 13 ilustra tal situação. Neste caso, a rede é ajustada, com base na comparação da saída da rede e o resultado desejado, até que a saída da rede corresponda a este resultado. Normalmente, são necessários muitos destes pares de entrada-resultado desejado para o treinamento de uma rede.

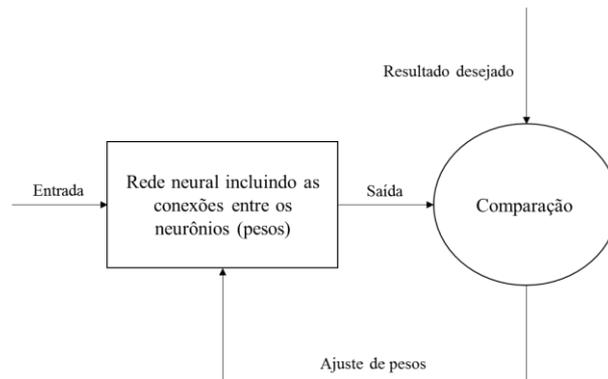


Figura 13 – Diagrama em blocos do funcionamento básico do treinamento de uma rede neural. Fonte: Adaptado de (The Mathworks).

Basicamente, todos os tipos de redes neurais apresentam a mesma unidade de processamento: um neurônio artificial, que simula o comportamento do neurônio biológico. Apesar desta inspiração, atualmente as RNAs estão muito distantes das redes neurais naturais e frequentemente, as semelhanças são mínimas. A Figura 14 mostra um modelo de neurônio artificial.

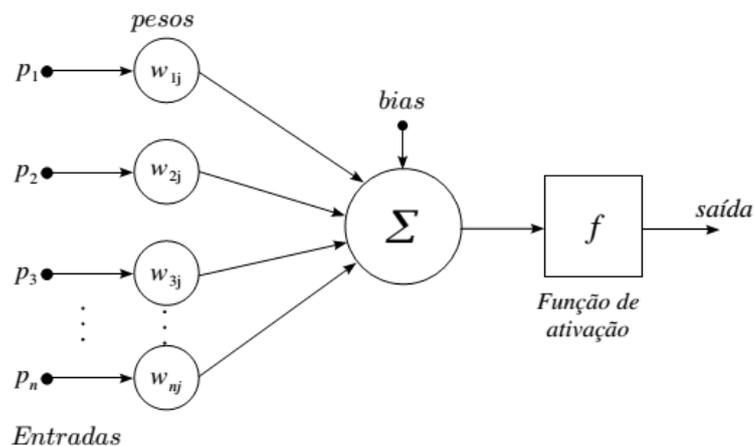


Figura 14 – Modelo de um neurônio artificial. Fonte: Adaptado de (Haykin, 1994).

Os neurônios artificiais possuem várias entradas, que correspondem às conexões sinápticas com outras unidades similares a ele, ou que recebem estímulos do exterior. Os neurônios que se comunicam com o meio externo são chamados de saída e os demais são chamados de neurônios intermediários ou escondidos. A camada composta por estes neurônios é a responsável por conferir à rede a capacidade de resolver problemas não-linearmente separáveis (Haykin, 1994).

O modelo ilustrado na Figura 14 inclui um sinal adicional *bias* (b), que favorece ou limita a possibilidade de ativação do neurônio. O processo sináptico é representado pelos pesos (ω), que amplificam cada um dos sinais recebidos. A chamada função de ativação, (f) modela a forma como o neurônio responde ao nível de excitação, limitando e definindo a saída da rede neural. A função de ativação pode ter diferentes representações, sendo as mais comuns as funções linear e sigmoide.

A capacidade de aprender é uma característica peculiar em relação aos sistemas inteligentes, biológicos ou não. Em sistemas artificiais, a aprendizagem é vista como o processo de atualização da representação interna do sistema, em resposta a estímulos externos, de modo que ele possa realizar uma tarefa específica. Isso inclui modificar a arquitetura de rede, o que envolve ajustar os pesos das conexões, cortar ou criar novas conexões e/ou alterar as regras de ativação dos neurônios (Schalkoff, 1997).

A aprendizagem das RNAs é realizada de forma iterativa, à medida que novos padrões de treinamento são apresentados a rede.

Pode-se dizer que o processo de aprendizagem foi realizado com sucesso em um sistema baseado em RNAs este for capaz de

- i. lidar com informações imprecisas, distorcidas, ruidosas, e probabilísticas sem efeito adverso perceptível na qualidade de resposta, e

- ii. generalizar a partir das tarefas que tenha aprendido para padrões desconhecidos (Jain *et al.*, 1996).

Os métodos de aprendizagem podem ser divididos em dois grupos: aprendizado supervisionado e aprendizado não-supervisionado. O aprendizado supervisionado é aquele onde um agente externo indica a resposta desejada para o padrão de entrada. Já no aprendizado não-supervisionado, também chamado de auto-organização, não existe uma indicação de resposta desejada, sendo a própria máquina responsável por determinar as saídas.

3.1.1 Regra de aprendizado por retropropagação (*backpropagation*)

O algoritmo de aprendizado mais conhecido e utilizado para treinamento de redes neurais é o algoritmo *backpropagation*. Trata-se de um algoritmo supervisionado que utiliza pares do tipo (entrada, saída desejada) para, por meio de um mecanismo de correções de erros, ajustar os pesos da rede.

O treinamento com o algoritmo *backpropagation* ocorre em duas fases: propagação direta e retropropagação. A fase da propagação direta é utilizada para definir a saída da rede para um determinado padrão de entrada. A fase de retropropagação utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos e conexões (Braga *et al.*, 2000). A Figura 15 ilustra estas fases.

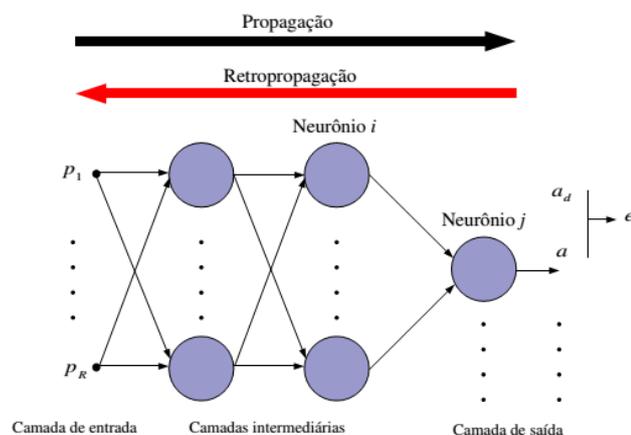


Figura 15 – Ilustração do fluxo de processamento do algoritmo *backpropagation*. Fonte: (Matsunaga, 2012)

O algoritmo *backpropagation* é baseado na regra delta proposta por Widrow-Hoff, sendo também chamado de regra delta generalizada. Este algoritmo propõe uma forma de definir o erro nos nós das camadas intermediárias, possibilitando o ajuste de seus pesos. Este ajuste de pesos é realizado utilizando-se o método do gradiente.

A função de custo a ser minimizada é uma função de erro, definida pela equação 1, define o erro total cometido pela rede.

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (1)$$

em que E é a medida do erro total, p é o número de padrões, k é o número de unidades de saída, d_i é a i -ésima saída desejada e y_i é a i -ésima saída gerada pela rede.

Embora o erro total, E , seja definido pela soma dos erros dos nós de saída para todos os padrões, Braga *et al.* (2000) considera que a minimização do erro de cada padrão individualmente, implica na minimização do erro total. Assim, o erro passa a ser definido pela equação 2.

$$E = \frac{1}{2} \sum_{j=1}^k (d_j - y_j)^2 \quad (2)$$

A regra delta sugere que a variação dos pesos seja definida de acordo com o gradiente descendente do erro com relação ao peso, conforme indicado na equação 3:

$$\Delta w_{ji} \propto - \frac{\partial E}{\partial w_{ji}} \quad (3)$$

É necessário definir como cada um dos pesos, de cada nó da rede, deve ser ajustado de forma a diminuir o erro total gerado pela rede. Utilizando a regra da cadeia tem-se que:

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad (4)$$

em que $net_j = \sum_{i=1}^n x_i w_{ji}$.

A última derivada da equação 4, $\frac{\partial net_j}{\partial w_{ji}}$, é calculada através da seguinte equação:

$$\frac{\partial net_j}{\partial w_{ji}} = \frac{\partial \sum_{i=1}^n x_i w_{ji}}{w_{ji}} = x_i \quad (5)$$

A derivada do lado direito da equação 4 é usualmente abreviada para δ_j . Esta derivada mede o erro do nó j , e também pode ser calculada através da regra da cadeia:

$$\delta_j = \frac{\partial E}{\partial net_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} \quad (6)$$

em que a derivada $\frac{\partial y_j}{\partial net_j}$, é definida por:

$$\frac{\partial y_j}{\partial net_j} = \frac{\partial f(net_j)}{\partial net_j} = f'(net_j) \quad (7)$$

Já a derivada $\frac{\partial E}{\partial y_j}$ da equação 6, utiliza o erro, e depende da camada onde o nó j se encontra. Se o nó em questão estiver na última camada, o erro pode ser atualizado por meio da equação 2, desta forma:

$$\frac{\partial E}{\partial y_j} = \frac{\partial \left(\frac{1}{2} \sum_{i=1}^k (d_i - y_i)^2 \right)}{\partial y_j} = (d_j - y_j) \quad (8)$$

sendo a equação 8 igual a regra delta original.

Substituindo os dois termos do lado direito da equação 6 pelas equações 7 e 8, obtêm-se:

$$\delta_j = (d_j - y_j) f'(net_j) \quad (9)$$

Se o nó j não se encontrar na camada de saída, utiliza-se a regra da cadeia para escrever:

$$\frac{\partial E}{\partial y_j} = \sum_{l=1}^M \frac{\partial E}{\partial net_l} \frac{\partial (net_l)}{\partial y_j} = \sum_{l=1}^M \frac{\partial E}{\partial net_l} \frac{\partial \sum_{i=1}^n w_{li} y_i}{\partial y_j} = \sum_{l=1}^M \frac{\partial E}{\partial net_l} w_{jl} \quad (10)$$

em que:

$$\sum_{l=1}^M \frac{\partial E}{\partial net_l} w_{jl} = \sum_{l=1}^M \delta_l w_{jl} \quad (11)$$

Através de substituições das equações 8 e 10 em 6, determina-se que para os nós situados nas camadas intermediárias, o erro é definido por:

$$\delta_j = f'(net_j) \sum \delta_l w_{jl} \quad (12)$$

Pode-se então generalizar a fórmula de ajuste de pesos da equação 3 como:

$$\Delta w_{ji} = \eta \delta_j x_i \quad (13)$$

ou:

$$w_{ji}(t + 1) = w_{ji}(t) + \eta \delta_j(t) x_i(t) \quad (14)$$

Caso o nó j pertença a camada de saída, o erro δ_j deve ser calculado através da equação 9, caso contrário δ_j será definido pela equação 12.

3.1.2 *Overfitting*

Um dos problemas que podem ocorrer durante o treinamento de uma rede neural é chamado de *overfitting*, ilustrado na Figura 16. A curva azul representa a função objetivo sem ruídos. Os círculos representam os pontos de uma função objetivo ruidosa. A curva preta representa a rede treinada e os círculos preenchidos com cruces (+) representam a resposta da rede aos padrões de treinamento (Hagan *et al.*, 1996).

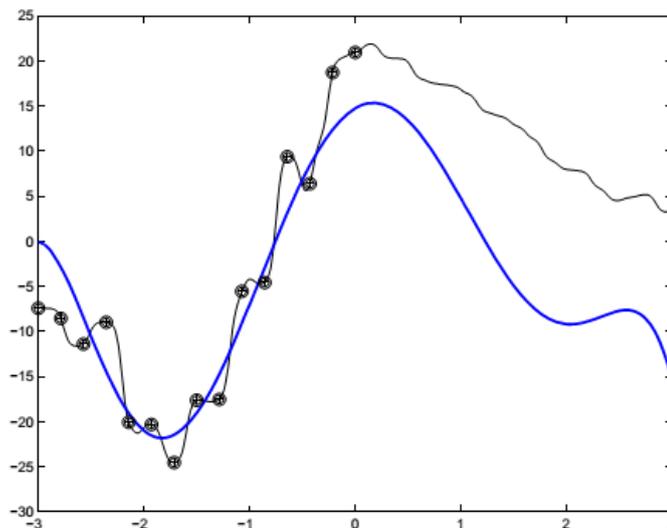


Figura 16 – Exemplo de *Overfitting*. Fonte: (Hagan *et al.*, 1996).

A resposta da rede corresponde aos pontos de treinamento, no entanto, a generalização da rede é ruim. Este problema é chamado de *overfit*. O erro no conjunto de treinamento é direcionado para um valor muito pequeno, mas quando novos dados são apresentados à rede, o erro é grande. A rede memorizou os exemplos de treinamento, mas não aprendeu a generalizar a novas situações.

Dois principais erros podem ser causados pelo *overfit*, ambos podem ser observados na Figura 16. O primeiro erro pode ser interpretado a partir da análise do intervalo $(-3,0)$. Nesta região estão concentrados todos os pontos de treinamento. A resposta da rede fica especializada no conjunto de treinamento e não funciona bem para novos valores de entrada (valores que não fazem parte do conjunto de treinamento). A rede, portanto, faz um trabalho de interpolação ruim, não conseguindo aproximar a função com precisão perto dos pontos de treinamento.

O segundo tipo de erro ocorre para valores de entrada localizados no intervalo $(0,3)$. A rede não funciona bem nesta região porque não há dados de treinamento ali. A rede está extrapolando o intervalo de entrada de dados.

Várias abordagens têm sido propostas para solucionar o problema de *overfitting*. Dentre estas, aquelas já consolidadas na literatura são conhecidas como Parada Antecipada e Regularização Bayesiana.

3.1.2.1 Parada Antecipada

O método padrão para melhorar a generalização da rede é conhecido como Parada Antecipada (Wang et al., 1994). A ideia por trás desta técnica é que a medida que o treinamento avança, a rede usa cada vez mais seus pesos, até que todos os pesos sejam totalmente utilizados quando o treinamento alcança um valor mínimo da superfície de erro. Ao aumentar o número de iterações de treinamento, aumenta-se também a complexidade da rede resultante.

Se o treinamento é interrompido antes que o mínimo seja alcançado, a rede estará usando menos parâmetros e o acontecimento do *overfit* será menos provável.

Para usar o critério de parada antecipada de maneira efetiva, precisamos utilizar uma técnica de validação cruzada. Nesta técnica, os dados disponíveis são divididos em três subconjuntos. O primeiro subconjunto é o conjunto de treinamento, utilizado para calcular o gradiente e atualizar os pesos da rede.

O segundo subconjunto é o conjunto de validação. O erro no conjunto de validação é monitorado durante todo o processo de treinamento. Este erro normalmente diminui durante a fase inicial de treinamento, assim como o erro de conjunto de treinamento. No entanto, quando a rede começa a sobreajustar os dados (*overfitting*), o erro de validação começa a subir.

Quando o erro de validação segue aumentando durante um determinado número de iterações o treinamento é interrompido. A Figura 17 mostra uma representação hipotética de erros de treinamento e validação, mostrando uma possível região de acontecimento de *overfit*.

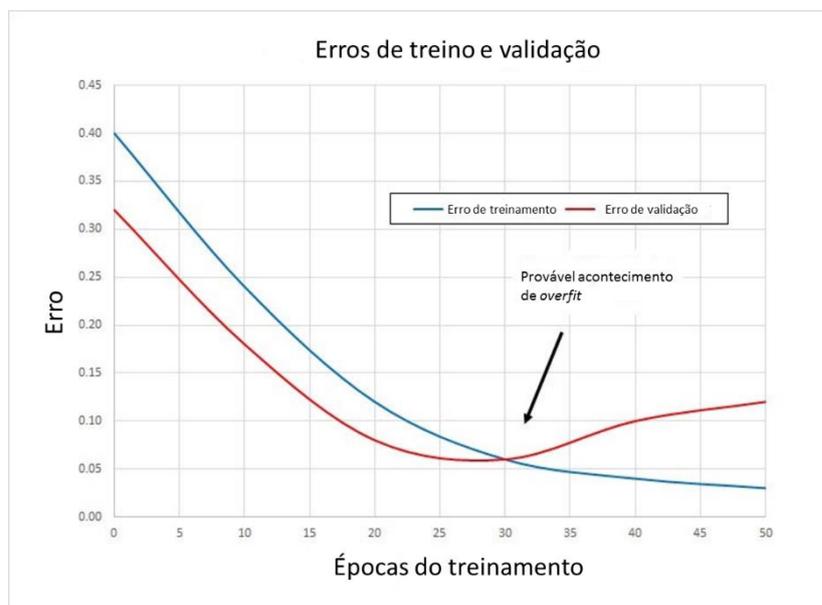


Figura 17 – Curvas hipotéticas de erros de treinamento e de validação. Fonte: Adaptado de (Hagan *et al.*, 1996)

O terceiro conjunto é o de testes. Este conjunto é utilizado para testar o desempenho do classificador gerado. O erro do conjunto de teste não é utilizado durante o treinamento, mas é usado para comparar diferentes modelos de redes.

3.1.2.2 Regularização Bayesiana

Outro método para melhorar a generalização é chamado de regularização. A regularização implica na modificação da função de desempenho, que é normalmente o erro médio quadrático da rede no conjunto de treinamento.

A função típica de desempenho usada para treinar redes neurais é a dos erros médios quadráticos da rede, indicada na equação 15.

$$F = mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - \alpha_i)^2 \quad (15)$$

É possível melhorar a generalização da rede modificando a sua função de desempenho por meio da adição de um termo que consiste na média da soma do quadrado dos pesos da rede, como mostrado na equação 16.

$$mse_{reg} = \gamma \times msw + (1 - \gamma) \times mse \quad (16)$$

em que γ é o desempenho e msw é definido pela equação 17, abaixo:

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (17)$$

Usar esta função de desempenho faz com que a rede tenha pesos menores, o que implica em uma resposta mais suave e diminui a probabilidade de acontecer o sobreajuste (Demuth *et al.*, 2013).

3.2 MÁQUINAS DE VETORES DE SUPORTE

Máquinas de Vetores de Suporte (SVM) são algoritmos de aprendizado supervisionado poderosos utilizados para tarefas de classificação ou de regressão. Os fundamentos de SVM são provenientes da Teoria de Aprendizagem Estatística desenvolvida inicialmente por Vapnik (1999).

Máquinas de vetores de suporte são máquinas de aprendizado binárias utilizadas para separar dados pertencentes a duas classes diferentes, a partir de um hiperplano que maximiza a

margem de separação entre elas. A margem representa a largura máxima entre vetores paralelos ao hiperplano que não contém nenhum ponto no seu interior. (Haykin, 2009). Estes conceitos são ilustrados na Figura 18.

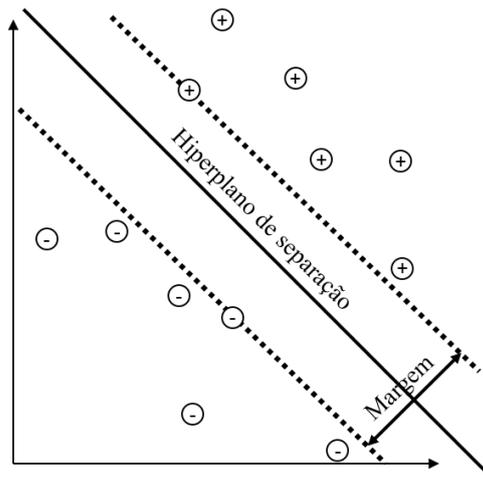


Figura 18- Ilustração dos conceitos de hiperplano de separação e margem. Fonte: adaptado de (Theodoridis e Koutroumbas, 2008)

Seja um vetor x composto pelas N características de um conjunto de treinamento. Considerando estas características pertencentes a duas diferentes classes, w_1 e w_2 , linearmente separáveis. O objetivo então é encontrar um hiperplano, descrito a partir da equação 18, que classifique corretamente todos os vetores de treinamento.

$$g(x) = w^T x + w_0 = 0 \quad (18)$$

Tal hiperplano não é único. Outros algoritmos, como o do *perceptron*, podem convergir para uma das possíveis soluções. Um hiperplano ideal é aquele que tem a mesma distância para os pontos, pertencentes às classes w_1 e w_2 , mais próximos.

Todo hiperplano é caracterizado pela sua direção (determinada por w) e pela sua posição exata no espaço (determinada por w_0). Para encontrar o hiperplano que garanta a maior margem de separação deve-se primeiro considerar que a distância de um ponto a um hiperplano é dada por:

$$z = \frac{|g(x)|}{\|w\|} \quad (19)$$

Deve-se escalar w e w_0 de forma que o valor de $g(x)$ nos pontos mais próximos das classes w_1 e w_2 seja igual a 1 para w_1 e portanto, igual a -1 para w_2 . O que é equivalente a ter uma margem que satisfaça as seguintes condições:

$$\frac{1}{\|w\|} + \frac{1}{\|w\|} = \frac{2}{\|w\|} \quad (20)$$

$$w^T x + w_0 \geq 1, \quad \forall x \in w_1 \quad (21)$$

$$w^T x + w_0 \leq -1, \quad \forall x \in w_2 \quad (22)$$

De acordo com Theodoridis e Koutroumbas (2008), para classes separáveis, os parâmetros do hiperplano que maximizam a margem são calculados através da determinação do vetor de pesos w e da polarização w_0 , tal que a expressão 23 seja minimizada satisfazendo as condições de Karush-Kuhn-Tucker (KKT):

$$J(w) = \frac{1}{2} \|w\|^2 \quad (23)$$

Para classes não separáveis, os mesmos parâmetros podem ser calculados minimizando a expressão 24, onde novas variáveis ξ_i , conhecidas como variáveis de folga são introduzidas. O objetivo agora é tornar a margem tão grande quanto possível, mas ao mesmo tempo manter o menor número de pontos com $\xi > 0$ (Theodoridis e Koutroumbas, 2008).

$$J(w, w_0, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (24)$$

O parâmetro C na expressão 24 é uma constante positiva que estabelece um compromisso entre a variável de folga e a margem (Bishop, 2006).

Máquinas de vetores de suporte utilizam *kernels* para mapear o vetor de características para um espaço de dimensão maior e, desta forma, permitir a classificação de dados não separáveis linearmente.

Segundo Theodoridis e Koutroumbas (2008), não existe nenhum método prático e eficiente para seleção do *kernel* mais apropriado para a classificação de um problema. Esta

ainda é uma questão de pesquisa sem solução, mas desafiadora. Abaixo estão listados alguns dos principais *kernels* utilizados para desenvolvimento de classificadores SVM.

- *Kernel* linear

O *kernel* linear é a função mais simples que um *kernel* pode assumir. É definido pelo produto interno $\langle x, y \rangle$ mais uma constante opcional C , conforme mostra a equação 25.

$$k(x, y) = x^T y + C \quad (25)$$

- *Kernel* polinomial

O *kernel* polinomial é não-estacionário e é recomendado para dados normalizados, sua definição encontra-se na equação 26. É possível ajustar alguns dos seus parâmetros, como a inclinação α , a constante C e o grau do d .

$$k(x, y) = (\alpha x^T y + C)^d \quad (26)$$

- *Kernel* gaussiano

O *kernel* gaussiano é uma função de base radial (GRBF). Sua descrição é encontrada na equação 27. É necessário um cuidado especial no ajuste do parâmetro γ , se superestimado, o *kernel* pode apresentar um comportamento linear. Caso subestimado, a função não terá regularização e será muito sensível aos ruídos presentes nos padrões de treinamento.

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \quad (27)$$

- *Kernel* multiquadrático

O *kernel* multiquadrático tem um custo computacional menor que o *kernel* gaussiano, podendo ser utilizado como uma alternativa em alguns casos. Sua função está definida na equação 28:

$$k(x, y) = \sqrt{\|x - y\|^2 + c^2} \quad (28)$$

Uma vez que um *kernel* é adotado, os chamados parâmetros de *kernel*, como o parâmetro de folga, C , são selecionados de modo a otimizar o desempenho do classificador. Este conjunto

de parâmetros, também conhecido como hiperparâmetros, é de fundamental importância para a capacidade de generalização do classificador (Theodoridis e Koutroumbas, 2008).

3.3 TÉCNICAS DE REDUÇÃO DE DIMENSIONALIDADE

O termo dimensionalidade se refere ao número de características utilizadas para representar determinados padrões, ou seja, a dimensão do espaço de características. A redução de dimensionalidade é uma etapa de pré-processamento comum em tarefas de classificação de dados, uma vez que remove características de pouca representatividade, impactando diretamente no desempenho do classificador (Jain *et al.*, 2000).

Esta redução pode ser feita através do uso de critérios objetivos que mensuram matematicamente o peso de cada característica na representação dos dados. Alguns destes critérios são a análise dos componentes principais, análise discriminante linear e seleção escalar de características.

3.3.1 Análise dos componentes principais

A análise de componentes principais (PCA) é considerada um dos resultados mais valiosos da álgebra linear aplicada. PCA é utilizada abundantemente em todas as formas de análise, da neurociência a computação gráfica, por se tratar de um método simples e não-paramétrico de extrair informações relevantes a partir de conjuntos de dados confusos (Shlens, 2014).

A técnica foi desenvolvida por (Pearson, 1901) e (Hotelling, 1933) e apesar de antiga, continua sendo bastante utilizada e forma a base para outras técnicas de seleção de características mais avançadas. A melhor referência moderna para o tema é (Jolliffe, 2013).

A ideia central da técnica PCA é reduzir a dimensão de um conjunto de dados em que há um grande número de variáveis inter-relacionadas, mantendo o máximo possível da variação presente no conjunto de dados.

Esta redução é obtida através da transformação dos dados originais em um novo conjunto de variáveis, os componentes principais, que apresentam as seguintes propriedades: cada componente principal é uma combinação linear de todas as variáveis originais, os componentes são não correlacionados e estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos da variação total contida nos dados.

Utilizando PCA, o número de componentes extraídos é igual ao número de variáveis analisadas. Considerando um conjunto de 'p' características de 'n' indivíduos de uma população π , as características observadas são representadas pelas variáveis $X_1, X_2, X_3 \dots X_p$ e constituem a matriz X. O resultado da análise será uma matriz Y também de dimensões 'n x p'.

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix} \quad (29)$$

A estrutura de interdependência entre as variáveis da matriz de dados é representada pela matriz de covariância S, ou pela matriz de correlação R. No entanto, o entendimento dessa estrutura através das variáveis $X_1, X_2, X_3 \dots X_p$, pode ser uma tarefa árdua. Assim, o objetivo da análise de componentes principais é representar esta estrutura através do uso das variáveis $Y_1, Y_2, Y_3, \dots Y_p$ não correlacionadas e com variâncias ordenadas, para que seja possível comparar os indivíduos da população utilizando apenas as variáveis Y_{is} que apresentam a maior variância (Varella, 2008).

A partir da matriz de dados X, pode-se estimar a matriz de covariância S, da população π . A matriz de covariância é simétrica, de ordem 'p x p'.

$$S = \begin{bmatrix} \widehat{Var}(x_1) & \widehat{Cov}(x_1x_2) & \widehat{Cov}(x_1x_3) & \dots & \widehat{Cov}(x_1x_p) \\ \widehat{Cov}(x_2x_1) & \widehat{Var}(x_2) & \widehat{Cov}(x_2x_3) & \dots & \widehat{Cov}(x_2x_p) \\ \widehat{Cov}(x_3x_1) & \widehat{Cov}(x_3x_2) & \widehat{Var}(x_3) & \dots & \widehat{Cov}(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \widehat{Cov}(x_px_1) & \widehat{Cov}(x_px_2) & \widehat{Cov}(x_px_3) & \dots & \widehat{Var}(x_p) \end{bmatrix} \quad (30)$$

De acordo com Pearson *et al.* (1901), é recomendado que os dados sejam padronizados para garantir que todas as variáveis $X_i (i = 1, 2, 3, \dots, p)$ tenham o mesmo peso na análise. A padronização pode ser feita com média zero e variância 1, ou com variância 1 e média qualquer:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s(x_j)}, i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p \quad (31)$$

$$z_{ij} = \frac{x_{ij}}{s(x_j)}, i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p \quad (32)$$

em que \bar{x}_j e $s(x_j)$ são, respectivamente, a estimativa da média e do desvio padrão da característica j :

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (33)$$

$$s(x_j) = \sqrt{\hat{V}ar(x_j)}, j = 1, 2, \dots, p \quad (34)$$

$$\sqrt{\hat{V}ar(x_j)} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n - 1} \quad (35)$$

Após a padronização, obtemos uma nova matriz de dados, Z :

$$\begin{bmatrix} z_{11} & z_{12} & z_{13} & \dots & z_{1p} \\ z_{21} & z_{22} & z_{23} & \dots & z_{2p} \\ z_{31} & z_{32} & z_{33} & \dots & z_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & z_{n3} & \dots & z_{np} \end{bmatrix} \quad (36)$$

A matriz das variáveis padronizadas z_j é igual a matriz de correlação de dados X . Para determinar os componentes principais normalmente parte-se da matriz de correlação R . Os componentes principais são determinados resolvendo-se a equação característica da matriz, isto é:

$$\det|R - \lambda I| = 0 \quad (37)$$

$$R = \begin{bmatrix} 1 & r(x_1x_2) & r(x_1x_3) & \dots & r(x_1x_p) \\ r(x_2x_1) & 1 & r(x_2x_3) & \dots & r(x_2x_p) \\ r(x_3x_1) & r(x_3x_2) & 1 & \dots & r(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(x_px_1) & r(x_px_2) & r(x_px_3) & \dots & 1 \end{bmatrix} \quad (38)$$

Sejam os autovalores, ou as raízes características da matriz R, representados por $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$, então:

$$\lambda_1 > \lambda_2 > \lambda_3 \dots \lambda_p \quad (39)$$

Para cada autovalor λ_i , existe um autovetor \tilde{a}_i :

$$\tilde{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix} \quad (40)$$

Os autovalores \tilde{a}_i são normalizados, ou seja, a soma dos quadrados dos coeficientes é igual a 1 e os mesmos são ortogonais entre si. Sendo \tilde{a}_i o autovetor correspondente ao autovalor λ_i , então o i -ésimo componente principal é dado por:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (41)$$

A contribuição C_i de cada componente principal Y_i é expressa em porcentagem. É calculada dividindo-se a variância de Y_i pela variância total. Representa a proporção de variância total explicada pelo componente principal Y_i (Varella, 2008).

A importância de um componente principal é avaliada por meio de sua contribuição, isto é, pela proporção de variância total explicada pelo componente. A soma dos primeiros k autovalores representa a proporção de informação retida na redução de p para k dimensões.

Com essa informação pode-se decidir quantos componentes serão utilizados na análise, isto é, quantos componentes serão utilizados para diferenciar os indivíduos. Não existe um modelo estatístico que ajude nesta decisão. Para aplicações em diversas áreas do conhecimento

o número de componentes utilizados tem sido aquele que acumula 70% ou mais de proporção da variância total (Varella, 2008).

3.3.2 Análise discriminante linear

A análise discriminante linear (LDA), é comumente utilizada como técnica de redução de dimensionalidade na etapa de pré-processamento em aplicações de reconhecimento de padrões e aprendizado de máquina. O objetivo é projetar um conjunto de dados em um espaço de dimensões menores, maximizando a separabilidade entre classes.

Formulado por Fisher(1936), o discriminante também tem alguns usos práticos como classificador. O discriminante linear original foi descrito para um problema de duas classes. Mais tarde foi generalizado para problemas multi-classes por Rao (1948).

A abordagem LDA é muito semelhante a PCA, mas além de buscar maximizar a variância dos dados, o LDA também almeja maximizar a separação entre várias classes existentes no conjunto de dados. Ou seja, a abordagem LDA visa encontrar uma representação das características na qual a distância dentro das classes seja mínima enquanto que a distância entre as classes é maximizada (Duda *et al.*, 2000).

Pode-se reduzir a dimensionalidade de um vetor d -dimensional a apenas uma dimensão meramente projetando os dados d -dimensionais em uma linha. No entanto, ainda que as amostras originais sejam bem separadas, compactar dados simplesmente os projetando em uma linha arbitrária pode misturar as amostras de todas as classes, podendo afetar negativamente o desempenho do classificador. Contudo, movendo esta linha, pode ser possível encontrar uma orientação adequada para que as amostras projetadas sejam bem separadas. Este é o objetivo da análise discriminante clássica (Duda *et al.*, 2000).

Considere um vetor \mathbf{x} no espaço d -dimensional, assumindo que seus pontos são originários de duas classes distintas. Deseja-se comprimir as informações de \mathbf{x} em um número menor de características (neste exemplo apenas uma). Esta operação corresponde a gerar uma

característica y como uma combinação linear dos componentes de x (Theodoridis e Koutroumbas, 2008).

Para alcançar este objetivo, precisa-se encontrar a orientação w no espaço d -dimensional, de forma que as duas classes sejam separadas. A projeção de x em w será definida através da equação:

$$y = w^T x \quad (42)$$

Em seguida, adota-se a relação discriminante de Fisher (FDR), que quantifica a capacidade de separabilidade de características individuais dos componentes de uma classe.

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (43)$$

em que μ_1 e μ_2 são as médias e σ_1 e σ_2 são os desvios padrões das classes ω_1 e ω_2 respectivamente, depois da projeção em w . O maior valor da razão FDR representa a característica que melhor separa duas classes distintas (Theodoridis e Koutroumbas, 2008).

Usando a definição mostrada na equação 42, pode-se afirmar que:

$$\mu_i = w^T \mu_i \quad i = 1, 2 \quad (44)$$

em que μ_i , $i = 1, 2$, é o valor médio dos dados em w_i no espaço d -dimensional.

Sejam S_w e S_b as matrizes de covariância entre classes e interclasses respectivamente, dadas por:

$$S_w = \sum_{i=1}^M P_i (x - \mu_i)(x - \mu_i)^T \quad (45)$$

$$S_b = \sum_{i=1}^M P_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T \quad (46)$$

em que P_i é a probabilidade da classe w_i e μ_0 é o vetor médio global. Assumindo que as classes são equiprováveis pode-se demonstrar que:

$$(\mu_1 - \mu_2)^2 = w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad w \propto w^T S_b w \quad (47)$$

Analisando o denominador da equação 41, pode-se concluir que:

$$\sigma_1^2 = E[(y - \mu_i)^2] = E[w^T(x - \mu_i)(x - \mu_i)^T w] \quad (48)$$

$$\sigma_1^2 + \sigma_2^2 \propto w^T S_w w \quad (49)$$

Combinando as equações 47, 45 e 41, temos que a orientação ótima pode ser obtida maximizando o critério de Fisher em relação a w , assumindo que w seja invertível:

$$FDR = \frac{w^T S_b w}{w^T S_w w} \quad (50)$$

$$w = S^{-1}(\mu_1 - \mu_2) \quad (51)$$

De acordo com Theodoridis e Koutroumbas (2008), para problemas envolvendo mais de duas classes, adota-se o critério J_3 de separabilidade. Dado por:

$$J_3 = \text{traço} \left(\frac{S_b}{S_w} \right) \quad (52)$$

A fim de maximizar o critério J_3 , uma matriz U , de dimensão d_{xl} , é formada pelos auto vetores que correspondem aos maiores autovalores da matriz $S_w^{-1} S_b$. A redução de dimensionalidade é obtida considerando $l < d$. Quanto menos autovetores forem utilizados na formação da matriz U , maior a redução de dimensionalidade alcançada (Theodoridis e Koutroumbas, 2008).

3.3.3 Seleção escalar de características

O método de seleção escalar de características (Theodoridis e Koutroumbas, 2008) é utilizado em reconhecimento de padrões para selecionar as variáveis que melhor separam um conjunto de amostras em n classes distintas.

Nesta técnica, as variáveis são consideradas de forma individual, ou seja, como escalares. Uma medida de separabilidade de classes é o critério utilizado nesse processo de seleção, em seguida esses valores são ordenados e as melhores características são selecionadas.

De acordo com Theodoridis e Koutroumbas (2008), a razão de discriminação de Fisher FDR pode ser utilizada para validar o grau de separação individual de características, conforme definida na equação 41.

Para selecionar a primeira característica, calcula-se a medida FDR ($C(k)$) para todas as características, ordenam-se as mesmas em ordem decrescente, e seleciona-se a característica com o maior resultado. Define-se então, o coeficiente de correlação cruzada entre duas características i e j como:

$$\rho_{ij} = \frac{\sum_{n=1}^N x_{ni}x_{nj}}{\sqrt{\sum_{n=1}^N x_{ni}^2 \sum_{n=1}^N x_{nj}^2}} \quad (53)$$

em que, $n = 1, 2, \dots, N$ e $k = 1, 2, \dots, m$, x_{nk} é a k -ésima característica do n -ésimo padrão. O parâmetro ρ_{ij} é obtido entre a primeira característica selecionada (x_{i1}) e as $m - 1$ restantes.

A segunda característica selecionada (x_{i2}) é obtida de forma que:

$$i_2 = \arg \max_j \{ \alpha_1 C(j) - \alpha_2 |\rho_{i_1 j}| \} \text{ para todo } j \neq i \quad (54)$$

em que, α_1 e α_2 são fatores que regulam a importância de cada termo. Neste trabalho os valores determinados foram $\alpha_1 = \alpha_2 = 0.5$.

As demais características (x_{ik} , onde $k = 3, \dots, m$) são selecionadas de forma que:

$$i_k = \arg \max_j \left\{ \alpha_1 C(j) - \frac{\alpha_2}{k-1} \sum_{r=1}^{k-1} |\rho_{i_r j}| \right\} \quad (55)$$

MATERIAIS E MÉTODOS

Este capítulo descreve os materiais e as diferentes metodologias utilizadas nesta dissertação. As metodologias foram desenvolvidas com base nos objetivos específicos que norteiam o trabalho.

4.1 CONJUNTO DE DADOS

O material utilizado neste trabalho para realização da tarefa de classificação de bifurcações em imagens coronarianas é aquele disponibilizado por Macedo *et al.* (2015). Tratando-se, portanto, do conjunto de características geométricas extraídas a partir do lúmen segmentado de imagens IVOCT.

Macedo *et al.* (2015) analisaram imagens de 9 *pullbacks* de 9 diferentes pacientes. As imagens foram divididas em três diferentes conjuntos, levando em consideração o nível de dificuldade de classificação, conforme exemplos apresentados na Figura 19.

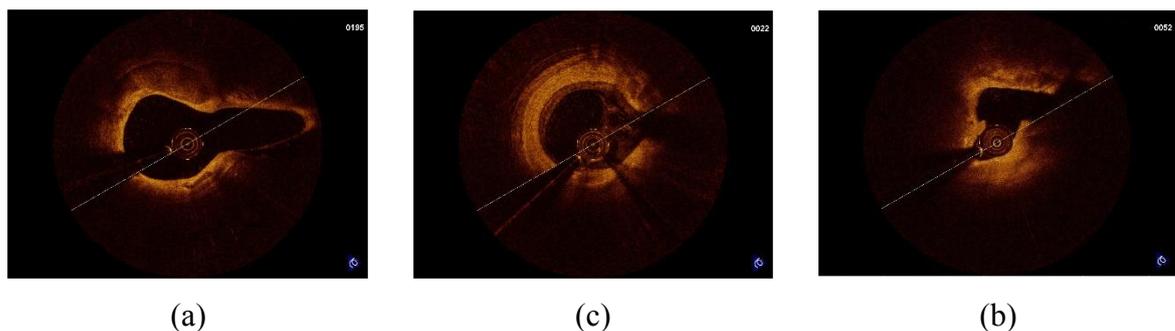


Figura 19 – Exemplos de imagens (a) conjunto 1: com bifurcação normal (b) conjunto 2: bifurcação com sangue residual (c) conjunto 3: bifurcação com placa complexa.

A constituição dos referidos conjuntos foi apresentada, previamente, na Tabela 7 do capítulo 2.

Os dados (características geométricas do lúmen segmentado) utilizados no presente trabalho são derivados destas imagens. Os descritores utilizados para extrair características geométricas do contorno do lúmen segmentado foram:

1. Distância ao centroide
2. Circularidade
3. Energia de curvatura
4. Curvatura (máxima, mínima, média)
5. Variância radial
6. Relação entre os eixos
7. Variância da assinatura
8. Variância setorial
9. Diferença de área entre imagens
10. Área triangular média.
11. Outros descritores foram determinados por meio de derivações e combinações destes, totalizando 104 características diferentes.

4.2 MÉTODOS

4.2.1 Balanceamento dos conjuntos

O desbalanceamento de classes pode ser prejudicial dependendo do problema em questão, resultando em uma tendência de melhores resultados para as classes majoritárias em detrimento das minoritárias. Entretanto, em muitos casos, o que importa é ter um bom desempenho para as classes minoritárias.

Como pode-se observar na Tabela 11, os conjuntos formados por Macedo et al. (2015) são desbalanceados. A classe BR, correspondente a imagens com bifurcação, representa apenas cerca de 14% dos dados em cada um dos conjuntos, e a classe majoritária, NBR corresponde a aproximadamente 85%.

Tabela 11 – Percentual de amostras das classes BR e NBR em cada conjunto.

Conjunto	#NBR ¹	#BR ²	%NBR ¹	%BR ²
Conjunto 1	1184	171	85,56%	14,44%
Conjunto 2	1256	177	85,91%	14,09%
Conjunto 3	1282	178	86,12%	13,88%

¹ NBR –imagens em região sem bifurcação

² BR –imagens em região de bifurcação

Após experimentos iniciais, identificou-se a necessidade de realizar um balanceamento de dados intraconjunto, uma vez que o objetivo principal deste trabalho é identificar imagens com a presença de bifurcação (pertencentes a classe BR).

Dentre as abordagens para tratamento de dados desbalanceados, podemos destacar as técnicas: reamostragem aleatória, SMOTE (Chawla *et al.*, 2002), seleção unilateral (Kubat e Matwin, 1997) e sobreamostragem baseada em agrupamento (Yen e Lee, 2009).

Algumas das principais técnicas de balanceamento de dados introduzem padrões sintéticos ao conjunto original. Tal fato, associado ao objetivo de realizar uma análise comparativa do presente estudo com o implementado por Macedo *et al.*(2015), nos levou a optar por uma abordagem que preservasse os conjuntos de treinamento e teste aos moldes do que foi utilizado previamente, a sobreamostragem.

Desta forma, as imagens da classe NBR foram divididas de forma aleatória, proporcionalmente, em sete partes. As imagens foram BR replicadas e então adicionadas a cada uma destas partes, formando 7 subconjuntos balanceados. A Figura 20 mostra como foi feito o balanceamento dos conjuntos.

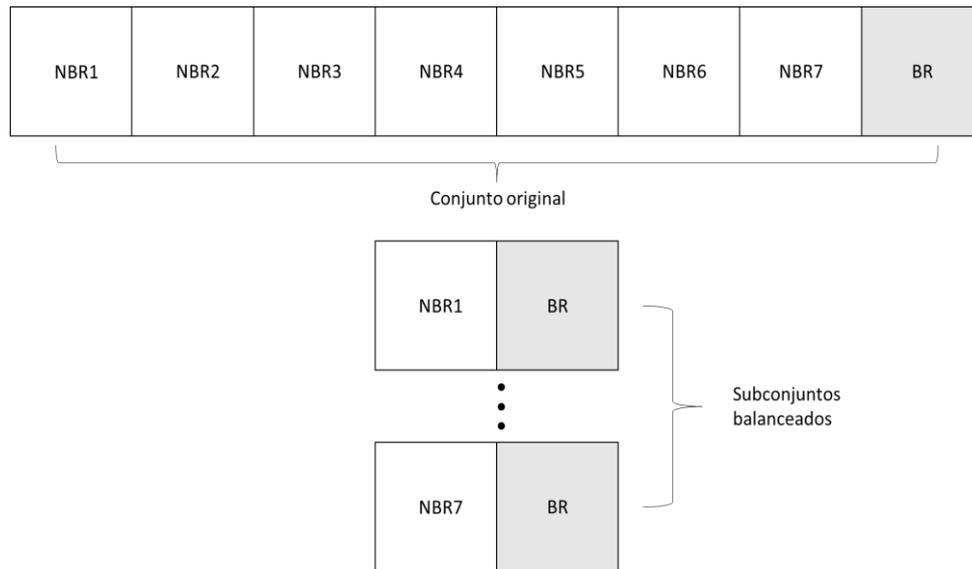


Figura 20 - Ilustração do processo de balanceamento dos conjuntos compilados por Macedo *et al.* (2015)

Cada conjunto definido por Macedo *et al.* (2015), de acordo com o nível de dificuldade, foi tratado individualmente. Em outras palavras, os conjuntos 1, 2 e 3, foram divididos em 7 diferentes subconjuntos cada. Na etapa de classificação das amostras, cada um dos subconjuntos gerados é utilizado e a resposta final do classificador é a média dos resultados obtidos nos subconjuntos.

4.3 DIVISÃO DO BANCO DE DADOS PARA TREINAMENTO E TESTE

A separação de dados em conjuntos de teste e treinamento é uma parte importante da avaliação de modelos de classificadores. Normalmente, a maior parte dos dados é usada para treinamento e uma parte menor dos dados é usada para teste. Desta forma, pode-se avaliar o desempenho do classificador quando dados não empregados no treinamento são apresentados. Neste trabalho, a validação cruzada *k-fold* é utilizada para separação dos conjuntos.

No *k-fold*, o conjunto de dados é dividido aleatoriamente em subconjuntos mutuamente exclusivos de tamanhos aproximadamente iguais. O classificador é treinado e testado *k* vezes (Kohavi, 1995). A Figura 21 detalha como são realizadas as divisões dos grupos em cada experimento, segundo a técnica *10-fold cross validation*.

Cada subconjunto balanceado (conforme descrito na 4.2.1) é dividido através da técnica *10-fold-cross-validation* e utilizado para treinamento e teste dos classificadores e ao final do processo, um resultado médio é calculado para a avaliação do desempenho do classificador.

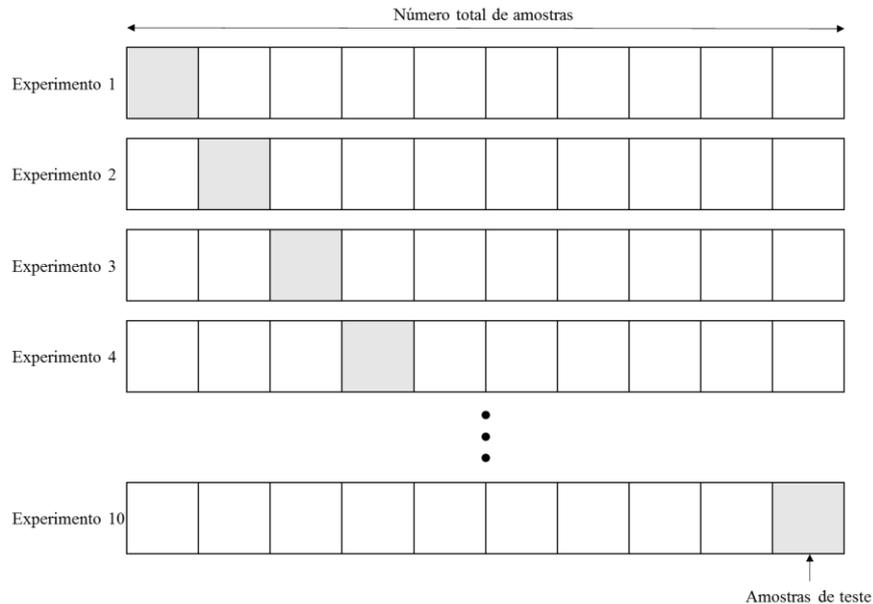


Figura 21 – Ilustração da técnica K-fold Cross Validation, com $k=10$.

4.4 CLASSIFICAÇÃO DOS DADOS

Para ambos classificadores, Redes Neurais Artificiais e Máquinas de Vetores de Suporte, o desenvolvimento da metodologia foi dividido em quatro fases. As fases se distinguem pelo conjunto de dados utilizado como entrada para os classificadores implementados. A Figura 22 ilustra as quatro fases com respeito ao conjunto de dados.

A primeira fase é caracterizada por utilizar o conjunto de dados originais com o total de 104 características. A segunda fase, por utilizar extração de características com o método PCA, selecionando as melhores 5, 6, 7, ..., 20 características. A terceira fase, análoga a segunda, porém com o auxílio do método LDA. E finalmente, a quarta fase, análoga as anteriores, porém utilizando a técnica de seleção escalar de características.

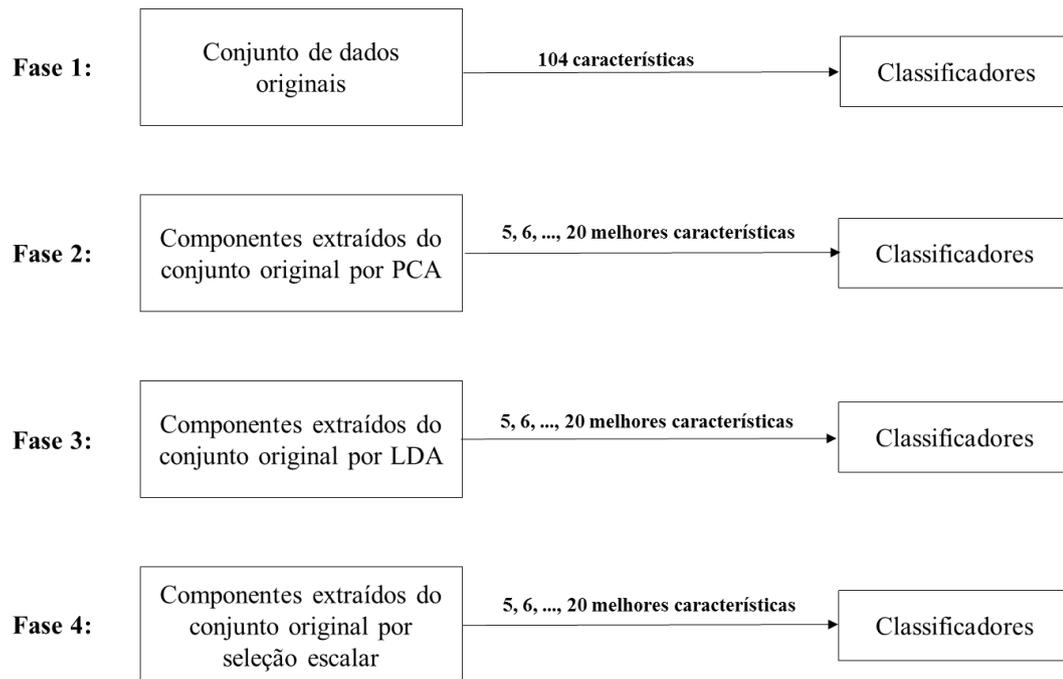


Figura 22- Quatro fases da classificação de dados e os diferentes conjuntos utilizados como entrada para os classificadores implementados.

Os detalhes do desenvolvimento dos classificadores utilizando cada uma das técnicas propostas são discutidos a seguir.

4.4.1 Redes Neurais Artificiais

As arquiteturas das redes neurais são caracterizadas pelo número de neurônios na camada de entrada, número de neurônios nas camadas intermediárias e pelo número de neurônios na camada de saída.

O número de neurônios na camada de entrada é igual ao número de coordenadas do vetor de características, ou seja, igual ao número de características utilizadas. Neste trabalho, conforme representado na Figura 22, o vetor de características é formado por 104 coordenadas na Fase 1 e nas Fases 2, 3 e 4, varia de 5 a 20 características.

Para o desenvolvimento da tarefa de classificação, são utilizadas duas camadas escondidas, hl_1 e hl_2 . Neste trabalho, hl_1 e hl_2 são sempre valores iguais e podem assumir três combinações diferentes: 5-5, 8-8 e 10-10. Na camada de saída existe apenas um neurônio. A Figura 23 mostra um diagrama em blocos genérico das arquiteturas utilizadas neste trabalho.

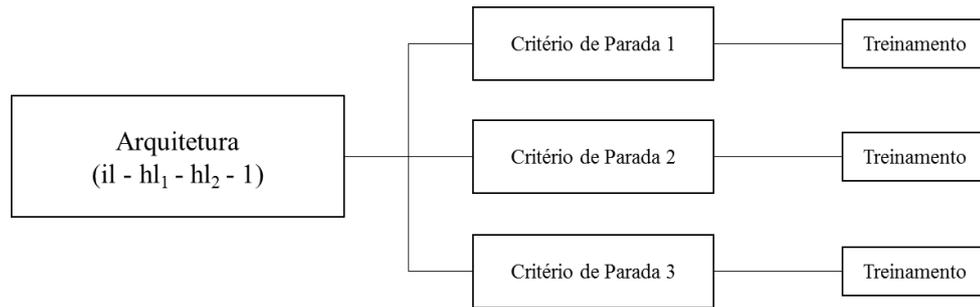


Figura 23 – Diagrama em blocos genérico utilizado em cada fase do trabalho. il representa o número de neurônios na camada de entrada (igual ao número de características, podendo valer 104 e o intervalo de 5 a 20), hl_1 e hl_2 representam os números de neurônios nas camadas escondidas (podendo valer 5-5, 8-8 e 10-10).

Cada arquitetura é treinada com os critérios de parada erro médio quadrático, parada antecipada e regularização, representados na Figura 23 como critérios de parada 1, 2 e 3, respectivamente. A Figura 24 mostra um exemplo de arquitetura utilizada quando o vetor de características possui 104 componentes, ou seja, quando o conjunto de dados original é utilizado. Na Tabela 12, são detalhados os passos metodológicos implementados em cada uma das arquiteturas de redes neurais desenvolvidas.

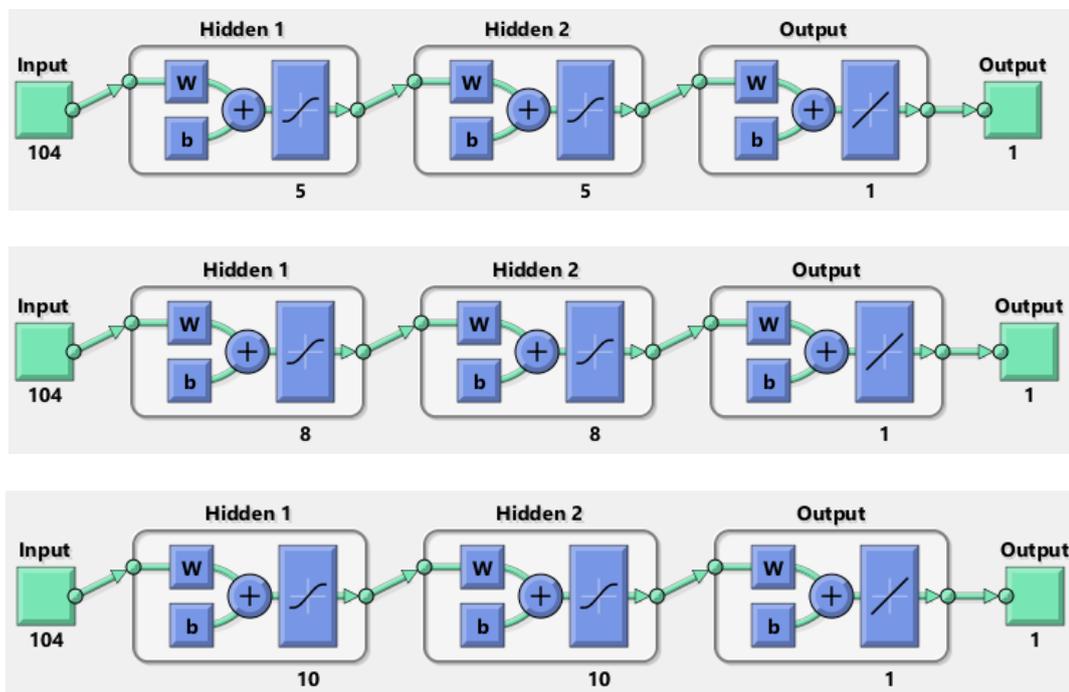


Figura 24 – Arquiteturas utilizadas para o treinamento dos classificadores considerando o conjunto total de características como entrada

Tabela 12 – Passos metodológicos a serem implementados em cada uma das arquiteturas de redes neurais utilizadas

Critério de Parada da rede neural	Metodologias de treinamento e teste	Método de seleção de características
Erro médio quadrático	10-fold-cross-validation	Conjunto de dados completo
		PCA – melhores 5 a 20 características
		LDA - melhores 5 a 20 características
		Seleção escalar – melhores 5 a 20 características
Parada Antecipada	Treinamento: 1/3 do conjunto Validação: 1/3 do conjunto Teste: 1/3 do conjunto	Conjunto de dados completo
		PCA – melhores 5 a 20 características
		LDA - melhores 5 a 20 características
		Seleção escalar – melhores 5 a 20 características
Regularização	10-fold-cross-validation	Conjunto de dados completo
		PCA – melhores 5 a 20 características
		LDA - melhores 5 a 20 características
		Seleção escalar – melhores 5 a 20 características

Cada rede neural tem sua arquitetura variada devido ao número de características na entrada (5, 6, 7, ... 20, 104) e ao número de neurônios nas camadas escondidas (5, 8 ou 10). Além disso, cada rede é treinada com diferentes critérios de parada: erro médio quadrático, parada antecipada e regularização, totalizando 153 diferentes classificadores para a tarefa de identificação de bifurcações em imagens IVOCT com redes neurais.

A seguir serão detalhadas as particularidades de cada um dos diferentes critérios de parada utilizados.

4.4.1.1 Erro médio quadrático

Quando utilizando o critério de parada do erro médio quadrático, o treinamento é finalizado quando quaisquer das seguintes condições ocorrerem:

- a) O mínimo erro quadrático de 10^{-6} é alcançado;

- b) O número máximo de 5000 épocas (iterações) é atingido.

4.4.1.2 Parada Antecipada

Para implementar o critério Parada Antecipada, os conjuntos de treinamento validação e teste foram divididos da seguinte forma: Conjunto de treinamento: 50% dos dados, conjunto de validação: 25% dos dados, conjunto de teste: 25% dos dados.

O treinamento com parada antecipada é finalizado quando uma das seguintes condições é atingida:

- a) O número máximo de 1000 épocas é atingido;
- b) O erro médio quadrático de 10^{-6} é alcançado;
- c) O erro de validação aumentou por mais de 10 épocas consecutivas.

4.4.1.3 Regularização

Os critérios de parada do treinamento quando utilizando a regularização são os mesmos do erro médio quadrático, o treinamento é finalizado quando quaisquer das seguintes condições ocorrerem:

- a) O mínimo erro quadrático de 10^{-6} é alcançado;
- b) O número máximo de 5000 épocas (iterações) é atingido.

4.4.2 Máquinas de Vetores de Suporte

Para implementação dos classificadores SVM, é preciso primeiramente determinar quais serão os *kernels* utilizados. Não existem técnicas estabelecidas para seleção de *kernels* apropriados para uma determinada tarefa de classificação. No entanto, existem diversas funções de *kernel* disponíveis para o treinamento de um classificador SVM.

Neste trabalho, foram escolhidos os *kernels* GRBF (*gaussian radial basis function*) e polinomial (P) após uma análise preliminar dos resultados obtidos quando diferentes funções

de mapeamento foram utilizadas para gerar classificadores. O processo que levou a escolha dos *kernels* é ilustrado na Figura 25.

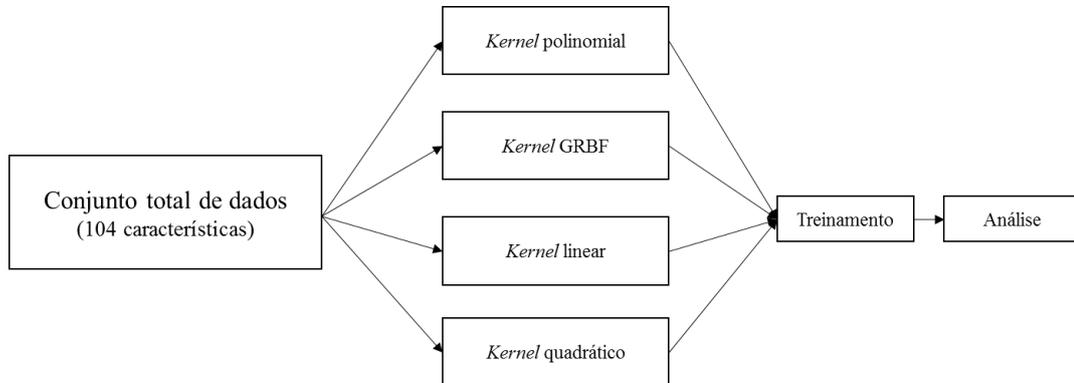


Figura 25- Metodologia utilizada para a escolha de *kernels* para a classificação

O conjunto total de dados foi usado como entrada para diferentes classificadores com diferentes *kernels*. Os classificadores foram treinados e, posteriormente, seus desempenhos foram analisados. Aqueles *kernels* responsáveis pelos melhores desempenhos foram selecionados para as próximas etapas do trabalho.

Com a escolha dos *kernels* mais apropriados para a solução do problema, passamos para a etapa de otimização dos hiperparâmetros. Para isso, a técnica do *grid-search*, ou busca exaustiva, é utilizada.

Dois hiperparâmetros são ajustados a fim de determinar os classificadores com melhores desempenhos, a constante de regularização C e o hiperparâmetro γ do *kernel*. A busca exaustiva seleciona um conjunto finito de valores para cada um dos parâmetros, neste caso:

$$c = \{1, 2, 3, 4\} \quad (56)$$

$$\gamma = \{2^{-5}, 2^{-3}, 2^{-1}, 2^1, 2^3, 2^5\} \quad (57)$$

Em seguida, um classificador SVM é treinado para cada par possível (C, γ) . Este processo está representado na Figura 26. Os *kernels* utilizados nestes classificadores são aqueles anteriormente mencionados, P e GRBF.

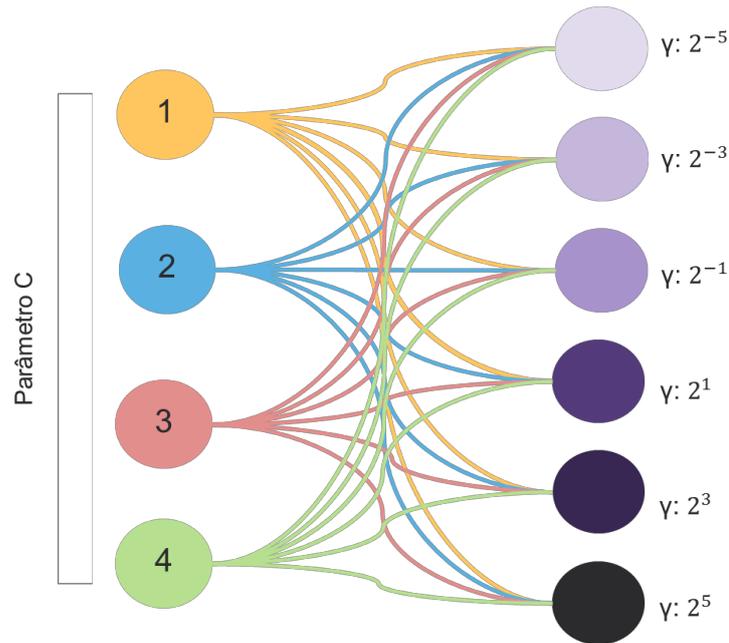


Figura 26 - Ilustração do processo de busca exaustiva

Os desempenhos dos classificadores gerados são avaliados em um conjunto de teste e, finalmente, o algoritmo de busca indica as configurações que obtiveram o melhor desempenho no procedimento de avaliação.

O parâmetro γ varia exponencialmente, portanto é preciso delimitar os pares (C, γ) que apresentam os melhores desempenhos. Para isso, é feita uma nova busca exaustiva, dentro do intervalo avaliado como ótimo. Estes valores não são pré-definidos, uma vez que é esperado que haja uma diferença de comportamento causada pelos conjuntos de dados que compõem a entrada dos classificadores (conjuntos 1, 2 e 3 definidos na seção Materiais) e pelos *kernels* utilizados, porém, o incremento de γ na nova busca é de 0,2.

A Tabela 13 mostra os passos metodológicos implementados para cada classificador delimitado através da técnica busca exaustiva.

Tabela 13- Passos metodológicos a serem implementados em cada dos classificadores obtidos através da busca exaustiva

<i>Kernels</i>	Metodologias de treinamento e teste	Método de seleção de características
GRBF	10-fold-cross-validation	Conjunto de dados completo
		PCA – melhores 5 a 20 características
		LDA - melhores 5 a 20 características
		Seleção escalar – melhores 5 a 20 características
Polinomial	10-fold-cross-validation	Conjunto de dados completo
		PCA – melhores 5 a 20 características
		LDA - melhores 5 a 20 características
		Seleção escalar – melhores 5 a 20 características

Ao todo, foram treinados 336 classificadores utilizando Máquinas de Vetores de Suporte e o *kernel* GRBF e 216 classificadores utilizando *kernel* polinomial, totalizando 552 classificadores SVM.

4.5 AMBIENTE DE IMPLEMENTAÇÃO

O desenvolvimento da metodologia proposta foi realizado no Laboratório de Reconhecimento de Padrões do Centro de Tecnologia Eletrônica e da Informação – CETELI da Universidade Federal do Amazonas – UFAM, utilizando-se o ambiente de desenvolvimento Matlab® (The MathWorks, Inc. Software) versão 2014. O computador utilizado foi um computador com processador Intel Core i7-4790 de 3.60 GHz e 16,0 GBytes de memória RAM.

RESULTADOS

Neste capítulo são apresentados os resultados obtidos, fazendo uso de Redes Neurais Artificiais e Máquinas de Vetores de Suporte para a classificação de imagens IVOCT como pertencentes ou não a região de bifurcação. Os resultados de cada tipo de classificador são mostrados separadamente.

Conforme descrito na seção Materiais e Métodos, o conjunto total de imagens encontra-se dividido em três subconjuntos, de acordo com o nível de dificuldade de classificação. Os resultados são apresentados para cada um destes três subconjuntos.

Inicialmente são apresentados os resultados obtidos na primeira fase da pesquisa, que utiliza o conjunto de características original, totalizando 104 padrões para cada imagem. Posteriormente, são apresentados os resultados utilizando as técnicas PCA, LDA e seleção escalar de características, divididos em segunda, terceira e quarta fase da pesquisa.

Os resultados são apresentados em forma de gráficos e tabelas, destacando os seguintes parâmetros: taxa de falsos positivos, sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* (F1) e área sob a curva ROC.

Ao todo foram treinados 704 classificadores RNA e SVM. Neste capítulo encontram-se listados apenas os resultados dos classificadores com melhor desempenho.

5.1 CLASSIFICAÇÃO COM REDES NEURAIAS

5.1.1 Primeira fase

A primeira fase é aquela em que o classificador é treinado utilizando o conjunto original de dados, as 104 características compiladas por Macedo *et al.* (2015).

5.1.1.1 Redes neurais com erro médio quadrático

A Figura 27 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios nas camadas escondidas varia entre 5, 8 e 10 e o critério de parada é o Erro Médio Quadrático (MSE). Pode-se observar que para os três conjuntos, a acurácia na classificação é maior quando o número de neurônios nas camadas escondidas é igual a 10.

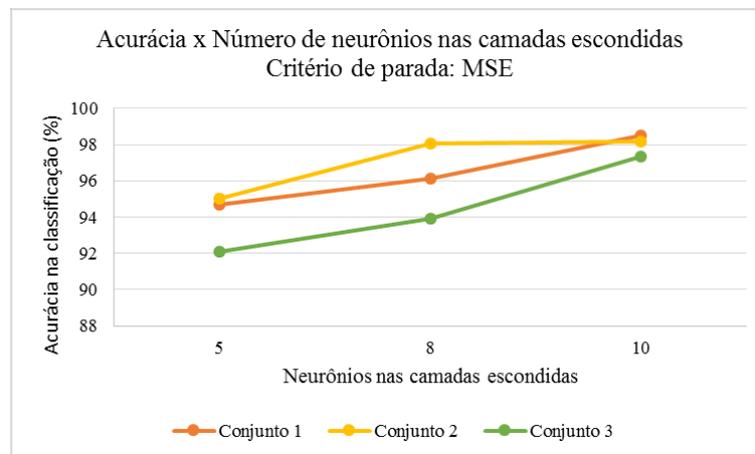


Figura 27- Acurácia x Números de neurônios nas camadas escondidas com critério de parada erro médio quadrático

Na Tabela 14, são mostrados valores detalhados dos melhores classificadores projetados para cada conjunto, com o critério de parada erro médio quadrático, utilizando as 104 características como entrada.

Tabela 14 – Melhores resultados obtidos utilizando redes neurais com erro médio quadrático e as 104 características como entrada para cada conjunto.

Conjunto	N	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	10	1,09 ± 0,73	98,07 ± 1,36	98,48 ± 0,9	98,91 ± 0,73	98,95 ± 0,71	1,09 ± 0,73	98,43 ± 0,96	0,99 ± 0,01
C2	10	1,28 ± 1,49	97,59 ± 2,05	98,15 ± 1,75	98,72 ± 1,49	98,67 ± 1,56	1,3 ± 1,51	98,08 ± 1,83	0,99 ± 0,01
C3	10	2,24 ± 1,1	96,88 ± 2,94	97,32 ± 1,83	97,76 ± 1,1	97,69 ± 1,17	2,36 ± 1,17	97,18 ± 2,05	0,98 ± 0,01

N- número de neurônios nas camadas intermediárias; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.1.2 Redes neurais com parada antecipada

A Figura 28 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios nas camadas escondidas varia entre 5, 8 e 10 e o critério de parada é a Parada Antecipada (ES), utilizando o conjunto total de

características. Observa-se que para os conjuntos 1, o maior valor de acurácia é obtido quando $n = 5$, para o conjunto 2 quando $n = 8$, e para o conjunto 3, quando $n = 10$.

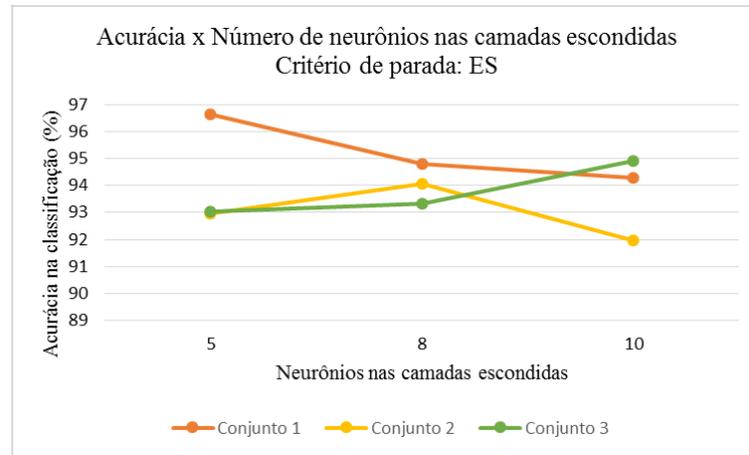


Figura 28 - Acurácia x Números de neurônios nas camadas escondidas com critério de parada antecipada

A Tabela 15, a seguir, mostra valores detalhados dos melhores classificadores projetados para cada conjunto, com parada antecipada, utilizando as 104 características como entrada.

Tabela 15 - Melhores resultados obtidos utilizando redes neurais com parada antecipada e as 104 características como entrada para cada conjunto.

Conjunto	N	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	5	2,32 ± 2,12	95,64 ± 4,15	96,64 ± 2,98	97,68 ± 2,12	97,5 ± 2,38	2,44 ± 2,27	96,54 ± 3,06	0,99 ± 0,01
C2	8	3,21 ± 3,15	91,41 ± 5,46	94,06 ± 4,19	96,79 ± 3,15	96,53 ± 3,4	3,23 ± 3,18	93,87 ± 4,36	0,98 ± 0,02
C3	10	3,15 ± 3,42	92,97 ± 5,57	94,92 ± 4,44	96,85 ± 3,42	96,56 ± 4,36	3,29 ± 4,23	94,71 ± 4,79	0,98 ± 0,02

N- número de neurônios nas camadas intermediárias; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.1.3 Redes neurais com regularização

A Figura 29 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios nas camadas escondidas varia entre 5, 8 e 10 e o critério de parada é a Regularização (REG). Pode-se observar que para os três conjuntos, a maior acurácia é obtida quando $n = 10$.

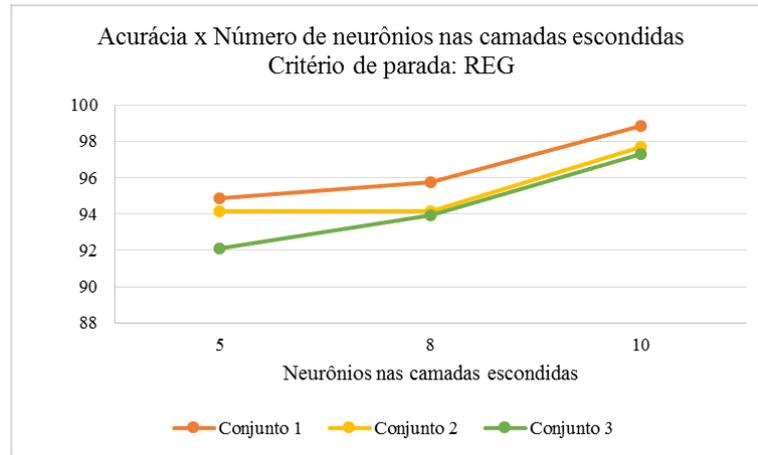


Figura 29 - Acurácia x Números de neurônios nas camadas escondidas com critério de parada regularização

Os valores detalhados dos melhores classificadores projetados para cada conjunto, com o critério de parada erro médio quadrático, utilizando todas as 104 características como entrada encontram-se na Tabela 16.

Tabela 16 - Melhores resultados obtidos utilizando redes neurais com regularização e as 104 características como entrada para cada conjunto.

Conjunto	N	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	10	1,26 ± 0,49	98,99 ± 0,87	98,87 ± 0,64	98,74 ± 0,49	98,76 ± 0,5	1,26 ± 0,49	98,86 ± 0,65	0,99 ± 0,01
C2	10	2,07 ± 2,29	97,43 ± 1,92	97,68 ± 1,84	97,93 ± 2,29	97,98 ± 2,21	2,09 ± 2,31	97,65 ± 1,82	0,99 ± 0,01
C3	10	2,02 ± 1,37	97,11 ± 1,61	97,55 ± 1,26	97,98 ± 1,37	97,94 ± 1,35	2,1 ± 1,45	97,46 ± 1,32	0,99 ± 0,01

N- número de neurônios nas camadas intermediárias; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.2 Segunda Fase

Nesta fase são selecionadas as 20 melhores características do conjunto através da técnica PCA. O classificador é então treinado utilizando-se das 5 até as 20 melhores características como entrada. As seguintes métricas são apresentadas: sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e área sob a curva ROC.

5.1.2.1 Redes Neurais com erro médio quadrático

A Figura 30 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de características de entrada varia entre 5, 6, 7, ... 20, selecionados por PCA. O critério de parada é o erro médio quadrático e o número de neurônios nas camadas escondidas é $n = 10$ para todos os conjuntos.

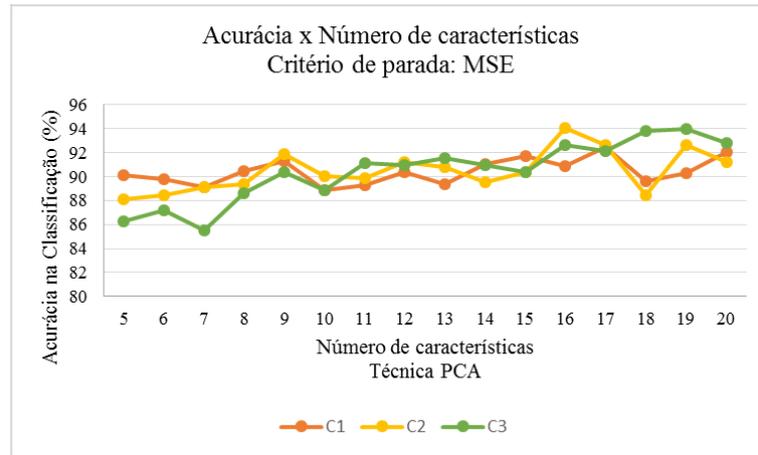


Figura 30- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados. O critério de parada é o erro médio quadrático, as redes têm 10 neurônios nas camadas escondidas.

Os valores detalhados dos melhores classificadores projetados para cada conjunto, com o critério de parada erro médio quadrático, utilizando conjuntos de características reduzidos por meio da técnica PCA, encontram-se na Tabela 17.

Tabela 17 - Desempenho dos classificadores RNA com erro médio quadrático e conjuntos reduzidos de características por PCA.

Conjunto	N	#C	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	10	17	8,62 ± 5,95	93,55 ± 6,27	92,48 ± 5,73	91,38 ± 5,95	92,04 ± 5,38	8,5 ± 5,89	92,63 ± 5,6	0,95 ± 0,05
C2	10	16	4,36 ± 2,2	92,51 ± 5,26	94,08 ± 3,01	95,64 ± 2,2	95,42 ± 2,34	4,43 ± 2,32	93,71 ± 3,46	0,96 ± 0,03
C3	10	19	4,55 ± 3,86	92,55 ± 3,09	94,01 ± 3,27	95,45 ± 3,86	95,23 ± 3,78	4,66 ± 3,92	93,73 ± 3,32	0,96 ± 0,03

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.2.2 Redes Neurais com parada antecipada

A Figura 31 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios características de entrada varia entre 5, 6, 7, ... 20, selecionados por PCA. O critério de parada é a parada antecipada (ES) e o número de neurônios nas camadas escondidas é, 5 para o conjunto 1, 8 para o conjunto 2 e 10 para o conjunto 3. O desempenho detalhado destes classificadores encontra-se na Tabela 18.

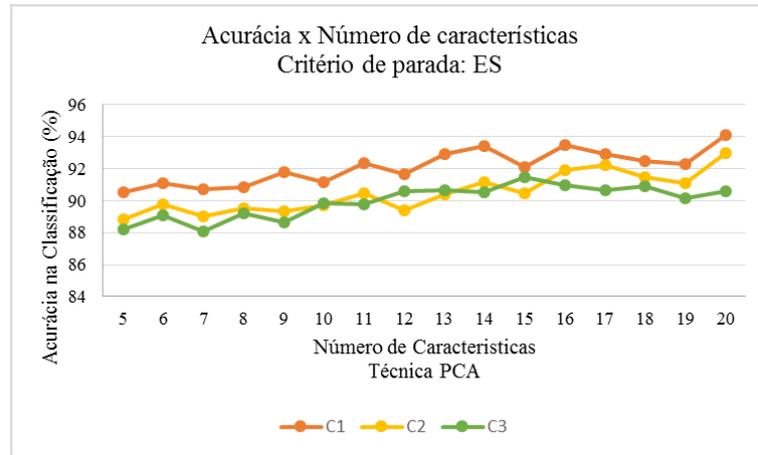


Figura 31- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados com parada antecipada (ES).

Tabela 18- Desempenho dos classificadores RNA com parada antecipada e conjuntos reduzidos de características por PCA.

Conjunto	N	#C	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	5	20	6,83 ± 5,77	95,06 ± 3,89	94,12 ± 4,68	93,17 ± 5,77	93,7 ± 5,21	6,76 ± 5,72	94,22 ± 4,53	0,97 ± 0,03
C2	8	20	5,96 ± 4,77	91,99 ± 4,56	93,01 ± 4,33	94,04 ± 4,77	94,15 ± 4,62	6,07 ± 4,85	92,85 ± 4,39	0,97 ± 0,03
C3	10	13	7,07 ± 4,66	88,28 ± 7,53	90,63 ± 5,74	92,93 ± 4,66	92,55 ± 5,11	7,31 ± 4,84	90,13 ± 6,17	0,95 ± 0,03

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.2.3 Redes Neurais com regularização

A Figura 32 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios características de entrada varia entre 5, 6, 7, ... 20, selecionados por PCA. O critério de parada é a Regularização (REG) e o número de neurônios nas camadas é escondidas $n = 10$ para todos os conjuntos.

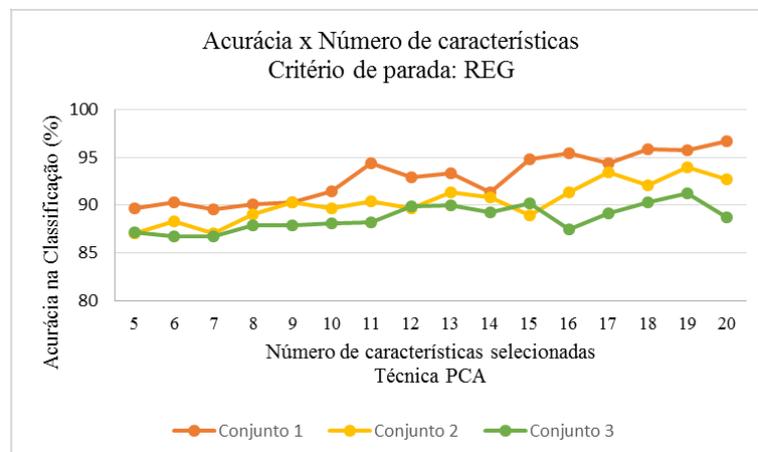


Figura 32- Acurácia x Número de características selecionadas por LDA para cada conjunto de dados. O critério de parada é a regularização

O desempenho detalhado destes classificadores encontra-se na Tabela 19.

Tabela 19- Desempenho dos classificadores RNA com regularização e conjuntos reduzidos de características por PCA.

Conjunto	N	#C	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	10	20	2,88 ± 1,73	96,14 ± 3,28	96,64 ± 2,35	97,12 ± 1,73	97,3 ± 1,64	2,84 ± 1,72	96,57 ± 2,48	0,98 ± 0,02
C2	10	19	5,9 ± 2,88	93,81 ± 4,14	93,94 ± 2,92	94,1 ± 2,88	94,2 ± 2,64	6,13 ± 3,03	93,82 ± 3,07	0,96 ± 0,02
C3	10	19	9,44 ± 6,2	91,97 ± 5,12	91,26 ± 5,16	90,56 ± 6,2	90,98 ± 5,76	9,72 ± 6,41	91,21 ± 5,2	0,94 ± 0,04

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.3 Terceira Fase

Nesta fase são selecionadas as 20 melhores características do conjunto através da técnica LDA. O classificador é então treinado utilizando-se das 5 até as 20 melhores características como entrada. As seguintes métricas são apresentadas: sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e área sob a curva ROC.

5.1.3.1 Redes Neurais com erro médio quadrático

A Figura 33 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de características de entrada varia entre 5, 6, 7, ... 20, selecionados por LDA. O critério de parada é o erro médio quadrático e o número de neurônios nas camadas escondidas é $n = 10$ para todos os conjuntos. As demais métricas de desempenho destes classificadores encontram-se na Tabela 20.

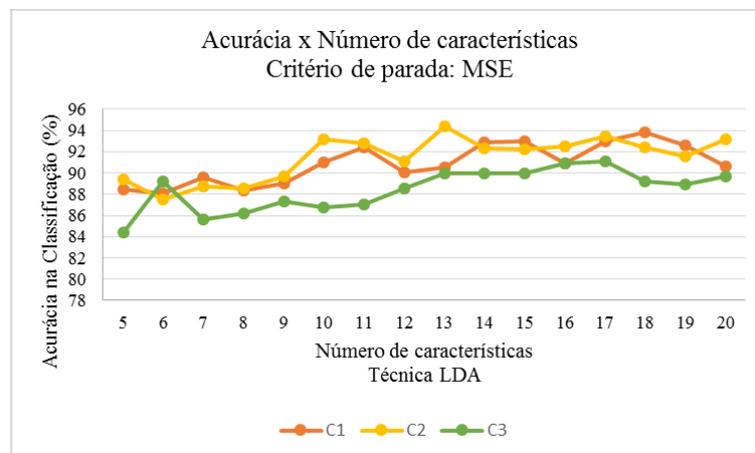


Figura 33- Acurácia x Número de características selecionadas por LDA para cada conjunto de dados. O critério de parada é o erro médio quadrático, as redes têm 10 neurônios nas camadas escondidas.

Tabela 20- Desempenho dos classificadores RNA com erro médio quadrático e conjuntos reduzidos de características por LDA.

Conjunto	N	#C	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	10	15	6,01 ± 3,29	91,97 ± 3,18	92,98 ± 3,1	93,99 ± 3,29	94,27 ± 3,01	5,92 ± 3,26	92,93 ± 3,07	0,96 ± 0,03
C2	10	13	5,72 ± 2,88	94,51 ± 1,67	94,39 ± 1,85	94,28 ± 2,88	94,6 ± 2,58	5,85 ± 2,98	94,39 ± 1,81	0,97 ± 0,01
C3	10	17	7,26 ± 3,3	89,32 ± 4,77	91,06 ± 3,76	92,74 ± 3,3	92,61 ± 3,49	7,47 ± 3,45	90,7 ± 3,98	0,95 ± 0,02

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.3.2 Redes Neurais com parada antecipada

A Figura 34 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios características de entrada varia entre 5, 6, 7, ... 20, selecionados por LDA. O critério de parada é a parada antecipada (ES) e o número de neurônios nas camadas escondidas é, 5 para o conjunto 1, 8 para o conjunto 2 e 10 para o conjunto 3. As demais métricas de desempenho destes classificadores encontram-se na Tabela 21.

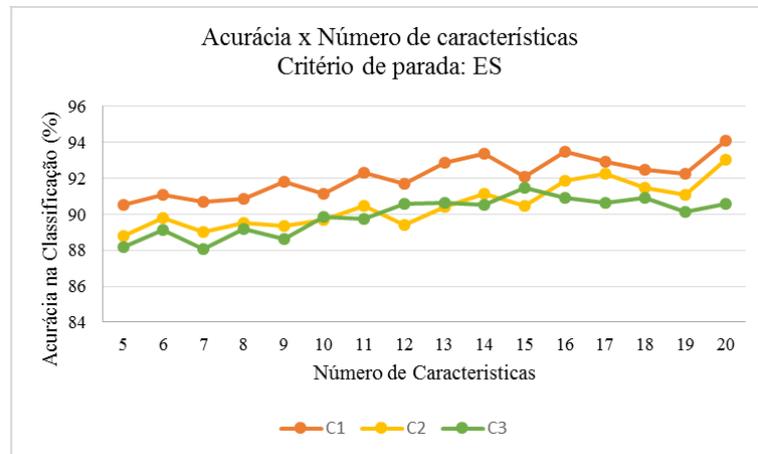


Figura 34- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados com parada antecipada (ES), o número de neurônios nas camadas escondidas é 5 para o conjunto 1, 8 para o conjunto 2 e 10 para o conjunto 3.

Tabela 21 - Desempenho dos classificadores RNA com parada antecipada e conjuntos reduzidos de características por LDA.

Conjunto	N	#C	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	5	18	6,86 ± 5,68	95,91 ± 2,15	94,54 ± 3,46	93,14 ± 5,68	93,94 ± 4,76	6,75 ± 5,56	94,72 ± 3,24	0,96 ± 0,02
C2	8	18	7,17 ± 6,86	92,32 ± 5,57	92,59 ± 5,61	92,83 ± 6,86	93,16 ± 6,18	7,25 ± 6,95	92,52 ± 5,53	0,97 ± 0,03
C3	10	13	7,73 ± 4,83	91,51 ± 6,05	91,88 ± 5,02	92,27 ± 4,83	92,35 ± 4,83	7,98 ± 5,02	91,71 ± 5,25	0,96 ± 0,03

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.3.3 Redes Neurais com regularização

A Figura 35 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de características de entrada varia entre 5, 6, 7, ... 20, selecionados por LDA. O critério de parada é a Regularização (REG) e o número de neurônios nas camadas escondidas $n = 10$.

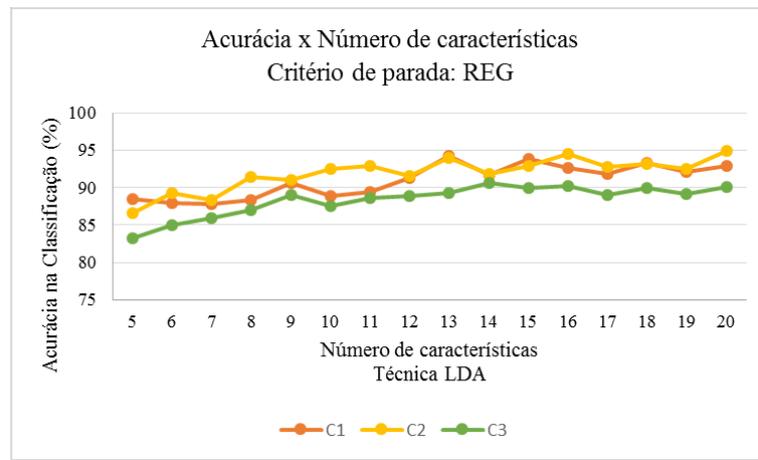


Figura 35- Acurácia x Número de características selecionadas por LDA para cada conjunto de dados. O critério de parada é a regularização, as redes têm 10 neurônios nas camadas escondidas.

As demais métricas de desempenho destes classificadores encontram-se na Tabela 22.

Tabela 22- Desempenho dos classificadores RNA com regularização e conjuntos reduzidos de características por LDA. As redes de todos os conjuntos possuem 10 neurônios nas camadas escondidas.

Conjunto	N	#C	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	10	13	4,47 ± 3,11	92,99 ± 3,47	94,24 ± 2,96	95,53 ± 3,11	95,64 ± 3	4,45 ± 3,09	94,13 ± 3,02	0,96 ± 0,03
C2	10	13	6,11 ± 2,54	93,92 ± 2,77	93,91 ± 2,33	93,89 ± 2,54	94,04 ± 2,37	6,21 ± 2,6	93,84 ± 2,38	0,97 ± 0,02
C3	10	19	6,03 ± 3	92,14 ± 3,95	90,58 ± 5,53	93,13 ± 4,1	92,77 ± 4,34	7,08 ± 4,25	90,03 ± 5,95	0,95 ± 0,04

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.4 Quarta fase

Nesta fase são selecionadas as 20 melhores características do conjunto através da técnica seleção escalar de características. O classificador é então treinado utilizando-se das 5 até as 20 melhores características como entrada. As seguintes métricas são apresentadas: sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e área sob a curva ROC.

5.1.4.1 Redes Neurais com erro médio quadrático

A Figura 36 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de características de entrada varia entre 5, 6, 7, ... 20, selecionados por seleção escalar de características. O critério de parada é o erro médio quadrático e o número de neurônios nas camadas escondidas é $n = 10$ para todos os conjuntos.

Tabela 23.

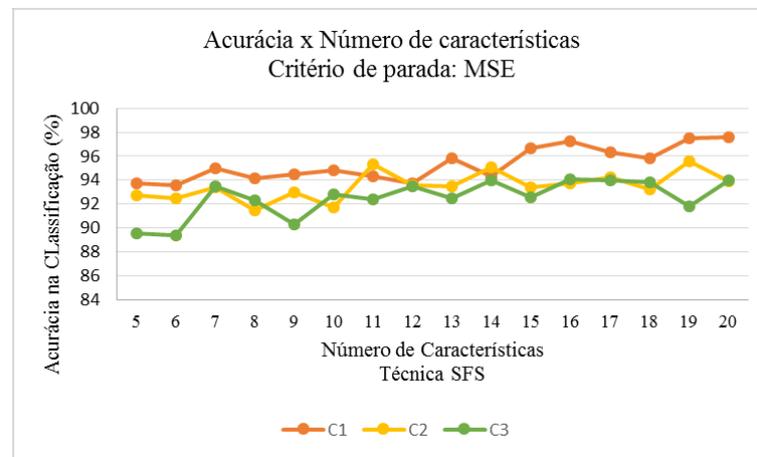


Figura 36- Acurácia x Número de características selecionadas com a técnica seleção escalar de características para cada conjunto de dados. O critério de parada é o erro médio quadrático, as redes têm 10 neurônios nas camadas escondidas.

Tabela 23- Desempenho dos classificadores RNA com erro médio quadrático e conjuntos reduzidos de características por seleção escalar. As redes de todos os conjuntos possuem 10 neurônios nas camadas escondidas.

Conjunto	N	#C	FPR	S	ACC	E	P	FA	F1	AUC
C1	10	20	1,6 ± 1,65	96,75 ± 2,66	97,57 ± 2,15	98,4 ± 1,65	98,37 ± 1,7	1,58 ± 1,64	97,51 ± 2,2	0,98 ± 0,02
C2	10	11	4,43 ± 2,95	95,14 ± 3,85	95,36 ± 2,94	95,57 ± 2,95	95,58 ± 3,02	4,51 ± 3	95,22 ± 3,13	0,96 ± 0,02
C3	10	20	4,6 ± 4,8	92,44 ± 4,52	93,95 ± 4,5	95,4 ± 4,8	95,4 ± 4,72	4,74 ± 4,9	93,72 ± 4,55	0,96 ± 0,04

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.4.2 Redes Neurais com parada antecipada

A Figura 37 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios características de entrada varia entre 5, 6, 7, ... 20, selecionados por seleção escalar. O critério de parada é a parada antecipada (ES) e o número de neurônios nas camadas escondidas é, 5 para o conjunto 1, 8 para o conjunto 2 e 10

para o conjunto 3. As demais métricas de desempenho destes classificadores encontram-se na Tabela 24.

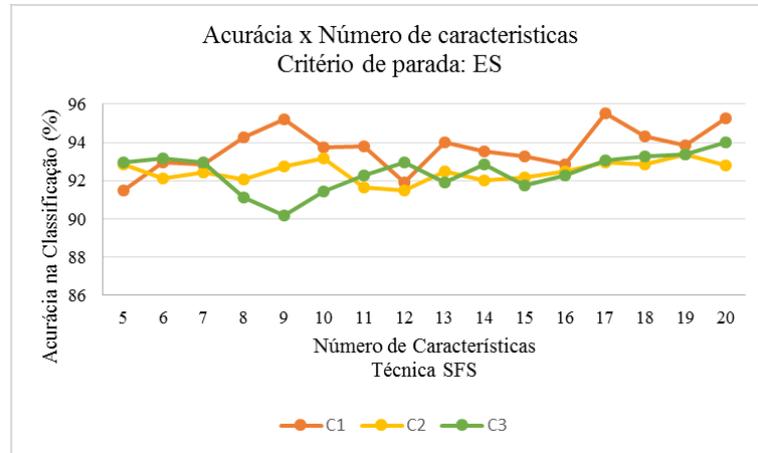


Figura 37- Acurácia x Número de características selecionadas com a técnica de seleção escalar de características para cada conjunto de dados com parada antecipada (ES), o número de neurônios nas camadas escondidas é 5 para o conjunto 1, 8 para o conjunto 2 e 10 para o conjunto 3

Tabela 24- Melhores resultados obtidos utilizando redes neurais com parada antecipada e conjuntos reduzidos por seleção escalar.

Conjunto	N	#C	FPR	S	ACC	E	P	FA	F1	AUC
C1	5	17	4,64 ± 3,26	95,68 ± 2,79	95,52 ± 2,39	95,36 ± 3,26	95,7 ± 2,88	4,58 ± 3,25	95,57 ± 2,35	0,97 ± 0,02
C2	8	10	5,58 ± 4,22	91,87 ± 5,46	93,15 ± 4,73	94,42 ± 4,22	94,26 ± 4,31	5,66 ± 4,29	92,94 ± 4,9	0,97 ± 0,02
C3	10	20	5,48 ± 3,24	93,43 ± 5,22	93,98 ± 3,77	94,52 ± 3,24	94,68 ± 3,12	5,6 ± 3,3	93,83 ± 3,97	0,97 ± 0,01

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.1.4.3 Redes Neurais com regularização

A Figura 38 mostra um comparativo dos valores de acurácia obtidos para cada um dos conjuntos de dados quando o número de neurônios características de entrada varia entre 5, 6, 7, ... 20, selecionados por seleção escalar de características. O critério de parada é a Regularização (REG) e o número de neurônios nas camadas escondidas $n = 10$ para todos os conjuntos. As demais métricas de desempenho destes classificadores encontram-se na Tabela 25.

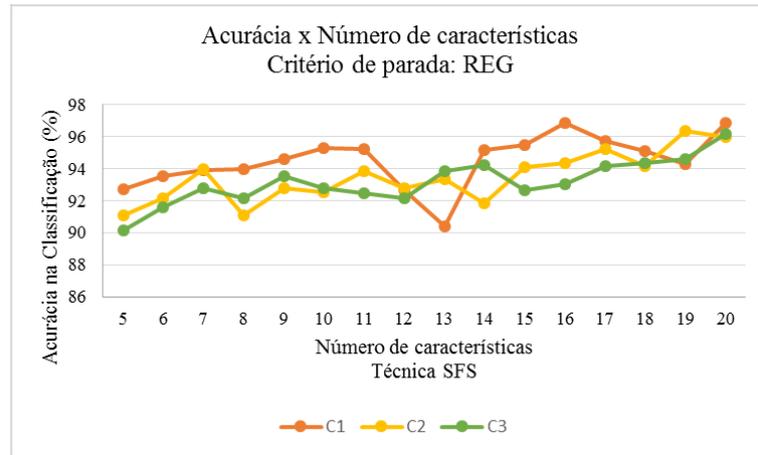


Figura 38- Acurácia x Número de características selecionadas com a técnica seleção escalar de características para cada conjunto de dados, o critério de parada é a regularização.

Tabela 25- Melhores resultados obtidos utilizando redes neurais com regularização e conjuntos reduzidos de características por seleção escalar. As redes de todos os conjuntos possuem 10 neurônios nas camadas escondidas.

Conjunto	N	#C	FPR	S	ACC	E	P	FA	F1	AUC
C1	10	20	1,75 ± 2,68	95,33 ± 4,47	96,89 ± 3,5	98,25 ± 2,68	97,89 ± 3,25	2,04 ± 3,14	96,48 ± 3,95	0,98 ± 0,04
C2	10	19	3,17 ± 2,37	95,79 ± 1,41	96,34 ± 1,84	96,83 ± 2,37	96,35 ± 2,69	3,77 ± 2,76	95,97 ± 1,98	0,97 ± 0,01
C3	10	20	3,54 ± 3,96	95,9 ± 2,02	96,19 ± 2,66	96,46 ± 3,96	96,59 ± 3,68	3,61 ± 4,06	96,09 ± 2,59	0,98 ± 0,02

N- número de neurônios nas camadas intermediárias; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2 CLASSIFICAÇÃO COM MÁQUINAS DE VETORES DE SUPORTE

5.2.1 Primeira fase

A primeira fase é aquela em que o classificador é treinado utilizando o conjunto original de dados, as 104 características compiladas por Macedo *et al.* (2015). Nesta fase, os hiperparâmetros são ajustados através da técnica busca exaustiva.

5.2.1.1 Kernel GRBF

A seguir são mostrados resultados de classificação em termos de acurácia para cada conjunto descrito na seção Materiais utilizando o kernel GRBF e o conjunto total de 104 características como entrada para o classificador. Nas figuras, são indicados resultados para diferentes pares (C, γ) , onde o parâmetro C é representado pelos índices K1, K2, K3 e K4 (ordem do kernel =1, ordem do kernel =2, ..., ordem do kernel = 4) e o eixo das abscissas corresponde ao parâmetro γ , que é variado de 0,03125 a 65536.

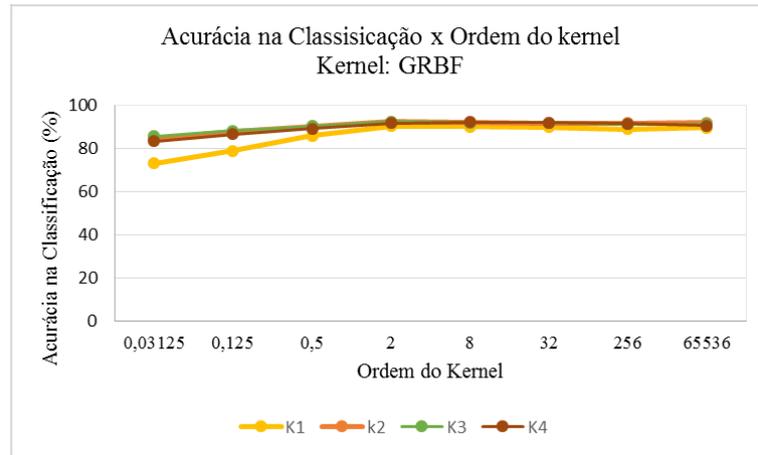


Figura 39- Acurácia de diferentes classificadores SVM utilizando o kernel GRBF obtidos através da busca exaustiva para classificação do Conjunto 1.

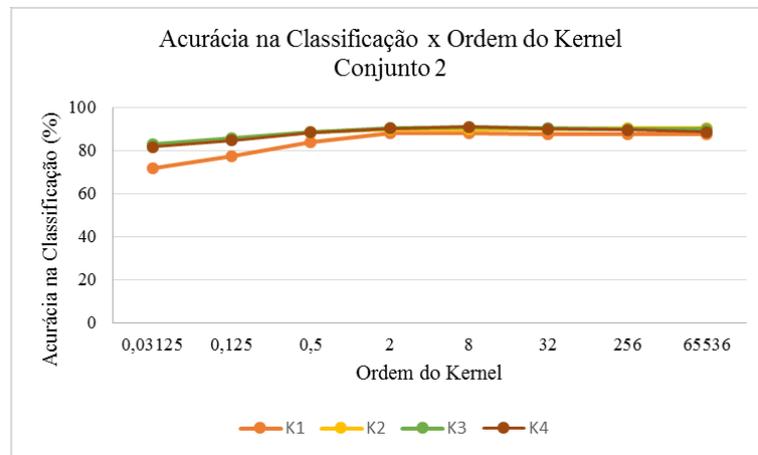


Figura 40- Acurácia de diferentes classificadores SVM utilizando o kernel GRBF obtidos através da busca exaustiva para classificação do Conjunto 2.

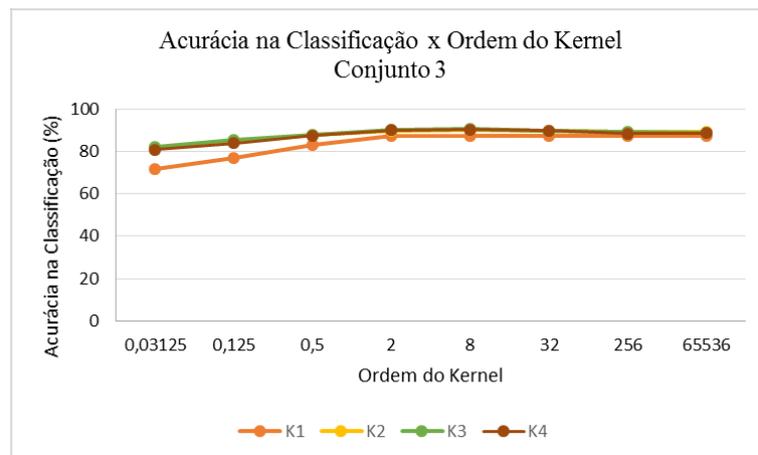


Figura 41- Acurácia de diferentes classificadores SVM utilizando o kernel GRBF obtidos através da busca exaustiva para classificação do Conjunto 3.

As figuras 39 a 41 indicam um crescimento na acurácia no intervalo $0,03123 < \gamma < 8$. Para refinar o desempenho do classificador, uma nova busca exaustiva foi realizada dentro do intervalo, para todos os conjuntos.

Na Tabela 26, são mostrados os melhores resultados de classificação obtidos para cada conjunto, utilizando o *kernel* GRBF quando todas as 104 características são utilizadas como entrada para o classificador.

Tabela 26- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel* GRBF e o conjunto completo de características.

Conjunto	C	γ	FPR	S	ACC	E	Precisão	FA	F1	AUC
C1	3	3,8	$6,48 \pm 3,72$	$90,18 \pm 5,25$	$91,96 \pm 4,08$	$93,52 \pm 3,72$	$92,54 \pm 4,28$	$7,52 \pm 4,39$	$91,13 \pm 4,57$	$0,92 \pm 0,04$
C2	4	4	$11,15 \pm 4,05$	$91,8 \pm 4,37$	$90,21 \pm 3,33$	$88,85 \pm 4,05$	$88,01 \pm 4,1$	$13,18 \pm 4,8$	$89,57 \pm 3,57$	$0,9 \pm 0,03$
C3	3	4,6	$7,75 \pm 3,37$	$88,58 \pm 5,87$	$90,58 \pm 4,11$	$92,25 \pm 3,37$	$90,89 \pm 3,9$	$9,31 \pm 3,99$	$89,48 \pm 4,66$	$0,9 \pm 0,05$

FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2.1.2 Kernel polinomial

Durante a busca exaustiva realizada utilizando-se o *kernel* polinomial não foi possível alcançar o valor máximo pré-estabelecido para o parâmetro γ devido a problemas de convergência quando a este era atribuído valores maiores que dois. Portanto, com o *kernel* polinomial, a busca exaustiva foi realizada dentro do conjunto:

$$\gamma = \{2^{-5}, 2^{-3}, 2^{-1}, 2^1\} \quad (58)$$

A seguir são mostrados resultados de classificação em termos de acurácia para cada conjunto descrito na seção Materiais utilizando o kernel polinomial e o conjunto total de 104 características como entrada para os classificadores.

Nas figuras 43, 44 e 45, são indicados resultados para diferentes pares (C, γ) , onde o parâmetro C é representado pelos índices K1, K2, K3 e K4 (ordem do kernel =1, ordem do kernel =2, ..., ordem do kernel = 4) e o eixo das abscissas corresponde ao parâmetro γ , que é variado de 0,03125 a 2.

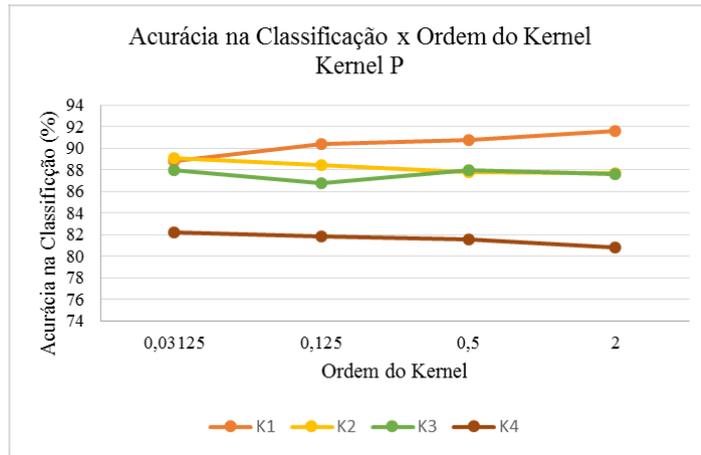


Figura 42 - Acurácia de diferentes classificadores SVM utilizando o kernel P obtidos através da busca exaustiva para classificação do Conjunto 1.

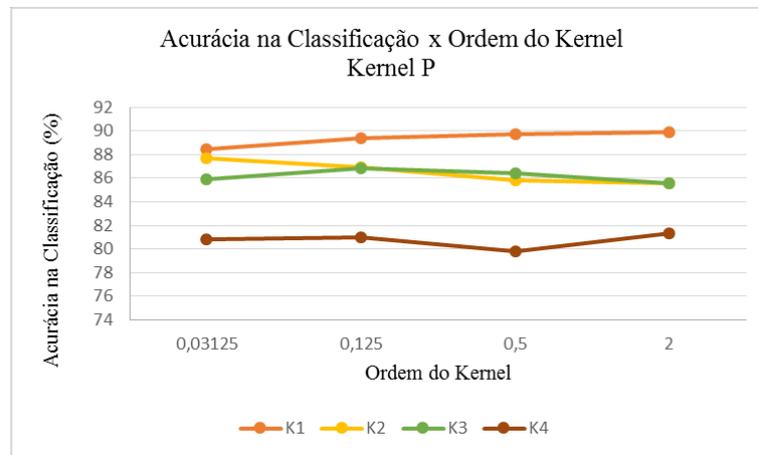


Figura 43- Acurácia de diferentes classificadores SVM utilizando o kernel P obtidos através da busca exaustiva para classificação do Conjunto 2.

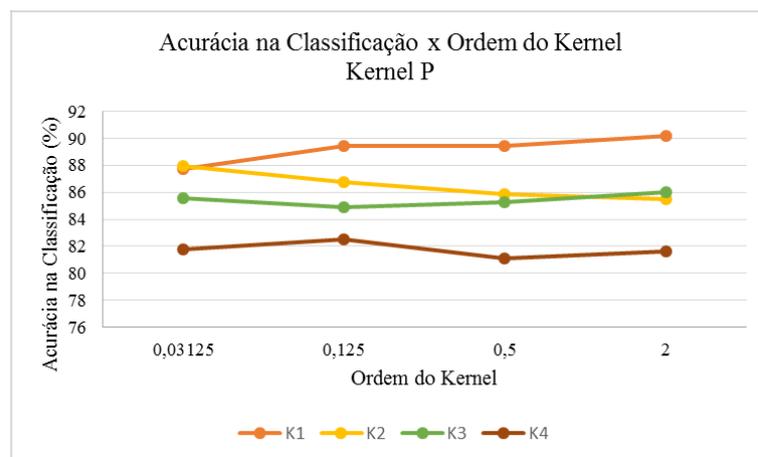


Figura 44 - Acurácia de diferentes classificadores SVM utilizando o kernel P obtidos através da busca exaustiva para classificação do Conjunto 3.

Podemos notar que o classificador com $C = 1$ apresenta resultados de acurácia superiores aos demais classificadores. Além disso, observa-se que o maior valor de acurácia

para esta configuração é obtido quando $\gamma = 2$. As métricas obtidas com melhor classificador implementado para os conjuntos 1, 2 e 3, utilizando o *kernel* polinomial e as 104 características são mostradas na Tabela, abaixo.

Tabela 27- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel* polinomial e o conjunto completo de características.

Conjunto	C	γ	FPR	S	ACC	E	P	FA	F1	AUC
C1	1	2	6,82 ± 3,65	88,8 ± 6,06	91,14 ± 4,68	93,18 ± 3,65	92,14 ± 4,52	7,89 ± 4,35	90,19 ± 5,3	0,91 ± 0,05
C2	1	2	6,99 ± 4,36	87,25 ± 6,45	90,37 ± 5,06	93,01 ± 4,36	91,59 ± 5,14	8,31 ± 5,15	89,08 ± 5,77	0,9 ± 0,5
C3	1	2	7,44 ± 4,07	86,35 ± 5,21	89,76 ± 4,38	92,56 ± 4,07	91,1 ± 4,95	8,88 ± 4,87	88,34 ± 5,03	0,89 ± 0,04

FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2.2 Segunda fase

Nesta fase são selecionadas as 20 melhores características do conjunto através da técnica PCA. O classificador é então treinado utilizando-se das 5 até as 20 melhores características como entrada. As seguintes métricas são apresentadas: sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e área sob a curva ROC.

5.2.2.1 Kernel GRBF

A Figura 45 mostra resultados de acurácia obtidos para cada classificador delimitado pelo processo de busca exaustiva realizado anteriormente. O conjunto 1 é treinado com um classificador com par (3, 3,8), o conjunto dois com o par (4,4) e o conjunto 3 com o par (3, 4,6).

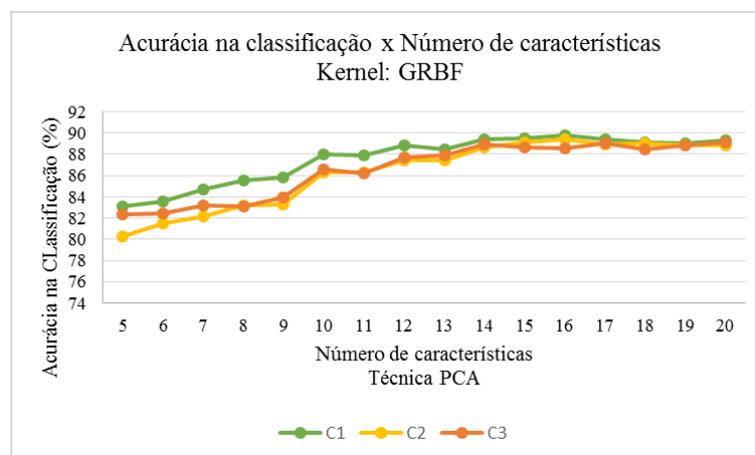


Figura 45- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados utilizando o *kernel* GRBF.

Na Tabela 28 encontram-se os melhores resultados de classificação obtidos para cada conjunto, extraídos a partir dos valores de acurácia extraídos da Figura 45.

Tabela 28- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel* GRBF e diferentes conjuntos de características na entrada do classificador selecionadas por PCA.

Conjunto	C	γ	f	FPR	S	ACC	E	P	FA	F1	AUC
C1	3	3,8	16	4,97 ± 2,77	84,65 ± 5,94	89,8 ± 4,3	95,03 ± 2,77	94,72 ± 2,94	4,93 ± 2,76	89,07 ± 4,76	0,9 ± 0,04
C2	4	4	16	5,03 ± 4,01	83,77 ± 7,37	89,41 ± 5,21	94,97 ± 4,01	94,52 ± 4,2	5,1 ± 4,05	88,41 ± 5,87	0,89 ± 0,05
C3	3	4,6	20	6,73 ± 3,56	84,78 ± 6,9	89,1 ± 5,03	93,27 ± 3,56	92,61 ± 4,15	6,91 ± 3,62	88,2 ± 5,54	0,89 ± 0,05

f – número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2.2.2 Kernel polinomial

A Figura 46 mostra resultados de acurácia obtidos para cada classificador delimitado pelo processo de busca exaustiva realizado anteriormente. Todos os conjuntos são treinados com o par (1,2).

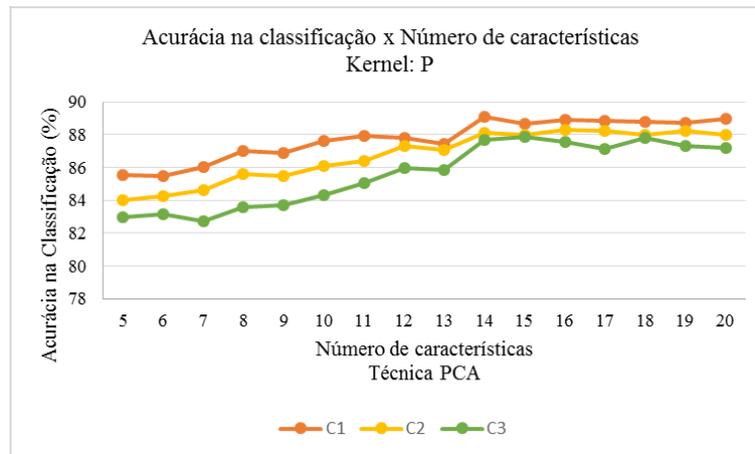


Figura 46- Acurácia x Número de características selecionadas com a técnica PCA para cada conjunto de dados utilizando o *kernel* P.

Na Tabela 29 encontram-se os melhores resultados de classificação obtidos para cada conjunto, extraídos a partir dos valores de acurácia extraídos da Figura 46.

Tabela 29- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel* P e diferentes conjuntos de características na entrada do classificador selecionadas por PCA.

Conjunto	C	γ	f	FPR	S	ACC	E	P	FA	F1	AUC
C1	1	2	14	5,58 ± 4,12	83,79 ± 5,73	89,08 ± 4,73	94,42 ± 4,12	94,22 ± 4,12	5,53 ± 4,08	88,39 ± 5,02	0,89 ± 0,05
C2	1	2	16	6,54 ± 4,92	83,01 ± 7,17	88,3 ± 5,65	93,46 ± 4,92	93,06 ± 5,52	6,59 ± 5	87,25 ± 6,22	0,88 ± 0,06
C3	1	2	15	7,27 ± 4,92	82,81 ± 6,91	87,86 ± 5,36	92,73 ± 4,92	92,08 ± 5,26	7,44 ± 5,04	86,84 ± 5,93	0,88 ± 0,05

f – número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2.3 Terceira fase

Nesta fase são selecionadas as 20 melhores características do conjunto através da técnica LDA. O classificador é então treinado utilizando-se das 5 até as 20 melhores características como entrada. As seguintes métricas são apresentadas: sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e área sob a curva ROC.

5.2.3.1 Kernel GRBF

A Figura 47 mostra resultados de acurácia obtidos para cada classificador delimitado pelo processo de busca exaustiva realizado anteriormente. O conjunto 1 é treinado com um classificador com par (3, 3,8), o conjunto dois com o par (4,4) e o conjunto 3 com o par (3, 4,6).

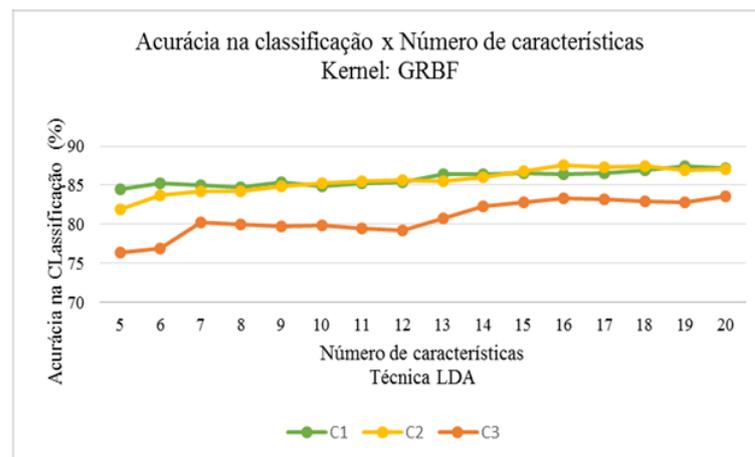


Figura 47- Acurácia x Número de características selecionadas com a técnica LDA para cada conjunto de dados utilizando o *kernel* GRBF

Na Tabela 30 encontram-se os melhores resultados de classificação obtidos para cada conjunto, extraídos a partir dos valores de acurácia extraídos da Figura 47.

Tabela 30- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel* GRBF e diferentes conjuntos de características na entrada do classificador selecionadas por LDA

Conjunto	C	γ	f	FPR	S	ACC	E	P	FA	F1	AUC
C1	3	3,8	19	5,99 ± 5,27	80,87 ± 8,84	87,4 ± 6,57	94,01 ± 5,27	93,36 ± 5,93	5,93 ± 5,22	86,2 ± 7,34	0,87 ± 0,07
C2	4	4	16	5,89 ± 3,17	81,2 ± 9,73	87,57 ± 6,22	93,86 ± 3,22	93,04 ± 4,03	6,19 ± 3,2	86,25 ± 7,25	0,88 ± 0,06
C3	3	4,6	20	6,98 ± 3,77	73,8 ± 8,76	83,52 ± 6,05	93,02 ± 3,77	91,42 ± 4,97	7,23 ± 3,86	81,1 ± 7,31	0,83 ± 0,06

f - número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2.3.2 Kernel polinomial

A Figura 48 mostra resultados de acurácia obtidos para cada classificador delimitado pelo processo de busca exaustiva realizado anteriormente. Todos os conjuntos são treinados com o par (1,2).

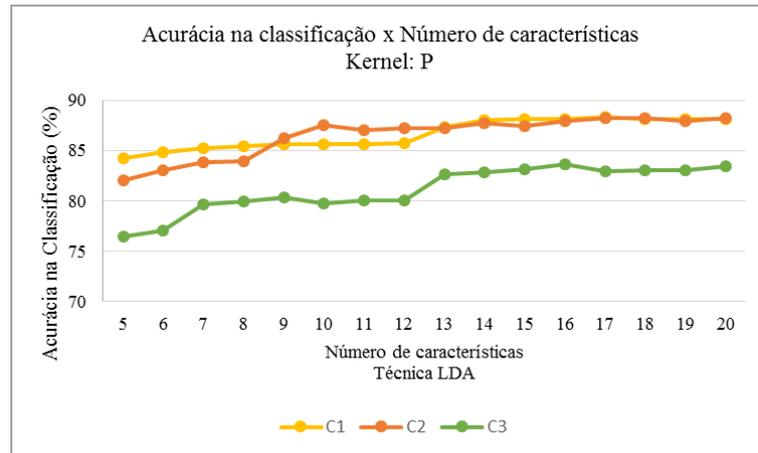


Figura 48- Acurácia x Número de características selecionadas com a técnica LDA para cada conjunto de dados utilizando o *kernel P*

Na Tabela 31 encontram-se os melhores resultados de classificação obtidos para cada conjunto, extraídos a partir dos valores de acurácia extraídos da Figura 48.

Tabela 31- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel P* e diferentes conjuntos de características na entrada do classificador selecionadas por LDA

Conjunto	C	γ	f	FPR	S	ACC	E	P	FA	F1	AUC
C1	1	2	17	6,34 \pm 4,91	83,03 \pm 8,57	88,32 \pm 6,68	93,66 \pm 4,91	92,93 \pm 5,77	6,28 \pm 4,84	87,47 \pm 7,29	0,88 \pm 0,07
C2	1	2	20	6,68 \pm 4,04	83,04 \pm 9,5	88,22 \pm 6,66	93,32 \pm 4,04	92,66 \pm 4,84	6,77 \pm 4,09	87,14 \pm 7,52	0,88 \pm 0,07
C3	1	2	16	5,79 \pm 2,76	72,65 \pm 11,74	83,6 \pm 6,82	94,21 \pm 2,76	92,7 \pm 3,98	5,92 \pm 2,85	80,61 \pm 9,2	0,83 \pm 0,07

f – número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2.4 Quarta fase

Nesta fase são selecionadas as 20 melhores características do conjunto através da técnica de seleção escalar de características. O classificador é então treinado utilizando-se das 5 até as 20 melhores características como entrada. As seguintes métricas são apresentadas: sensibilidade, acurácia, especificidade, precisão, taxa de falsos alarmes, *f-measure* e área sob a curva ROC.

5.2.4.1 Kernel GRBF

A Figura 49 mostra resultados de acurácia obtidos para cada classificador delimitado pelo processo de busca exaustiva realizado anteriormente. O conjunto 1 é treinado com um classificador com par (3, 3,8), o conjunto dois com o par (4,4) e o conjunto 3 com o par (3, 4,6).

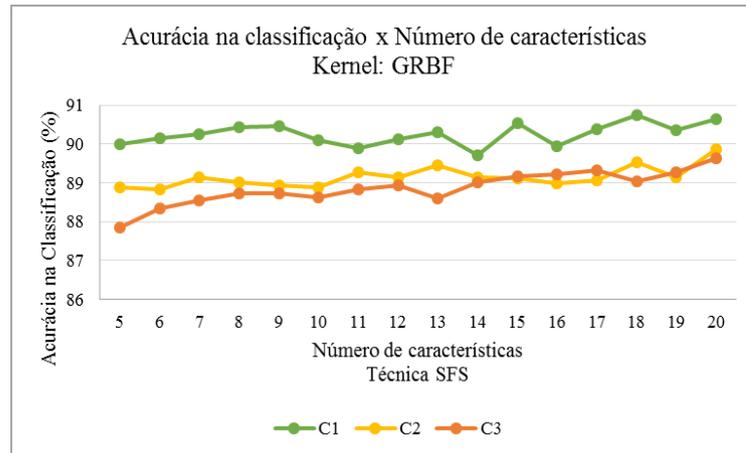


Figura 49- Acurácia x Número de características selecionadas com a técnica de seleção escalar para cada conjunto de dados utilizando o *kernel* GRBF

Na Tabela 32 encontram-se os melhores resultados de classificação obtidos para cada conjunto, extraídos a partir dos valores de acurácia extraídos da Figura 47.

Tabela 32- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel* GRBF e diferentes conjuntos de características na entrada do classificador selecionadas por seleção escalar.

Conjunto	C	γ	f	FPR	S	ACC	E	P	FA	F1	AUC
C1	3	3,8	18	6,11 ± 4,34	87,64 ± 6,44	90,75 ± 5,24	93,89 ± 4,34	93,89 ± 4,47	6,02 ± 4,29	90,35 ± 5,48	0,91 ± 0,05
C2	4	4	20	7,4 ± 5,14	87,12 ± 7,06	89,86 ± 5,98	92,6 ± 5,14	92,34 ± 5,51	7,54 ± 5,21	89,36 ± 6,32	0,9 ± 0,06
C3	3	4,6	20	7,8 ± 5,73	86,37 ± 7,38	89,63 ± 6,17	92,2 ± 5,73	91,59 ± 6,36	8,02 ± 5,91	88,69 ± 6,77	0,89 ± 0,06

f – número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

5.2.4.2 Kernel polinomial

A figura mostra resultados de acurácia obtidos para cada classificador delimitado pelo processo de busca exaustiva realizado anteriormente. Todos os conjuntos são treinados com o par (1,2).

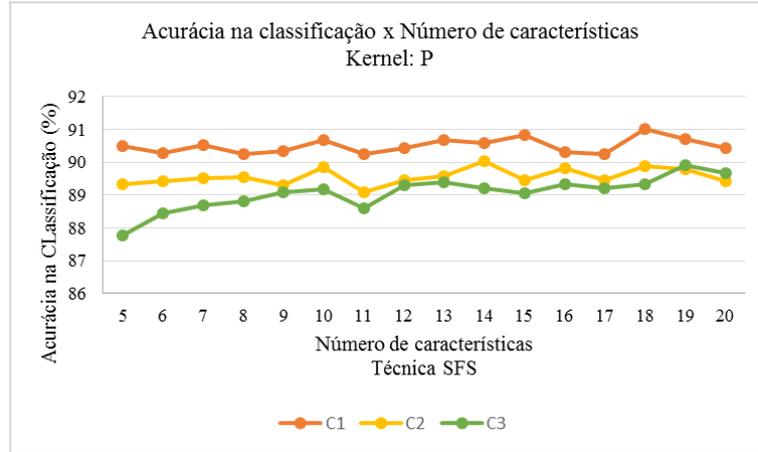


Figura 50- Acurácia x Número de características selecionadas com a técnica de seleção escalar para cada conjunto de dados utilizando o *kernel P*

Na Tabela 33 encontram-se os melhores resultados de classificação obtidos para cada conjunto, extraídos a partir dos valores de acurácia extraídos da Figura 50.

Tabela 33- Melhores resultados de classificação obtidos para cada conjunto utilizando SVM com o *kernel P* e diferentes conjuntos de características na entrada do classificador selecionadas por seleção escalar.

Conjunto	C	γ	f	FPR	S	ACC	E	P	FA	F1	AUC
C1	1	2	14	6 ± 4,27	88,07 ± 6	91,01 ± 5,08	94 ± 4,27	93,89 ± 4,51	5,93 ± 4,23	90,68 ± 5,32	0,91 ± 0,05
C2	1	2	14	7,33 ± 5,13	87,35 ± 7,25	90,03 ± 5,99	92,67 ± 5,13	92,2 ± 5,55	7,43 ± 5,16	89,56 ± 6,31	0,9 ± 0,06
C3	1	2	19	7,09 ± 5,22	86,84 ± 7,25	89,9 ± 6,03	92,91 ± 5,22	92,52 ± 5,57	7,32 ± 5,37	89,26 ± 6,47	0,9 ± 0,06

f – número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

DISCUSSÃO E CONCLUSÕES

Dois classificadores do estado da arte foram analisados neste trabalho: redes neurais artificiais e máquinas de vetores de suporte. Foram utilizados três métodos para aperfeiçoar a generalização das redes neurais: erro médio quadrático, regularização e parada antecipada. Os classificadores SVM foram treinados com diferentes *kernels*, GRBF e polinomial. Usou-se a técnica de busca exaustiva para encontrar as melhores combinações (C , γ) para a tarefa de classificação das regiões de bifurcação.

Uma peça chave do desenvolvimento deste trabalho é a etapa de balanceamento de dados. A decisão pelo balanceamento foi tomada após uma análise dos resultados parciais obtidos com os classificadores RNA, primeira etapa desenvolvida no trabalho.

A Figura 51 mostra os efeitos do balanceamento de dados no desempenho dos classificadores. Neste exemplo, são comparadas acurácia e sensibilidade de classificadores RNA com critério de parada regularização e dez neurônios nas camadas escondidas.

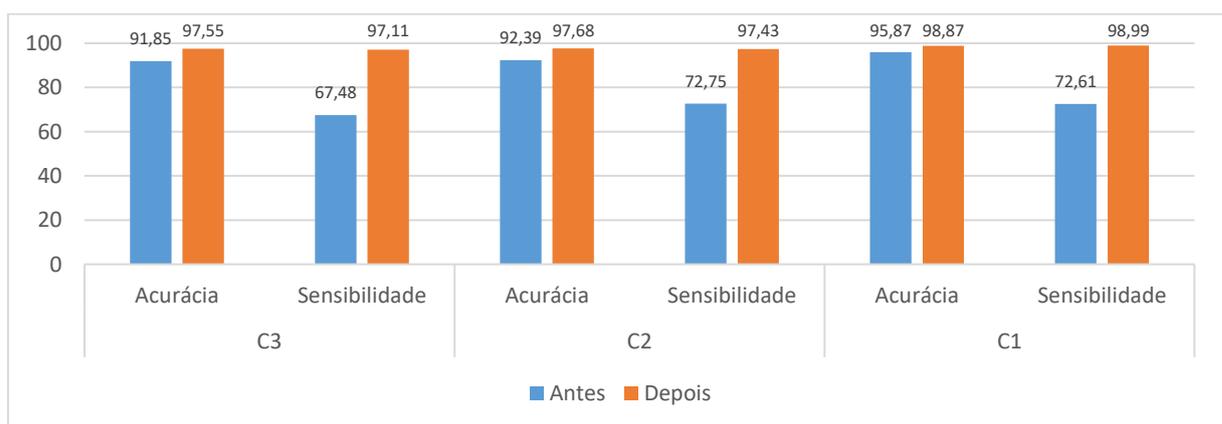


Figura 51- Efeitos do balanceamento de dados no desempenho dos classificadores RNA (reg, $n=10$), são analisadas medidas de acurácia e sensibilidade.

Nota-se que no conjunto 3, devido a etapa de balanceamento, a acurácia na classificação subiu de 91,85% para 97,55%. A sensibilidade aumentou de 67,48%, taxa muito próxima a de um classificador aleatório, para 97,11%. O mesmo padrão de comportamento foi observado em

outros conjuntos mostrando, portanto, a importância desta etapa no desenvolvimento deste trabalho, uma vez que a sensibilidade é uma métrica de extrema importância na avaliação de classificadores dedicados ao auxílio de tomadas de decisões médicas.

A escolha da arquitetura da RNA é uma etapa crucial no desenvolvimento dos classificadores. Foram testadas redes com 5, 8 e 10 neurônios nas camadas intermediárias. Os parâmetros das redes foram obtidos após sucessivos testes. A Tabela 34 mostra uma comparação entre os melhores classificadores obtidos utilizando RNA na primeira fase do trabalho.

Tabela 34- Comparação entre os melhores resultados obtidos com classificadores RNA utilizando o conjunto total de características.

Critério	N	#C	FPR	S	ACC	E	Precisão	FA	F1	AUC
Conjunto 1										
MSE	10	104	1,09 ± 0,73	98,07 ± 1,36	98,48 ± 0,9	98,91 ± 0,73	98,95 ± 0,71	1,17 ± 0,28	98,43 ± 0,96	0,99 ± 0,01
ES	5	104	2,32 ± 2,12	95,64 ± 4,15	96,64 ± 2,98	97,68 ± 2,12	97,5 ± 2,38	2,44 ± 2,27	96,54 ± 3,06	0,99 ± 0,01
REG	10	104	1,26 ± 0,49	98,99 ± 0,87	98,87 ± 0,64	98,74 ± 0,49	98,76 ± 0,5	1,26 ± 0,49	98,86 ± 0,65	0,99 ± 0,01
Conjunto 2										
MSE	10	104	1,28 ± 1,49	97,59 ± 2,05	98,15 ± 1,75	98,72 ± 1,49	98,67 ± 1,56	1,3 ± 1,51	98,08 ± 1,83	0,99 ± 0,01
ES	8	104	3,21 ± 3,15	91,41 ± 5,46	94,06 ± 4,19	96,79 ± 3,15	96,53 ± 3,4	3,23 ± 3,18	93,87 ± 4,36	0,98 ± 0,02
REG	10	104	2,07 ± 2,29	97,43 ± 1,92	97,68 ± 1,84	97,93 ± 2,29	97,98 ± 2,21	2,09 ± 2,31	97,65 ± 1,82	0,99 ± 0,01
Conjunto 3										
MSE	10	104	2,24 ± 1,1	96,88 ± 2,94	97,32 ± 1,83	97,76 ± 1,1	97,69 ± 1,17	2,36 ± 1,17	97,18 ± 2,05	0,98 ± 0,01
ES	10	104	3,15 ± 3,42	92,97 ± 5,57	94,92 ± 4,44	96,85 ± 3,42	96,56 ± 4,36	3,29 ± 4,23	94,71 ± 4,79	0,98 ± 0,02
REG	10	104	2,02 ± 1,37	97,11 ± 1,61	97,55 ± 1,26	97,98 ± 1,37	97,94 ± 1,35	2,1 ± 1,45	97,46 ± 1,32	0,99 ± 0,01

N- número de neurônios na camada intermediária; #C –número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

A partir de uma análise da Tabela 34 é possível notar uma pequena diferença entre o desempenho dos classificadores em cada conjunto utilizado, sendo o conjunto o que apresenta as melhores métricas. Atribui-se esse comportamento aos diferentes níveis de dificuldade de classificação presentes em cada conjunto. Também pode-se notar que o desempenho dos classificadores com critérios de parada MSE e REG foram superiores aos classificadores com critério de parada ES.

Foi realizada uma análise de significância estatística entre as médias das acurácias dos classificadores RNA com diferentes critérios de parada. O teste t de Student mostrou que não

há diferença entre os classificadores que utilizam os critérios erro médio quadrático e regularização ($p > 0,05$). A comparação entre os classificadores MSE e REG com o classificador ES mostrou que os classificadores que utilizam o critério de parada ES têm um desempenho inferior aos demais ($p < 0,05$).

A Tabela 35 mostra uma comparação entre os melhores resultados obtidos com classificadores RNA mostrando cada técnica de seleção de características utilizada. De uma forma geral, os classificadores que utilizaram como entrada subconjuntos obtidos por meio da seleção escalar de características apresentaram melhor desempenho.

Tabela 35- Comparação entre os melhores resultados obtidos com classificadores RNA e diferentes técnicas de seleção de características.

Técnica	CP	N	#C	FPR	S	ACC	E	P	FA	F1	AUC
Conjunto 1											
PCA	REG	10	20	2,88 ± 1,73	96,14 ± 3,28	96,64 ± 2,35	97,12 ± 1,73	97,3 ± 1,64	2,84 ± 1,72	96,57 ± 2,48	0,98 ± 0,02
LDA	ES	5	18	6,86 ± 5,68	95,91 ± 2,15	94,54 ± 3,46	93,14 ± 5,68	93,94 ± 4,76	6,75 ± 5,56	94,72 ± 3,24	0,96 ± 0,02
SFS	MSE	10	20	1,6 ± 1,65	96,75 ± 2,66	97,57 ± 2,15	98,4 ± 1,65	98,37 ± 1,7	1,58 ± 1,64	97,51 ± 2,2	0,98 ± 0,02
Conjunto 2											
PCA	MSE	10	16	4,36 ± 2,2	92,51 ± 5,26	94,08 ± 3,01	95,64 ± 2,2	95,42 ± 2,34	4,43 ± 2,32	93,71 ± 3,46	0,96 ± 0,03
LDA	MSE	10	13	5,72 ± 2,88	94,51 ± 1,67	94,39 ± 1,85	94,28 ± 2,88	94,6 ± 2,58	5,85 ± 2,98	94,39 ± 1,81	0,97 ± 0,01
SFS	REG	10	19	3,17 ± 2,37	95,79 ± 1,41	96,34 ± 1,84	96,83 ± 2,37	96,35 ± 2,69	3,77 ± 2,76	95,97 ± 1,98	0,97 ± 0,01
Conjunto 3											
PCA	MSE	10	19	4,55 ± 3,86	92,55 ± 3,09	94,01 ± 3,27	95,45 ± 3,86	95,23 ± 3,78	4,66 ± 3,92	93,73 ± 3,32	0,96 ± 0,03
LDA	ES	10	13	7,73 ± 4,83	91,51 ± 6,05	91,88 ± 5,02	92,27 ± 4,83	92,35 ± 4,83	7,98 ± 5,02	91,71 ± 5,25	0,96 ± 0,03
SFS	REG	10	20	3,54 ± 3,96	95,9 ± 2,02	96,19 ± 2,66	96,46 ± 3,96	96,59 ± 3,68	3,61 ± 4,06	96,09 ± 2,59	0,98 ± 0,02

CP- critério de parada; N- número de neurônios nas camadas intermediárias; #C- número de características; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

Foi realizada uma análise estatística entre as médias das acurácias dos classificadores RNA utilizando o conjunto total de características e conjuntos reduzidos de características na entrada do classificador. O teste t de Student mostrou que há uma diferença significativa entre os classificadores ($p < 0,05$), portanto, conclui-se que classificadores que utilizam conjunto reduzidos de características como entrada tem um desempenho inferior aos que utilizam o conjunto completo.

Em geral, os classificadores RNA obtiveram melhores resultados que os SVM. Atribui-se a este efeito os métodos utilizados para melhorar o desempenho da generalização de redes neurais e à escolha da arquitetura multicamadas.

A Tabela 36 mostra uma comparação entre os melhores resultados obtidos utilizando os classificadores SVM e os *kernels* polinomial (P) e GRBF com diferentes pares (C, γ). O teste t de Student mostrou que não há significância estatística entre os classificadores SVM com diferentes *kernels* ($p > 0,05$).

Tabela 36- Comparação entre os melhores resultados obtidos utilizando o classificador SVM e diferentes *kernels*.

Kernel	C	γ	FPR	S	ACC	E	P	FA	F1	AUC
Conjunto 1										
GRBF	3	3,8	6,48 ± 3,72	90,18 ± 5,25	91,96 ± 4,08	93,52 ± 3,72	92,54 ± 4,28	7,52 ± 4,39	91,13 ± 4,57	0,92 ± 0,04
P	1	2	6,82 ± 3,65	88,8 ± 6,06	91,14 ± 4,68	93,18 ± 3,65	92,14 ± 4,52	7,89 ± 4,35	90,19 ± 5,3	0,91 ± 0,05
Conjunto 2										
GRBF	4	4	11,15 ± 4,05	91,8 ± 4,37	90,21 ± 3,33	88,85 ± 4,05	88,01 ± 4,1	13,18 ± 4,8	89,57 ± 3,57	0,9 ± 0,03
P	1	2	6,99 ± 4,36	87,25 ± 6,45	90,37 ± 5,06	93,01 ± 4,36	91,59 ± 5,14	8,31 ± 5,15	89,08 ± 5,77	0,9 ± 0,5
Conjunto 3										
GRBF	3	4,6	7,75 ± 3,37	88,58 ± 5,87	90,58 ± 4,11	92,25 ± 3,37	90,89 ± 3,9	9,31 ± 3,99	89,48 ± 4,66	0,9 ± 0,05
P	1	2	7,44 ± 4,07	86,35 ± 5,21	89,76 ± 4,38	92,56 ± 4,07	91,1 ± 4,95	8,88 ± 4,87	88,34 ± 5,03	0,89 ± 0,04

FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

Conforme verificado com os classificadores RNA, com os classificadores SVM, o desempenho de classificação com o conjunto completo de características é melhor que o obtido utilizando conjuntos reduzidos.

A Tabela 37 mostra uma comparação entre os melhores resultados em termos de acurácia e sensibilidade obtidos nesta dissertação e os resultados de trabalhos que buscam identificar bifurcações em imagens vasculares publicados na literatura.

O melhor resultado obtido em termos de acurácia ultrapassa os resultados reportados até o presente momento. No que diz respeito a sensibilidade, o trabalho de Macedo *et al.* (2015) tem um desempenho superior a este.

Tabela 37- Identificação de bifurcações em imagens médicas: comparações com métodos publicados na literatura em termos de acurácia e sensibilidade.

Autores	Ano	Imagem	Técnica	S	ACC
Alberti <i>et al.</i>	2011	IVUS	AdaBoost	75,09± 13,7%	94±4,5%
Baboiu and Hamarneh	2012	Angiografia de retina	Filtro de bifurcação	95%±0,4	95% ~97%
Wang <i>et al.</i>	2014	IVOCT	Cálculo da distância	94%	Não apresentado
Macedo <i>et al.</i>	2015	IVOCT	<i>Adaboost</i>	98,89±0,67%	95,58±2,08%
Este trabalho	2017	IVOCT	RNA - MSE	97,36±0,81%	98,99 ± 0,87

S- sensibilidade; ACC- acurácia.

Discutimos com maiores detalhes as comparações entre este trabalho e o de Macedo *et al.* (2015), uma vez que o mesmo banco de dados é utilizado em ambos os trabalhos.

Na Tabela 38 é mostrado um resumo das melhores métricas reportadas por Macedo *et al.* (2015) para fins de comparação com os resultados obtidos nesta dissertação.

Tabela 38- Resumo dos melhores resultados de classificação obtidos por Macedo *et al.* (2015)

Classificador	FPR	S	ACC	E	P	FA	F1	AUC
Conjunto 1								
AB	27,48±13,87	98,90±0,89	95,58±2,08	75,52±13,87	96,17±1,86	3,97±1,98	97,51±1,15	96,41±3,18
AB+FROLS	38,46±13,78	98,99±0,67	94,25±1,85	61,54±13,78	94,70±1,86	5,57±2,07	96,79±1,01	94,96±4,13
Conjunto 2								
AB+FROLS	43,01±11,56	98,89±1,08	93,72±1,89	56,99±11,56	94,25±1,46	6,05±1,60	96,51±1,05	94,71±3,14
AB+MFROLS	52,55±14,51	99,13±0,87	92,74±1,81	47,45±14,51	93,08±1,82	7,41±2,08	96,00±0,98	92,60±3,18
Conjunto 3								
AB	39,93±11,41	98,05±1,62	93,42±0,94	60,07±11,41	94,69±1,45	5,54±1,58	96,32±0,53	94,36±4,19

AB- *AdaBoost*; FROLS - regressão progressiva ortogonal dos quadrados mínimos; MFROLS - múltiplos FROLS; FPR- taxa de falsos positivos; S- sensibilidade; ACC- acurácia; E- especificidade; P- precisão; FA- taxa de falsos alarmes; F1- *f-measure*; AUC- área sob a curva ROC.

No que diz respeito a acurácia, os resultados obtidos com RNA neste trabalho foram superiores aos reportados. Em termos de sensibilidade, no conjunto 1 conseguimos um empate e nos conjuntos 2 e três os resultados reportados por Macedo *et al.* (2015) sobrepujaram os aqui apresentados.

Os resultados apresentados neste trabalho em termos de taxa de falsos positivos (FPR) e especificidade são superiores aos reportados por Macedo *et al.* (2015). Nossas taxas médias de FPR encontram-se em torno de 2,07% com o classificador RNA, enquanto que no trabalho citado, os melhores resultados apresentam média de FPR de 40,2%. Em relação a

especificidade, os valores reportados por Macedo *et al.* (2015) chegam a níveis tão baixos quanto 47,45%. Dentre os melhores classificadores apresentados neste trabalho, o menor valor de especificidade medido foi de $88,85\% \pm 4,05$.

Nossos melhores classificadores (RNA) apresentaram melhores taxas de alarmes falsos e de área sob a curva ROC (AUC). O melhor valor de AUC reportado nesta dissertação, 0,99 é maior que aquele apresentado por Macedo *et al.* (2015). Realizamos um teste de significância estatística entre a diferença das áreas que se situam sob estas duas curvas e concluímos que os resultados aqui apresentados são superiores aos de Macedo *et al.* (2015) ($p < 0.05$).

Além das técnicas utilizadas para classificação dos dados, a principal diferença entre a metodologia abordada nesta dissertação e no trabalho de Macedo *et al.* (2015) é a etapa de balanceamento dos dados, que se revelou de extrema importância para o desempenho dos classificadores desenvolvidos.

Embora técnicas de redução de dimensionalidade sejam úteis no que diz respeito a eliminação de redundância do vetor de entrada e filtragem de ruído, neste trabalho, os classificadores que utilizam o conjunto completo de características mostraram melhores desempenhos quando comparado com aqueles que utilizam conjuntos reduzidos de características.

A partir de uma análise dos resultados obtidos, recomenda-se que, enquanto usando este banco de dados, a classificação seja feita com classificadores RNA com erro médio quadrático ou parada antecipada com 10 neurônios nas camadas intermediárias e o conjunto completo de características.

REFERÊNCIAS

AHA, A. H. A. Angina (Chest Pain) Dallas, 2015. Disponível em: < http://www.heart.org/HEARTORG/Conditions/HeartAttack/SymptomsDiagnosisofHeartAttack/Angina-Chest-Pain_UCM_450308_Article.jsp#.V6iM0PkrKUK >. Acesso em: 08/08/2016.

ALBERTI, M. et al. Automatic Branching Detection in IVUS Sequences. **Proceedings of Pattern Recognition and Image Analysis: 5th Iberian Conference, IbPRIA 2011**, Las Palmas de Gran Canaria, Spain, June 8-10, 2011. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p.126-133, 2012. ISBN 978-3-642-21257-4.

AZEVEDO-MARQUES, P. M. D. Diagnóstico auxiliado por computador na radiologia. **Radiologia Brasileira**, v. 34, p. 285-293, 2001. ISSN 0100-3984.

BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. **Journal of Microbiological Methods**, v. 43, n. 1, p. 3-31, 12/1/2000. ISSN 0167-7012.

BEZERRA, H. G. et al. Intracoronary Optical Coherence Tomography: A Comprehensive Review: Clinical and Research Applications. **JACC: Cardiovascular Interventions**, v. 2, n. 11, p. 1035-1046. 2009. ISSN 1936-8798.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. Springer, 2006. ISBN 9780387310732.

BOUMA, B. **Handbook of optical coherence tomography**. Informa Health Care, 2001. ISBN 0824705580.

BRAGA, A. D. P.; LUDERMIR, T. B.; CARVALHO, A. C. F. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro, RJ.: LTC - Livros Técnicos e Científicos, 2000.

BRATHWAITE, P. et al. Automatic classification and differentiation of atherosclerotic lesions in swine using IVUS and texture features. **Computers in Cardiology**, 2002, 2002, 22-25 Sept. 2002. p.109-112.

CARAMORI, P. R. A. et al. Long-term endothelial dysfunction after coronary artery stenting. **Journal of the American College of Cardiology**, v. 34, n. 6, p. 1675-1679, 11/15/1999. ISSN 0735-1097.

CARDIALYSYS, I. Angiography. **The Allround Trial Specialist in Cardiology**, The Netherlands, 2014. Disponível em: < <http://www.cardialysis.com/core-lab/angiography/> >.

CHAWLA, N. V. et al. SMOTE: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321-357, 2002. ISSN 1076-9757.

COSTA, L. D. F.; JR., R. M. C. **Shape Classification and Analysis: Theory and Practice**. CRC Press, Inc., 2009. 685 ISBN 0849379296, 9780849379291.

COSTOPOULOS, C. et al. Intravascular ultrasound and optical coherence tomography imaging of coronary atherosclerosis. **International Journal of Cardiovascular Imaging**, v. 32, n. 1, p. 189-200, Jan 2016. ISSN 1569-5794.

DOI, K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. **Computerized Medical Imaging and Graphics**, v. 31, n. 4-5, p. 198-211, 6// 2007. ISSN 0895-6111.

DUARTE, J. C. **O Algoritmo Boosting at Start e suas Aplicações**. 2009. (Doutorado). Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification (2nd Edition)**. Wiley-Interscience, 2000. ISBN 0471056693.

ELBASIONY, A.; LEVKOWITZ, H. Characterization of Atherosclerosis Plaque in OCT Images Using Texture Analysis and Parametric Equations. Information Visualisation (IV), 2011 15th International Conference on, 2011, 13-15 July 2011. p.237-240.

FISHER, R. A. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. **Annals of Eugenics**, v. 7, n. 2, p. 179-188, 1936. ISSN 2050-1439.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. In: VITÁNYI, P. (Ed.). **Computational Learning Theory: Second European Conference, EuroCOLT '95 Barcelona, Spain, March 13-15, 1995 Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995. p.23-37. ISBN 978-3-540-49195-8.

GAWAZ, M. et al. Platelet Activation and Coronary Stent Implantation. **Effect of Antithrombotic Therapy**, v. 94, n. 3, p. 279-285, 1996-08-01 00:00:00 1996.

GIGER, M.; BOONE, J.; CHAN, H. History and status of CAD and quantitative image analysis. **Med. Phys**, 2008.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Pearson Education, 2009. ISBN 9788131726952.

HAGAN, M. T.; DEMUTH, H. B.; BEALE, M. **Neural network design**. PWS Publishing Co., 1996. ISBN 0-534-94332-2.

DEMUTH, H. B.; BEALE, M.; HAGAN, M. T. "Improve Neural Network Generalization and Avoid Overfitting." 2013. Disponível em:
<<http://www.mathworks.com/help/nnet/ug/improve-neural-network-generalization-and-avoid-overfitting.html>>.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. Macmillan College Publishing Company, 1994. ISBN 9780132265560.

HAYKIN, S. S. **Neural Networks and Learning Machines**. Prentice Hall, 2009. ISBN 9780131471399.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, US, v. 24, n. 6, p. 417-441, 1933. ISSN 1939-2176(Electronic);0022-0663(Print).

HUANG, D. et al. Optical Coherence Tomography. **Science (New York, N.Y.)**, v. 254, n. 5035, p. 1178-1181, 1991. ISSN 0036-8075 1095-9203.

JAIN, A. K.; DUIN, R. P. W.; JIANCHANG, M. Statistical pattern recognition: a review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 1, p. 4-37, 2000. ISSN 0162-8828.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial Neural Networks: A Tutorial. **Computer**, v. 29, n. 3, p. 31-44, 1996. ISSN 0018-9162.

JOLLIFFE, I. T. **Principal Component Analysis**. Springer New York, 2013. ISBN 9781475719048.

KERKENI, A. et al. Automatic Bifurcation Detection in Coronary X-Ray Angiographies. 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), 2016, March 29 2016-April 1 2016. p.333-338.

KOHAVI, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. **International Joint Conference on Artificial Intelligence**, v. 2, p. 1137-1143, 1995.

KUBAT, M.; MATWIN, S. Addressing the curse of imbalanced training sets: one-sided selection. ICML, 1997, Nashville, USA. p.179-186.

LESAGE, D. et al. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. **Medical Image Analysis**, v. 13, n. 6, p. 819-845, 12// 2009. ISSN 1361-8415.

LI, Q.; NISHIKAWA, R. M. **Computer-Aided Detection and Diagnosis in Medical Imaging**. CRC Press, 2015. ISBN 9781439871775.

LINDEBERG, T. Principles for automatic scale selection. 1999.

MACEDO, M. M. **Aplicações Computacionais em Imagens Intravasculares de Tomografia por Coerência Óptica**. São Paulo: Instituto do Coração, Universidade de São Paulo 2014.

MACEDO, M. M. G. et al. A centerline-based estimator of vessel bifurcations in angiography images. 2013. p.86703K-86703K-7.

MACEDO, M. M. G; GUIMARÃES, W. V. N; GALON, M. Z; TAKIMURA, C. K; LEMOS, P.A, GUTIERRES, M. A. A bifurcation identifier for IV-OCT using orthogonal least squares and supervised machine learning. **Computerized Medical Imaging and Graphics**, v. 46, p. 237-248, 2015. ISSN 0895-6111.

MARQUES, C. M. G. **Confiabilidade Metrológica da Tomografia por Coerência Óptica em Aplicações Biomédicas**. 2012. (Mestrado). Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

MARTÍNEZ-TRINIDAD, J. F.; OCHOA, J. A. C.; KITTLER, J. **Progress in Pattern Recognition, Image Analysis and Applications: 11th Iberoamerican Congress on Pattern Recognition, CIARP 2006, Cancún, Mexico, November 14-17, 2006, Proceedings**. Springer Berlin Heidelberg, 2006. ISBN 9783540465577.

MATSUNAGA, V. Y. **Redes Neurais utilizando o MatLab**. Belém, PA, 2012.

MAUROVICH-HORVAT, P. et al. Differentiation of Early from Advanced Coronary Atherosclerotic Lesions: Systematic Comparison of CT, Intravascular US, and Optical Frequency Domain Imaging with Histopathologic Examination in ex Vivo Human Hearts. **Radiology**, v. 265, n. 2, p. 393-401, 2012/11/01 2012. ISSN 0033-8419.

MEHTRE, B. M.; KANKANHALLI, M. S.; WING FOON, L. Shape measures for content based image retrieval: A comparison. **Information Processing & Management**, v. 33, n. 3, p. 319-337, 1997/05/01 1997. ISSN 0306-4573.

MENGLIU, Z.; HAMARNEH, G. Bifurcation detection in 3D vascular images using novel features and random forest. **Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on**, 2014, April 29 2014-May 2 2014. p.421-424.

MURRAY, C. D. The Physiological Principle of Minimum Work: I. The Vascular System and the Cost of Blood Volume. **Proceedings of the National Academy of Sciences of the United States of America**, v. 12, n. 3, p. 207-214, 1926. ISSN 0027-8424/1091-6490.

OLIVEIRA, A. T. et al. Prevenção da reestenose pós-intervenções coronárias percutâneas: Os stents recobertos com drogas são a única opção terapêutica disponível? **Revista da Sociedade de Cardiologia do Rio Grande do Sul**, v. XIV, 2005.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, v. 2, n. 6, p. 559-572, // 1901.

PESARO, A. E. P. et al. Síndromes coronarianas agudas: tratamento e estratificação de risco. **Revista Brasileira de Terapia Intensiva**, v. 20, p. 197-204, 2008. ISSN 0103-507X.

RAO, C. R. The Utilization of Multiple Measurements in Problems of Biological Classification. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 10, n. 2, p. 159-203, 1948. ISSN 00359246.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. Editora Manole Ltda, 2003. ISBN 8520416837.

ROGOWSKA, J.; BREZINSKI, M. E. Evaluation of the adaptive speckle suppression filter for coronary optical coherence tomography imaging. **Medical Imaging, IEEE Transactions on**, v. 19, n. 12, p. 1261-1266, 2000. ISSN 0278-0062.

RUDZKI, M. Vessel detection method based on eigenvalues of the hessian matrix and its applicability to airway tree segmentation. Proceedings of the 11th International PhD Workshop OWD, 2009. p.100-105.

SCHALKOFF, R. J. **Artificial neural networks**. McGraw-Hill, 1997. ISBN 9780070571181.

SCHAPIRE, R. E.; FREUND, Y. **Boosting: Foundations and Algorithms**. MIT Press, 2012. ISBN 9780262017183.

SCHMITT, J. M. Intravascular imaging of atherosclerotic plaque with optical coherence tomography (oct). *Lasers and Electro-Optics, 2003. CLEO '03. Conference on, 2003, 6-6 June 2003.* p.407-409.

SERRUYS , P. W.; KUTRYK , M. J. B.; ONG , A. T. L. Coronary-Artery Stents. **New England Journal of Medicine**, v. 354, n. 5, p. 483-495, 2006.

SHLENS, J. A tutorial on principal component analysis. **arXiv preprint arXiv:1404.1100**, 2014.

SIU, S. O. M. Cerebrovascular diseases. **Vascular Surgery**, Illinois, 2012.

THE MATHWORKS, I. Neural Networks Overview. p. 06, Disponível em: < <http://www.mathworks.com/help/nnet/gs/neural-networks-overview.html> >. Acesso em: 17/02/2016.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition, Fourth Edition**. 4th. Academic Press, 2008. ISBN 1597492728, 9781597492720.

TSANTIS, S. et al. Automatic vessel lumen segmentation and stent strut detection in intravascular optical coherence tomography. **Medical physics**, v. 39, n. 1, p. 503-513, 2012. ISSN 0094-2405.

TU, S. et al. In vivo comparison of arterial lumen dimensions assessed by co-registered three-dimensional (3D) quantitative coronary angiography, intravascular ultrasound and optical coherence tomography. **The International Journal of Cardiovascular Imaging**, v. 28, n. 6, p. 1315-1327, 2012. ISSN 1573-0743.

TUNG, K.-P. **Coronary Segmentation in Intravascular Optical Coherence Tomography**. 2014. (Doctor of Philosophy). Department of Computing, Imperial College London, London.

TUNG, K. P. et al. AUTOMATIC VESSEL WALL DETECTION IN INTRAVASCULAR CORONARY OCT. **2011 8th Ieee International Symposium on Biomedical Imaging: from Nano to Macro**, p. 610-613, 2011. ISSN 1945-7928.

UPPUTURI, P. K. et al. Recent Developments in Vascular Imaging Techniques in Tissue Engineering and Regenerative Medicine. **BioMed Research International**, v. 2015, p. 9, 2015.

VAPNIK, V. **The Nature of Statistical Learning Theory**. Springer New York, 1999. ISBN 9780387987804.

VARELLA, C. A. A. **Análise de componentes principais Análise Multivariada Aplicada as Ciências Agrárias**: Universidade Federal Rural do Rio de Janeiro 2008.

VILES-GONZALEZ, J. F.; FUSTER, V.; BADIMON, J. J. Atherothrombosis: A widespread disease with unpredictable and life-threatening consequences. **European Heart Journal**, v. 25, n. 14, p. 1197-1207, 2004-07-01 00:00:00 2004.

WANG, A. et al. 3D assessment of stent cell size and side branch access in intravascular optical coherence tomographic pullback runs. **Computerized Medical Imaging and Graphics**, v. 38, n. 2, p. 113-122. 2014. ISSN 0895-6111.

WANG, C.; VENKATESH, S. S.; JUDD, J. S. Optimal stopping and effective machine complexity in learning. **Advances in neural information processing systems**, p. 303-303, 1994. ISSN 1049-5258.

WEBMD. Coronary Stent. **Heart Disease Health Center**, Atlanta, GA, 2016. Disponível em: <<http://www.webmd.com/heart-disease/coronary-stent>>.

WHO, World Health Organization. Cardiovascular diseases (CVDs), 2016. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs317/en/>>.

YEN, S.-J.; LEE, Y.-S. Cluster-based under-sampling approaches for imbalanced data distributions. **Expert Systems with Applications**, v. 36, n. 3, p. 5718-5727, 2009. ISSN 0957-4174.

ZARINS, C. et al. **Carotid bifurcation atherosclerosis. Quantitative correlation of plaque localization with flow velocity profiles and wall shear stress** 1983.

ZHOU, J. et al. VASCULAR STRUCTURE SEGMENTATION AND BIFURCATION DETECTION. 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007, 12-15 April 2007. p.872-875.

ANEXO

MÉTRICAS DE DESEMPENHO UTILIZADA

A seguir são descritas cada uma das métricas utilizadas para avaliar o desempenho dos classificadores.

- Taxa de falsos positivos

A taxa de falsos positivos mensura o número de elementos classificados incorretamente como positivos, sendo calculada como a razão entre o número de eventos negativos erradamente classificadas como positivos (falsos positivos) e o número total de eventos negativos reais.

$$FPR = \frac{FP}{VN + FP} \quad (59)$$

- Sensibilidade

A sensibilidade de um classificador refere-se à capacidade do classificador de detectar os indivíduos verdadeiramente positivos. Ou seja, no caso desta pesquisa, indicar corretamente as imagens pertencentes a região de bifurcação. O cálculo da sensibilidade é dado pela equação:

$$SENS = \frac{VP}{VP + FN} \quad (60)$$

- Acurácia

A acurácia mensura o grau pelo qual o classificador utilizado é capaz de determinar o verdadeiro valor daquilo que está sendo medido. O cálculo da acurácia é feito levando em consideração todos os acertos do classificador dividido pelo número total de amostras no conjunto.

$$ACC = \frac{(VP + VN)}{(VP + FP + VN + FN)} \quad (61)$$

- Precisão

A precisão, também chamada de valor preditivo positivo (PV+) é a probabilidade de um indivíduo avaliado com resultado positivo realmente pertencer a classe positiva. Neste trabalho,

a precisão mede a probabilidade de uma imagem classificada como BR realmente pertencer a classe BR.

$$PRE = \frac{VP}{VP + FP} \quad (62)$$

- Especificidade

A especificidade mede a capacidade do classificador em detectar corretamente os casos negativos. Nesta pesquisa a especificidade mede a taxa de acerto de classificação de imagens não pertencentes a região de bifurcação.

$$SPE = \frac{VN}{VN + FP} \quad (63)$$

- Taxa de falsos alarmes

A taxa de falsos alarmes (FAR) define o número de amostras classificadas incorretamente como positivas, dentre todas as amostras positivas de uma classe. Ou seja, o número de imagens erroneamente classificadas como BR dividido por todas as imagens BR existentes no conjunto (Martínez-Trinidad *et al.*, 2006).

$$FAR = \frac{FP}{VP + FN} \quad (64)$$

- *F-measure*

A *f-measure*, ou medida F_1 , transmite o equilíbrio entre a precisão e sensibilidade.

$$F1 = \frac{2PRE \cdot SENS}{(PRE + SENS)} \quad (65)$$

- Área sob a curva ROC

A área sob a curva ROC (AUC) é uma forma de mensurar as curvas ROC, que relaciona a sensibilidade e os falsos positivos (1-especificidade). Um classificador perfeito tem AUC igual a 1, enquanto que um classificador aleatório tem AUC 0,5. Por se tratar da área sob a curva, a AUC é calculada através da seguinte integral:

$$AUC = \int_0^1 f(x)dx \quad (66)$$

IDENTIFICAÇÃO DE REGIÕES DE BIFURCAÇÃO EM IMAGENS DE TOMOGRAFIA INTRAVASCULAR POR COERÊNCIA ÓTICA UTILIZANDO MÁQUINAS DE VETORES DE SUPORTE

C. D.L. Nascimento^{*}, M. G. F. Costa^{*}, M.M.G. Macedo^{**}, M. A. Gutierrez^{**}, C. F. F. Costa Filho^{*}

^{*} Centro de P&D em Tecnologia Eletrônica e da Informação, CETELI - UFAM, Manaus, Brasil

^{**} Instituto do Coração, InCor – HC, FM, USP, São Paulo, Brasil

e-mail: cdessana@gmail.com

Resumo: A angiografia coronária e o ultrassom intravascular têm sido utilizados na avaliação da aterosclerose. No entanto, uma nova modalidade de imagem médica, a tomografia intravascular por coerência ótica (IOCT), tem se diferenciado das demais por apresentar resolução espacial superior. Estudos têm demonstrado a importância da análise das regiões de bifurcação coronária em imagens médicas intravasculares, uma vez que estas regiões são mais propensas ao acúmulo de ateromas. Considerando que durante um exame de IOCT são adquiridas centenas de imagens, a rápida identificação das imagens correspondentes a regiões de bifurcação é, portanto, uma tarefa relevante, a qual é endereçada neste trabalho. Assim, analisa-se o desempenho de classificadores SVM implementados com diferentes parâmetros na tarefa de classificação de imagens IOCT correspondentes a regiões de bifurcação ou não. O método foi avaliado em imagens adquiridas de 9 exames de 9 pacientes. Descritores de forma foram utilizados como entrada para os classificadores. Conjuntos de treinamento compostos de 5 a 20 características foram selecionados segundo três métodos de seleção de características: seleção escalar, PCA e LDA. O melhor resultado obtido de acurácia foi 90,58%.

Palavras-chave: Doenças cardiovasculares, Tomografia Intravascular por Coerência Ótica (IOCT), Identificação de bifurcação, Máquina de Vetores de Suporte (SVM).

Abstract: *Coronary angiography and intravascular ultrasound has been used in the evaluation of atherosclerosis. However, a new medical imaging modality, intravascular optical coherence tomography (IOCT), has been differentiated from the others by having higher spatial resolution. Studies have shown the importance of coronary bifurcation regions analysis in intravascular medical image analysis, since these regions are more prone to atheroma accumulation. Whereas in*

the pullback of a patient hundreds of images are acquired, fast identification of images corresponding to bifurcation regions is therefore a significant task, which is addressed in this paper. Thus, we analyze the performance of SVM classifiers with different parameters in the classification task of IOCT images belonging to bifurcation regions or not. The method was evaluated using images acquired pullbacks 9 of 9 patients. Shape descriptors were used as input to the classifier. Training sets from 5 to 20 features were acquired according to three feature selection techniques: PCA, LDA and scalar feature selection. The best result obtained accuracy was 90.58%.

Keywords: *Cardiovascular disease, Optical Coherent Tomography, Bifurcation identifier, Support Vector Machines (SVM).*

Introdução

De acordo com a Organização Mundial da Saúde [1], as doenças cardiovasculares (DCV) são a principal causa de morte no mundo. Estima-se que 17,5 milhões de pessoas morreram de doenças cardiovasculares em 2012, representando 31% de todas as mortes globais. Destas, aproximadamente 7,4 milhões foram devido a doença cardíaca coronariana, também chamada de aterosclerose, caracterizada por espessamentos focais assimétricos nas artérias coronárias por um acúmulo gradual de placa dentro de suas paredes [2].

Transições na configuração da parede arterial nas curvas e bifurcações, também chamadas de ramificações, estão associadas a alterações na taxa do fluxo sanguíneo, que desempenha um papel relevante na progressão da aterosclerose [3]. Estas regiões também são predispostas ao desenvolvimento de espessamentos da íntima, incluindo a formação de placas ateroscleróticas. A presença de placas pode reduzir significativamente o

fluxo sanguíneo, levando a graves complicações de saúde, como anginas e acidentes vasculares cerebrais. Detectar e caracterizar estas regiões potencialmente vulneráveis pode ajudar a identificar a presença de placas e a compreender a progressão da doença.

Para a avaliação da aterosclerose duas grandes modalidades de imagem médica têm sido utilizadas, angiografia coronariana e ultrassom intracoronário (UIC) [4]. A tomografia intravascular por coerência ótica (IOCT) é uma nova modalidade de imagem médica que adquire imagens da microestrutura vascular em alta resolução, $12\mu\text{m}$ a $18\mu\text{m}$ em comparação com $150\mu\text{m}$ a $200\mu\text{m}$ para UIC, através da mensuração da magnitude e do tempo de retorno da luz infravermelha refletida pelos tecidos [5].

Durante a aquisição de imagens IOCT em aplicações clínicas, uma única corrida do cateter contém tipicamente centenas de imagens [5]. A automatização da identificação de bifurcações poderia acelerar o processo de análise destas imagens, indicando ao especialista quais delas mostram bifurcações das artérias e auxiliar métodos automatizados como análise de *stent* [6], análise de placa [7] e comparações entre as diferentes modalidades de imagens intravasculares [8].

Alguns estudos encontrados na literatura já propuseram métodos para identificação de bifurcações. Alberti et al. [9], extraíram um conjunto de características de textura das sequências de imagens UIC e utilizaram métodos de aprendizado de máquina supervisionado para identificação. Zhao and Hamaneh [10] utilizaram *Random Forest* para identificar bifurcações através de características geométricas em imagens vasculares 3D. Wang et al. [11] propuseram um método automatizado para identificar cortes de bifurcação das imagens IOCT usando a distância entre o centro do lúmen e o contorno do lúmen como um indicador de bifurcação. Macedo et al. [12] propuseram um método totalmente automatizado para diferenciar imagens IOCT como bifurcação compreendendo as tarefas de detecção automática do lúmen, extração de características geométricas e classificação utilizando e métodos de aprendizado de máquina.

Neste trabalho buscamos avaliar o desempenho de classificadores SVM na tarefa de identificação de bifurcações em imagens IOCT utilizando as características compiladas em [12]. Utilizou-se PCA, LDA e seleção escalar de características para reduzir a dimensionalidade do referido conjunto de dados. Combinações diversas de características e de *kernels* foram empregadas, no sentido de otimizar a taxa de classificação.

Materiais e métodos

Conjunto de dados – As características geométricas utilizadas neste trabalho, fornecidas por

Macedo et al. [12], foram compiladas de imagens coronarianas adquiridas utilizando um sistema IOCT no domínio de Fourier no Instituto do Coração (InCor), do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo. As imagens adquiridas são originárias de nove exames de nove pacientes, as quais foram adquiridas imediatamente antes do implante de *stent*, processo realizado com a finalidade de desobstrução das artérias.

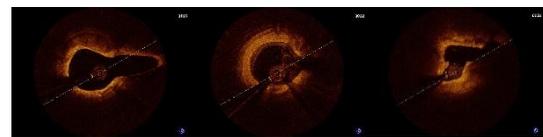
Os dados (características geométricas do lúmen segmentado) utilizados no presente trabalho são derivados de um conjunto de imagens com diferentes desafios de classificação, a saber: imagens com bifurcação, mas com aparente ausência de dificuldades de identificação da mesma; imagens contendo lesões ateroscleróticas complexas; e imagens contendo sangue residual. Na Figura 1, apresentam-se exemplos das categorias de imagens referidas. A Tabela 1 apresenta a composição do conjunto.

Os descritores utilizados para extrair características geométricas do contorno do lúmen segmentado foram: distância do centro da artéria ao centroide do contorno do lúmen, circularidade, energia de dobramento, curvatura, variância radial, taxa de eixo, variância da assinatura do contorno, desvio padrão da assinatura do contorno, variância radial, diferença de área entre cortes e área média de triângulos dentro de um contorno. Outros descritores foram determinados por meio de derivações e combinações destes, totalizando 104 características diferentes [12].

Tabela 1: Distribuição das imagens das quais foram extraídos os descritores de forma.

Número de imagens		
NBR ¹	BR ²	Total
1282	178	1460

¹NBR – imagens em região sem bifurcação;
²BR – imagens em região de bifurcação



(a) (b) (c)

Figura 1: Exemplos de imagens do conjunto (a) BR normal (b) BR com sangue residual (c) BR com placa complexa.

Balanceamento dos conjuntos – Uma vez que a classe BR (com bifurcação) do conjunto formado por Macedo et al. [12] representa cerca de 12% dos dados, identificou-se a necessidade de realizar um balanceamento dos dados intraconjunto. Assim, esse conjunto foi transformado em seis subconjuntos composto da seguinte forma: as imagens da classe NBR foram divididas, de forma aleatória, proporcionalmente em seis partes. As imagens BR

foram então adicionadas a cada um deles, formando então 6 subconjuntos balanceados.

Máquinas de vetores de suporte – Máquinas de Vetores de Suporte (SVM) são máquinas de aprendizado binárias usadas para separar dados pertencentes a duas classes, a partir de um hiperplano de separação. Para classes separáveis, os parâmetros do hiperplano que maximizam a margem são calculados através da determinação do vetor de pesos w e da polarização w_0 , tal que a expressão (1) seja minimizada satisfazendo as condições de Karush-Kuhn-Tucker (KKT) [14].

$$J(w) = \frac{1}{2} \|w\|^2 \quad (1)$$

Para classes não separáveis, os mesmos parâmetros podem ser calculados minimizando a expressão (2), onde novas variáveis ξ_i , conhecidas como variáveis de folga são introduzidas. O objetivo é maximizar a margem, mantendo o número de pontos com $\xi > 0$ o menor possível [14].

$$J(w, w_0, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (2)$$

O parâmetro C na expressão (2) é uma constante positiva que estabelece um compromisso entre a variável de folga e a margem [15].

SVMs utilizam *kernels* para mapear o vetor de características para um espaço de dimensão maior e dessa forma permitir a classificação em espaços não separáveis linearmente [14, 15]. Neste trabalho utilizamos os *kernels* GRBF (*gaussian radial basis function*) e polinomial (P) variando suas ordens de 1 a 4. Os valores do parâmetro C utilizados variaram entre 0,2 e 32. A fim de obter o melhor classificador SVM, as simulações foram realizadas combinando cada um dos *kernels* com os valores escolhidos para o parâmetro C . Para fins de treinamento e teste do classificador foi utilizado 10-fold-cross-validation [16].

Seleção de características – Com o objetivo de avaliar o efeito de um conjunto reduzido de variáveis de entrada sobre o desempenho dos classificadores, foram utilizados três métodos de seleção de características: análise de discriminante linear (LDA) [17], análise de componentes principais (PCA) [18] e seleção escalar de características (SFS) [14]. Conjuntos compostos por 5 a 20 características (obtidas através de cada método de seleção) foram utilizados como entrada para os classificadores.

Resultados

A seguir apresentam-se os resultados das diversas implementações do classificador SVM avaliados com os 6 subconjuntos balanceados de dados, em sua íntegra ou com características selecionadas desses subconjuntos, através das técnicas PCA, LDA e Seleção Escalar. Ressalta-se

que os resultados apresentados são resultados médios dos 6 subconjuntos.

A Tabela 2 mostra os melhores resultados de classificação obtidos utilizando os *kernels* polinomial e GRBF, variando a ordem do *kernel* e o parâmetro C , quando todas as 104 características são utilizadas.

Tabela 2: Resultados de acurácia, taxa de verdadeiros positivos, especificidade e área sob a curva ROC para classificadores SVM com diferentes *kernels* e parâmetros utilizando 104 características.

Kernel	Ordem	C	ACC ¹	TPR ²	SPE ³	AUC ⁴
P	1	2,2	89,92±5,36	87,32±6,08	92,45±4,77	0,90±0,06
P	2	0,03	88,21±4,66	83,76±5,35	92,49±4,15	0,88±0,05
GRBF	3	4,6	90,95±4,96	89,83±5,84	92,05±4,37	0,90±0,05
GRBF	1	4,4	87,96±4	95,82±2,23	80,27±6,8	0,88±0,04

¹ACC – acurácia (%); ²TPR – taxa de verdadeiros positivos (%)

³SPE – especificidade (%); ⁴AUC – área sob a curva ROC

A Tabela 3 mostra os melhores resultados obtidos para cada uma das técnicas de classificação utilizadas considerando as configurações de *kernel*, ordem do *kernel* e parâmetro C apresentados na Tabela 2, onde a configuração 1 corresponde ao *kernel* polinomial com C 2,5 e ordem 1, a configuração 2 ao *kernel* polinomial com C 1,2 e ordem 2, a configuração 3 ao *kernel* GRBF com C 2,4 e ordem 3 e a configuração 4 ao *kernel* GRBF com C 4 e ordem 1.

Tabela 3: Resultados de acurácia, taxa de verdadeiros positivos, especificidade e área sob a curva ROC para classificadores SVM com configurações e conjuntos reduzidos de características obtidos com as técnicas PCA, LDA e SFS.

Conf ¹	N ²	Técnica	ACC ³	TPR ⁴	SPE ⁵	AUC ⁶
1	18	PCA	88,34±5,16	83,24±5,95	93,27±5,06	0,88±0,05
1	20	LDA	83,46±7,15	72,96±11,75	94,13±3,39	0,83±0,07
1	20	SFS	89,65±6,55	87,01±6,99	92,74±5,15	0,90±0,06
2	15	PCA	85,75±6,5	74,63±10,88	96,56 ±3,22	0,86±0,07
2	20	LDA	89,06±7,99	73,81±12,80	93,81±4,1	0,82±0,03
2	12	SFS	88,41±7,09	85,40±8,19	91,80±5,28	0,88±0,07
3	20	PCA	89,06±4,86	84,74±6,61	93,26±3,64	0,89±0,05
3	19	LDA	83,41±6,2	73,81±9,22	93,28±3,58	0,83±0,07
3	18	SFS	89,63±6,97	87,09±7,99	92,12±6,05	0,90±0,07
4	16	PCA	89,57±5,15	89,23±4,28	90,59±5,52	0,90±0,05
4	17	LDA	85,71±5,11	79,77±7,43	91,51±3,98	0,86±0,05
4	12	SFS	89,14±6,08	89,67±6,11	90,07±6,17	0,89±0,06

¹Conf – configuração do classificador; ²N – número de características

³ACC – acurácia (%); ⁴TPR – taxa de verdadeiros positivos (%)

⁵SPE – especificidade (%); ⁶AUC – área sob a curva ROC

Discussão

Neste trabalho buscou-se avaliar o efeito da variação da ordem do *kernels* e do parâmetro C no

desempenho do classificador, resultando em 100 diferentes configurações para cada um dos *kernels* utilizados. Apesar desta ampla faixa de possibilidades, o *kernel* polinomial mostrou melhor desempenho variando sua ordem de 1 a 2 e *C* de 0,2 a 2,6. Já o *kernel* GRBF obteve melhor performance variando sua ordem de 1 a 4 e *C* de 2 a 16.

Os maiores valores de acurácia obtidos 90,95% e 89,92% respectivamente, foram alcançados utilizando o conjunto completo de características, conforme mostrado na Tabela 2. Estes valores são superiores aos apresentados em [12] que também utilizou SVM. Além dos inúmeros cenários de simulação realizados, credita-se esta melhora no desempenho do classificador SVM ao balanceamento do número de amostras das duas classes (BR e NBR) no conjunto de dados, proposto e implementado neste trabalho.

Ainda em relação a Tabela 2 observa-se que os classificadores com *kernel* GRBF obtiveram melhores resultados de TPR em comparação ao *kernel* polinomial. Como pode ser observado, algumas configurações apresentam alta especificidade e baixa taxa de verdadeiros positivos e outras o contrário. A definição de qual configuração apresenta melhor resultado depende de quais parâmetros são mais importantes para o especialista que fará uso do método aqui proposto.

Os valores de especificidade obtidos neste trabalho também superaram os apresentados por Macedo et al. [12] em até 35%. Entretanto, os resultados relativos a taxa de verdadeiros positivos obtidos por Macedo et al. [12] sobrepujaram os alcançados neste trabalho, sendo superados apenas pela configuração 4 quando todas as características foram utilizadas.

Comparando os resultados obtidos nas Tabelas 2 e 3 podemos analisar a influência da redução de características no desempenho do classificador. Em relação a especificidade, podemos notar que a configuração 4 apresenta uma melhoria de até 14% desta métrica quando um conjunto reduzido de características é utilizado, na configuração 2 notamos um aumento de 4,4% e avalia-se que as configurações 1 e 3 não sofreram variações significativas deste parâmetro. Nota-se também que a técnica de seleção de características LDA apresentou o menor desempenho de classificação dentre as utilizadas.

Em termos de acurácia observa-se que a redução de características não implicou em melhorias notáveis em relação aos resultados alcançados com o conjunto total de características em nenhuma das configurações avaliadas.

Conclusão

Classificadores SVM foram utilizados para auxiliar na identificação de imagens com a presença de bifurcações em sequências IOCT. O método foi

validado em um conjunto de dados oriundos de um conjunto robusto de 1460 imagens.

Embora técnicas de redução de dimensionalidade sejam úteis no que diz respeito a remoção de redundância do vetor de entrada e filtragem de ruídos, neste trabalho, o conjunto completo de 104 características apresentou os melhores desempenhos em relação aos obtidos com conjuntos reduzidos de características através das técnicas PCA, LDA e seleção escalar. O melhor resultado obtido (acurácia de 90,95%) supera os resultados apresentados na literatura.

Agradecimentos

Os autores gostariam de agradecer a Fundação de Amparo à Pesquisa do Amazonas (FAPEAM), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Samsung Eletrônica da Amazônia Ltda., através do Projeto de pesquisa e formação de RH nos termos da Lei Federal 8.387/91 pelo apoio financeiro.

Referências

- [1] WHO. "Cardiovascular diseases (CVDs)," 2016; <http://www.who.int/mediacentre/factsheets/fs317/en/>.
- [2] G. K. Hansson, "Inflammation, atherosclerosis, and coronary artery disease," *New England Journal of Medicine*, vol. 352, no. 16, pp. 1685-1695, 2005.
- [3] C. Zarins et al., "Carotid bifurcation atherosclerosis. Quantitative correlation of plaque localization with flow velocity profiles and wall shear stress," 1983.
- [4] A. Wahle et al., "Fusion of angiography and intravascular ultrasound in vivo: establishing the absolute 3-D frame orientation," *Biomedical Engineering, IEEE Transactions on*, vol. 46, no. 10, pp. 1176-1180, 1999.
- [5] H. G. Bezerra et al., "Intracoronary Optical Coherence Tomography: A Comprehensive Review Clinical and Research Applications," *JACC: Cardiovascular Interventions*, vol. 2, no. 11, pp. 1035-1046, 2009.
- [6] A. Wang et al., "3D assessment of stent cell size and side branch access in intravascular optical coherence tomographic pullback runs," *Computerized Medical Imaging and Graphics*, vol. 38, no. 2, pp. 113-122, 3//, 2014.
- [7] A. Elbasiony, and H. Levkowitz, "Characterization of Atherosclerosis Plaque in OCT Images Using Texture Analysis and Parametric Equations." pp. 237-240.
- [8] S. Tu et al., "In vivo comparison of arterial lumen dimensions assessed by co-registered three-dimensional (3D) quantitative coronary angiography, intravascular ultrasound and optical coherence tomography," *The International Journal of Cardiovascular Imaging*, vol. 28, no. 6, pp. 1315-1327, 2012.
- [9] M. Alberti et al., "Automatic Bifurcation Detection in Coronary IVUS Sequences," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 4, pp. 1022-1031, 2012.

- [10] W. Zhao et al., "3-D Stent Detection in Intravascular OCT Using a Bayesian Network and Graph Search," *Medical Imaging, IEEE Transactions on*, vol. 34, no. 7, pp. 1549-1561, 2015.
- [11] A. Wang et al., "Fully automated side branch detection in intravascular optical coherence tomography pullback runs," *Biomedical optics express*, vol. 5, no. 9, pp. 3160-3173, 2014.
- [12] M. M. G. Macedo et al., "A bifurcation identifier for IV-OCT using orthogonal least squares and supervised machine learning," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 237-248, 2015.
- [13] S. S. Haykin, *Neural Networks and Learning Machines*: Prentice Hall, 2009.
- [14] S. Theodoridis, and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th ed.: Academic Press, 2008.
- [15] C. M. Bishop, *Pattern Recognition and Machine Learning*: Springer, 2006.
- [16] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *International Joint Conference on Artificial Intelligence*, vol. 2, pp. 1137-1143, 1995.
- [17] S. Fengxi, M. Dayong, and L. Hongfeng, "Feature Selection Based on Linear Discriminant Analysis." pp. 746-749.
- [18] F. Song, Z. Guo, and D. Mei, "Feature Selection Using Principal Component Analysis." pp. 27-30

Classification of bifurcations regions in IVOCT images using support vector machine and artificial neural network models

C. D.N. Porto*^a, C. F. F. Costa Filho^a, M.M.G. Macedo^b, M. A. Gutierrez^b, M. G. F. Costa^a.

^aCenter for Research and Development in Electronic and Information Technology, Federal University of Amazonas, Manaus,-AM, Brazil ^bHeart Institute, University of São Paulo, São Paulo-SP, Brazil.

ABSTRACT

Studies in intravascular optical coherence tomography (IV-OCT) have demonstrated the importance of coronary bifurcation regions in intravascular medical imaging analysis, as plaques are more likely to accumulate in this region leading to coronary disease. A typical IV-OCT pullback acquires hundreds of frames, thus developing an automated tool to classify the OCT frames as bifurcation or non-bifurcation can be an important step to speed up OCT pullbacks analysis and assist automated methods for atherosclerotic plaque quantification. In this work, we evaluate the performance of two state-of-the-art classifiers, SVM and Neural Networks in the bifurcation classification task. The study included IV-OCT frames from 9 patients. In order to improve classification performance, we trained and tested the SVM with different parameters by means of a grid search and different stop criteria were applied to the Neural Network classifier: mean square error, early stop and regularization. Different sets of features were tested, using feature selection techniques: PCA, LDA and scalar feature selection with correlation. Training and test were performed in sets with a maximum of 1460 OCT frames. We quantified our results in terms of false positive rate, true positive rate, accuracy, specificity, precision, false alarm, f-measure and area under ROC curve. Neural networks obtained the best classification accuracy, 98.83%, overcoming the results found in literature. Our methods appear to offer a robust and reliable automated classification of OCT frames that might assist physicians indicating potential frames to analyze. Methods for improving neural networks generalization have increased the classification performance.

[1] **Keywords:** Cardiovascular Diseases, Intravascular Optical Coherence Tomography, Neural Networks, Support Vector Machines, Bifurcation, Detection, Coronary

INTRODUCTION

According to World Health Organization¹, cardiovascular diseases (CVD) are the leading cause of death globally. An estimated 17.5 million people died from CVDs in 2012, representing 31% of all global deaths. Of these deaths, an estimated 7.4 million were due to coronary heart disease, also known as atherosclerosis. Atherosclerosis is characterized by the blockage or interruption of blood supply to the heart due to an accumulation of fatty substances, called atheroma, in the coronary arteries. As a result from atherosclerosis, some other diseases may be developed including angina and cerebrovascular accident².

Transitions in artery wall configuration at bends and bifurcations, also named vessel branching, are associated with modifications in rate of the blood flow, which plays a relevant role in the progression of atherosclerosis³. These regions are also predisposed to the development of intimal thickenings, including the formation of atherosclerotic plaques. Atherosclerotic plaques can grow, leading to a significant reduction of the blood flow. Detecting and characterizing these potentially vulnerable regions can help identifying the presence of plaques and understanding the progression of the disease.

For the assessment of atherosclerotic disease, two major medical imaging modalities has been used, coronary angiography (CA) and intravascular ultrasound (IVUS)⁴. Optical coherence tomography (OCT) is a new medical image modality that is becoming important as it has many advantages over other modalities due to its resolution, which ranges from 12 μ m to 18 μ m compared with 150 μ m to 200 μ m for IVUS⁵. OCT is based on the technology of low-coherence interferometry, using scattering of electromagnetic radiation as the signal source. Intravascular OCT is an in-vivo application based on the introduction of an intravascular catheter for viewing the inner wall of blood vessels. For clinical intravascular OCT imaging, one pullback typically contains hundreds of cross-sectional images⁵. Automated bifurcation identification could speed up the process of image analyzing by indicating to the

expert analyst which pullback frames and regions shows a vessel bifurcation and assist methods such as stent analysis⁶, plaque analysis⁷ and co-registration between different modalities⁸.

Some studies found in literature have already proposed methods for bifurcation identification. Alberti *et al.*⁹ extracted a set of textural features from IVUS image sequences and used supervised machine learning methods for identification. Mengliu and Hamaneh¹⁰ used a Random Forest classifier to identify bifurcations by geometrical features in 3D vascular images. Regarding IV-OCT images, Wang *et al.*¹¹ proposed an automated method to identify bifurcation slices for IV-OCT images using the distance between lumen center and lumen contour as a bifurcation indicator. Macedo *et al.*¹² proposed a fully automated method to distinguish frames as bifurcation or nonbifurcation for IV-OCT data sets using geometrical features and machine learning methods.

In this paper, we aim to evaluate different classification methodologies in the task of identifying bifurcation regions in IV-OCT images. This is done by investigating the results of new methods for improving neural networks generalization and the results of SVM classifiers with different kernels over the classification performance of bifurcation slices in human IV-OCT images. The performance of neural networks using three different training stop criteria are evaluated: mean square error criterion, regularization criterion and early stop criterion. Geometrical features compiled by Macedo *et al.*¹² were used as neural network inputs. We also tested the effect of dimensionality reduction of the input data in the classifier performance, using principal component analysis, linear discriminant analysis, and scalar feature selection using correlation techniques. While training and testing the classifier, the 10-fold-cross-validation method was used.

METHODS

Dataset

The data used in this work in order to classify frames as BR or NBR in coronary images is the one compiled by Macedo *et al.*¹². It is, therefore, a set of geometric features extracted from the segmented lumen of IVOCT images.

The coronary images were acquired using a Fourier-Domain OCT (FD-OCT) system (at the Heart Institute, University of Sao Paulo Medical School (INCOR-HC FMUSP). The system is equipped with a Dragonfly catheter with a diameter of 0.90mm, a Guide Wire (GW) with an outer diameter of 0.3556mm (ImagewireTM, LightLab Imaging), and working length of 135cm. 271 frames were acquired in a pullback speed of 20mm/sec over a distance of 54.0mm. Frames corresponding to trunk blood vessel were discarded.

The dataset was divided in three different sets of frames. Set 1 is composed of frames absent of difficulties such as atherosclerotic lesions, presence of blood inside the lumen and dissections, and has 1355 frames. Set 2 is composed of all frames from Set 1 plus complex atherosclerotic lesions and dissections, and has 1433 frames. Set 3 is composed of all frames from Set 2 plus frames with residual luminal blood, and has 1460 frames. Each of these sets are described in Table 1. Examples of IV-OCT images with such classifications difficulties can be found in Figure 1. In order to compare the classification results, the same sets are used in this work.

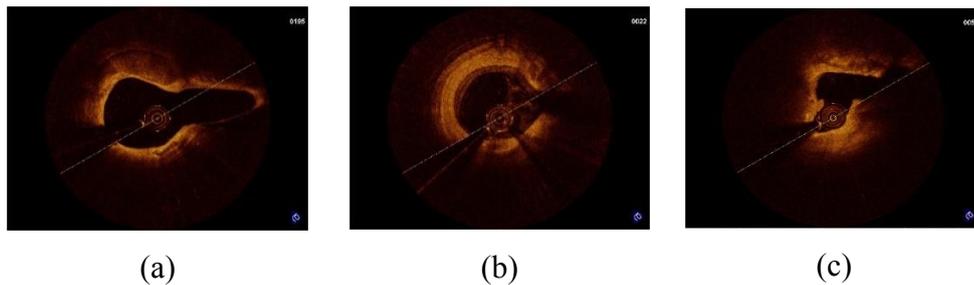


Figure 1. Examples of images of the set (a) with normal bifurcation (b) bifurcation with residual blood (c) bifurcation with complex lesions.

Table 1. Description of sets used in the classification task detailing difficulties and number of frames in each set.

Set	Inherent problems of IV-OCT	#NBR ¹	#BR ²	#Total
Set 1	-	1184	171	1355
Set 2	complex lesions, dissections	1256	177	1433
Set 3	complex lesions, dissections and presence of blood in the lumen	1282	178	1460

¹ NBR – non-bifurcation; ² BR – bifurcation

The hypothesis formulated by Macedo *et al.*¹² is that the cross-section of the blood vessel contains important information that can be used in the characterization and detection of bifurcations. Works found in the literature shows that this information can be extracted from the vessel texture as in Alberti *et al.*⁹ or from the lumen contour as in Wang *et al.*¹¹ Macedo *et al.*¹² used shape descriptors to extract geometrical features from the segmented lumen contour.

The following features were extracted: distance centroid, circularity, bending energy, curvature, radial variance, axis ratio, signature variance, signature standard deviation, sector variance, area difference between frames, mean triangular area. Other descriptors were determined using derivatives and paired combination, totaling 104 different features, excluding square features.

Dataset balancing

The dataset formed by Macedo *et al.*¹² is highly unbalanced since positive classes represent only about 12% of the data in each set, as can be seen in Table 1. Class imbalance often results in a better classification performance for majority classes over minority classes. However, in many cases what matters is to perform well for the minority classes, as in this bifurcation classification problem.

Aiming to solve this problem, we performed a data balancing by oversampling the minority class. This is done by dividing Set 1, Set 2 and Set 3 into seven small subsets each. The subsets are formed dividing non-bifurcation patterns proportionally and then adding the bifurcation patterns to each one of them, thus, forming seven balanced subsets. Figure 2 illustrate this process. In each set (Set 1, Set 2 and Set 3), the classification performance is calculated as a mean performance of the experiments for all the seven subsets.

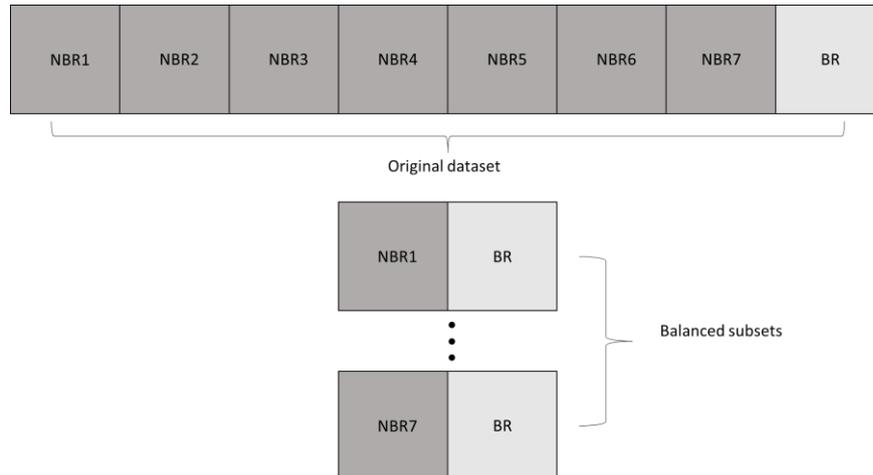


Figure 2. Illustration of the dataset balancing process

K-fold Cross Validation

In k-fold cross-validation, the dataset is randomly split into k equal sized mutually exclusive subsets, named folds. One fold the k folds is retained and the classifier is trained with the k-1 remaining folders. The classifier is then tested with the previously excluded folder. This process is repeated k times, until all the folders have been used to test the classifier. The classification performance from the k folds are averaged in order to produce a single estimation¹³.

In this study, each subset was divided into 10 folders. These folders were used to train and test the SVM classifier and neural networking with mean square error and regularization.

Feature Selection

With the aim of evaluating the effect of a reduced set of variables on the classifiers performance, three feature selection methods were employed: scalar feature selection with correlation¹⁴ (SFS), linear discriminant analysis¹⁵ (LDA) and principal component analysis¹⁶ (PCA). Sets of the best 5 to 20 features computed by each method were used as inputs of neural networks and support vector machines classifiers.

Neural Networks

Multilayers neural networks with nonlinear transfer functions allow the network to learn nonlinear relationships between input and output vectors increasing the space of hypotheses that it can represent and providing great computing power¹⁷. The number of artificial neurons per layer, as well as the number of layers, greatly influences the prediction abilities of the neural network. In this work, four-layer neural networks were employed for each of

the three sets. The designed networks architecture was f-n-n-1. Initially, the number of input variables, f, was set up to 104 (full set of features) and then, we vary f from 5 to 20. Additionally, n was set up to 5, 8 and 10.

The neural network training algorithm used in this work was the Levenberg Marquardt¹⁸. This algorithm approximates the error of the network with a second order expression, which contrasts to the former category that follows a first order expression.

The prediction error is minimized across many training cycles, known as epochs, until the network reaches specific level of accuracy. If a network is left to train for too long, however, it will be over trained, and will lose the ability to generalize. Three stop training criteria were employed for neural network training: mean square error¹⁹ (MSE), regularization¹⁹ (REG) and early stop²⁰ (ES).

With the mean square error criterion, the training was finished when its value reached 10e-6 or 1000 epochs. With the regularization criterion, aiming to work with more stable neural networks (networks with lower weights), a new term, proportional to the sum of the squared network weights, was added to the mean square error, according to expression (1):

$$msereg = \gamma mse + (1 - \gamma)msw \quad (1)$$

where:

γ - performance factor. A number between 0 and 1.

mse - mean square error. In this paper, $\gamma=0.5$.

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (2)$$

The regularization criterion in expression (1) causes lower neural network weights, enforcing a smooth network response and improving the generalization power of the neural network.

For neural network training and testing with MSE and REG criteria, the 10-fold-cross-validation was used in each of the seven subsets. In 10-fold-cross-validation, each subset is divide into ten groups. The neural network is then trained with data from nine groups and tested with data of only one group. This process is repeated 10 times. A different group is used for testing each time. The neural network performance is calculated as a mean performance of all these experiments.

With the early stop training criterion, each subset is divided into three groups: training, validation and testing. The main characteristic of this method is that during the training phase, although the validation group is not used, the mean square error is evaluated on it. When the mean square error grows in this data group, the neural network training is stopped. The neural network performance is calculated as a mean performance of the experiments for all the 7 testing subsets.

Support Vector Machines

SVM separates patterns belonging to two classes defining one hyperplane that maximizes the separating margin between these two classes. According to Theodoridis and Koutroumbas¹⁴, the hyperplane parameters that maximize the separating margin are the weight vector w and polarization w_0 that minimizes equation (3) and satisfies equation (4):

$$J(w, w_0) = \frac{1}{2} \|w\|^2 \quad (3)$$

$$y_i(w^T x_i + w_0) \geq 1, i = 1, 2, \dots, N, \quad (4)$$

For non-separating classes, the same parameters could be determined, minimizing the equation (5), where new variables, ξ_i , known as slack variables, are introduced. The optimizing task becomes more complex. The goal now is to make the margin as large as possible, but at the same time keep the number of points with $\xi > 0$ as small as possible.

$$J(w, w_0, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (5)$$

The C parameter in equation (5) is a positive constant that controls the relative influence of the two competing terms. SVMs use kernels for mapping characteristic vectors to a large dimension space vector where classes could be separated by hyperplanes. In this work polynomial function (P) and Gaussian radial basis function (GRBF) were used in association with the SVM classifier.

Simulations were carried out with each subset of features obtained in the feature selection step and with the original set, which includes all features, using the kernels mentioned above varying their degrees from 1 to 5. The values of C used to aid selecting the best classifier vary from 2e-5 to 2e8.

RESULTS

Both neural networks and SVM classifiers were trained and tested using the three sets described. The frames were previously classified by an expert as pertaining to BR or NBR regions. For SVM classifiers and neural networks with MSE and REG criteria, the classification performance is assessed by means of k-fold-cross-validation technique over $k = 10$. The classification performance was evaluated in terms of the following metrics:

- False positive rate (FPR)

$$FPR = \frac{FP}{TN + FP}$$

- True positive rate (TPR)

$$TPR = \frac{TP}{TP + FN}$$

- Accuracy (ACC)

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

- Specificity (SPE)

$$SPE = \frac{TN}{TN + FP}$$

- Precision (PRE)

$$PRE = \frac{TP}{TP + FP}$$

- False Alarm (FA)

$$FA = \frac{FP}{TP + FN}$$

- F-measure (F)

$$F = \frac{2 \cdot PRE \cdot TPR}{(PRE + TPR)}$$

- Area under ROC curve (AUC)

$$AUC = \int_0^1 f(x) dx$$

where TP= True Positive, FP= false positive, TN= true negative, FN= false negative, f is the ROC curve and x is the TPR value.

It is important to mention that in a detection task, the most significant parameter is the true positive rate (TPR) since it expresses the proportion of actual bifurcation samples, which are correctly identified as such.

For each simulation using SVM classifiers, a different combination of feature set, kernel, σ , and C was employed. Table 1 shows the best results obtained with SVM classifiers with two different kernels, Gaussian radial bases function (GRBF) and polynomial function (P). The three sets were tested, all 104 features were used as input variables.

We chose the two best models obtained from each kernel to present in our results. Although Set 03 is the one with most classification difficulties, the best TPR value, 95.59%, was obtained in this set, with GRBF function. The best accuracy, 91.96%, was also obtained with GRBF kernel, in Set 01.

Table 1 – SVM classification performance using two different models (M) for each kernel, polynomial (P) and Gaussian RBF function (GRBF). All 104 features were used as input to the classifiers.

Kernel	M	FPR (%)	TPR (%)	ACC (%)	SPE (%)	PRE (%)	FA (%)	F (%)	AUC
Set 01									
P	1	6.82 ± 3.65	88.8 ± 6.06	91.14 ± 4.68	93.18 ± 3.65	92.14 ± 4.52	7.89 ± 4.35	90.19 ± 5.3	0.91 ± 0.05
	2	5.97 ± 2.85	82.82 ± 6.31	88.83 ± 4.35	94.03 ± 2.85	92.57 ± 3.79	6.93 ± 3.4	87 ± 5.34	0.88 ± 0.05
GRBF	1	6.23 ± 3.75	88.97 ± 5.56	91.55 ± 4.21	93.77 ± 3.75	92.75 ± 4.24	7.21 ± 4.41	90.62 ± 4.69	0.91 ± 0.04
	2	6.48 ± 3.72	90.18 ± 5.25	91.96 ± 4.08	93.52 ± 3.72	92.54 ± 4.28	7.52 ± 4.39	91.13 ± 4.57	0.92 ± 0.04
Set 02									
Polynomial	1	6.99 ± 4.36	87.25 ± 6.45	90.37 ± 5.06	93.01 ± 4.36	91.59 ± 5.14	8.31 ± 5.15	89.08 ± 5.77	0.9 ± 0.5
	2	5.8 ± 4.4	81.04 ± 10.5	88.16 ± 7.02	94.2 ± 4.4	92 ± 6.39	6.9 ± 5.22	85.75 ± 8.72	0.88 ± 0.07
GRBF	1	6,61 ± 3,64	87,65 ± 6,56	90,76 ± 4,83	93,39 ± 3,64	92,17 ± 4,33	7,84 ± 4,34	89,58 ± 5,54	0,91 ± 0,05
	2	11,15 ± 4,05	91,8 ± 4,37	90,21 ± 3,33	88,85 ± 4,05	88,01 ± 4,1	13,18 ± 4,8	89,57 ± 3,57	0,9 ± 0,03
Set 03									
Polynomial	1	7.44 ± 4.07	86.35 ± 5.21	89.76 ± 4.38	92.56 ± 4.07	91.1 ± 4.95	8.88 ± 4.87	88.34 ± 5.03	0.89 ± 0.04
	2	9.7 ± 4.85	79.59 ± 4.43	85.42 ± 4.27	90.3 ± 4.85	87.94 ± 5.47	11.62 ± 5.7	83.1 ± 4.65	0.85 ± 0.04
GRBF	1	20.76 ± 6.3	95.59 ± 2.95	86.67 ± 4.09	79.24 ± 6.35	80.24 ± 5.35	24.96 ± 7.6	86.93 ± 3.81	0.87 ± 0.04
	2	7.75 ± 3.37	88.58 ± 5.87	90.58 ± 4.11	92.25 ± 3.37	90.89 ± 3.9	9.31 ± 3.99	89.48 ± 4.66	0.9 ± 0.05

Table 2 shows the best classification results obtained using SVM classifiers when reduced sets of features composed from the best 5 to 20 features, selected with PCA, LDA, and SFS techniques are used as input. The best TPR value obtained was 88.46% (Set 3) with the GRBF kernel, using 12 features, selected by SFS.

Regarding neural networks, the performance of the three training stop criteria, mean square error, regularization and early stop with all 104 features and different architectures, over Sets 1, 2, and 3 are shown in Table 3. Our best accuracy (98.15%) and TPR (97.36%) were obtained with the mean square error criterion, in Set 1. For Set 3, the set with most complex frames, our best accuracy and TPR values were 97.53% and 96.53% respectively, with the regularization criterion.

Table 2– Best classification results obtained using SVM with two different models (M) for each kernel, Polynomial (P) and GRBF with different sets of features (f) obtained by feature selection techniques (FST) PCA, LDA and SFS for sets 1, 2, and 3.

Kernel	M	FST	f	FPR (%)	TPR (%)	ACC (%)	SPE (%)	PRE (%)	FA (%)	F (%)	AUC
Set 01											
GRBF	1	PCA	14	5.48 ± 3.46	84.92 ± 6.16	90.07 ± 4.6	94.52 ± 3.46	93.38 ± 4.23	6.35 ± 4.09	88.65 ± 5.37	0.9 ± 0.05
	1	LDA	19	6.08 ± 4.99	81.11 ± 8.42	87.97 ± 6.07	93.92 ± 4.99	92.39 ± 6.21	7.01 ± 5.83	86.01 ± 7.09	0.88 ± 0.06
	1	SFS	17	6.56 ± 4.36	88.1 ± 6.41	90.97 ± 5.16	93.44 ± 4.36	92.38 ± 5.23	7.6 ± 5.14	89.84 ± 5.96	0.91 ± 0.05
	2	PCA	15	5.73 ± 4.06	84.8 ± 6.35	89.87 ± 4.92	94.27 ± 4.06	93.02 ± 5.06	6.64 ± 4.78	88.44 ± 5.7	0.9 ± 0.05
	2	LDA	20	6.15 ± 4.22	81.27 ± 7.94	88.25 ± 5.81	94.29 ± 5.02	92.79 ± 6.35	6.63 ± 5.89	86.22 ± 6.83	0.88 ± 0.06
	2	SFS	17	6.8 ± 4.57	87.55 ± 6.85	90.69 ± 5.47	93.44 ± 4.45	92.11 ± 5.56	7.6 ± 5.25	89.57 ± 6.25	0.9 ± 0.06
Set 02											
P	1	PCA	14	5.89 ± 3.75	83.82 ± 5.93	89.33 ± 4.52	94.11 ± 3.75	92.76 ± 4.68	6.82 ± 4.42	87.81 ± 5.39	0.89 ± 0.05
	1	LDA	15	6.15 ± 4.64	83.02 ± 8.71	88.65 ± 6.16	93.85 ± 4.64	92.31 ± 5.96	7.12 ± 5.44	86.83 ± 7.32	0.88 ± 0.06
	1	SFS	18	6.81 ± 4.51	87.82 ± 6.62	90.68 ± 5.36	93.19 ± 4.51	92.01 ± 5.49	7.91 ± 5.33	89.62 ± 6.03	0.91 ± 0.05
	2	PCA	14	5.26 ± 4.44	72.6 ± 7.7	84.6 ± 5.68	94.74 ± 4.44	92.45 ± 6.2	6.23 ± 5.28	80.91 ± 7.26	0.84 ± 0.06
	2	LDA	17	6.61 ± 3.37	72.53 ± 10.9	83.83 ± 6.42	93.39 ± 3.37	90.44 ± 5	7.81 ± 3.97	79.93 ± 8.51	0.83 ± 0.07
	2	SFS	15	7.17 ± 5.53	82.7 ± 8.65	88.17 ± 6.73	92.83 ± 5.53	91.03 ± 6.89	8.53 ± 6.63	86.32 ± 7.78	0.88 ± 0.07
Set 03											

GRBF	1	PCA	19	10.31 ± 5.4	88.28 ± 5.39	89.05 ± 4.94	89.69 ± 5.47	88.19 ± 5.96	12.39 ± 6.6	88.02 ± 5.29	0.89 ± 0.05
	1	LDA	20	8.12 ± 3.82	78.5 ± 7.74	85.82 ± 4.92	91.88 ± 3.82	89.33 ± 5.03	9.74 ± 4.58	83.06 ± 6.15	0.85 ± 0.05
	1	SFS	12	9.59 ± 4.4	88.46 ± 5.73	89.53 ± 4.91	90.41 ± 4.4	88.9 ± 5.13	11.55 ± 5.2	88.4 ± 5.45	0.89 ± 0.05
	2	PCA	16	5.94 ± 3.24	82.83 ± 7.11	89.2 ± 4.63	94.16 ± 3.62	92.48 ± 4.65	7.04 ± 4.38	87.32 ± 5.57	0.89 ± 0.05
	2	LDA	20	7.85 ± 3.85	72.28 ± 7.77	83.1 ± 5.14	92.15 ± 3.85	88.44 ± 5.89	9.52 ± 4.66	79.14 ± 6.87	0.82 ± 0.05
	2	SFS	20	7.33 ± 4.56	86.8 ± 6.76	89.99 ± 5.4	92.67 ± 4.56	91.05 ± 5.59	8.82 ± 5.47	88.67 ± 6.15	0.9 ± 0.05

Table 3 –Classification results according to the architecture (N) and stop criterion (SC) of neural network, using all 104 features (f) as input to the classifiers

SC	f	N (n-n-1)	FPR (%)	TPR (%)	ACC (%)	SPE (%)	P (%)	FA (%)	F (%)	AUC
Set 01										
MSE	104	5-5-1	5.28 ± 5.66	93.57 ± 3.75	94.19 ± 4.46	94.72 ± 5.66	94.32 ± 5.8	6.13 ± 6.62	93.73 ± 4.72	0.95 ± 0.04
	104	8-8-1	1.7 ± 1.47	97.36 ± 0.81	98.15 ± 0.79	98.83 ± 1.47	98.7 ± 1.56	1.37 ± 1.73	97.93 ± 0.81	0.99 ± 0.01
	104	10-10-1	2.62 ± 2.58	95.52 ± 2.36	96.51 ± 2.41	97.38 ± 2.58	97.14 ± 2.91	3.02 ± 3.02	96.17 ± 2.56	0.98 ± 0.02
REG	104	5-5-1	5.82 ± 4.83	95.81 ± 2.33	94.94 ± 3.52	94.18 ± 4.83	93.93 ± 4.9	6.75 ± 5.68	94.7 ± 3.6	0.96 ± 0.02
	104	8-8-1	3.46 ± 1.54	95.23 ± 2.65	95.93 ± 1.88	96.54 ± 1.54	96.19 ± 1.74	3.98 ± 1.8	95.59 ± 2.09	0.97 ± 0.02
	104	10-10-1	2.98 ± 2.53	94.54 ± 3.17	95.87 ± 2.72	97.02 ± 2.53	96.65 ± 2.84	3.42 ± 2.91	95.44 ± 3.01	0.97 ± 0.03
ES	104	5-5-1	3.14 ± 5.42	96.09 ± 3.18	96.41 ± 3.63	96.86 ± 5.42	96.19 ± 6.45	4.17 ± 7.19	96.04 ± 4.04	0.98 ± 0.03
	104	8-8-1	4.15 ± 2.98	96.85 ± 4.99	96.4 ± 3.5	95.85 ± 2.98	95.37 ± 3.51	4.67 ± 3.42	96.09 ± 4.15	0.99 ± 0.01
	104	10-10-1	7.2 ± 5.8	91.1 ± 9.16	92.09 ± 6.31	92.8 ± 5.8	90.48 ± 7.7	9.74 ± 8.31	90.67 ± 7.75	0.96 ± 0.04
Set 02										
MSE	104	5-5-1	0.9 ± 0.03	90.15 ± 4.72	87.41 ± 4.01	85.09 ± 6.7	84.43 ± 6.18	17.59 ± 7.8	86.89 ± 4.01	0.9 ± 0.03
	105	8-8-1	2.79 ± 2.18	95.88 ± 1.48	96.59 ± 1.46	97.21 ± 2.18	96.88 ± 2.2	3.32 ± 2.57	96.22 ± 1.56	0.98 ± 0.01
	104	10-10-1	2.71 ± 1.89	95.6 ± 2.05	96.51 ± 1.93	97.29 ± 1.89	97.01 ± 2.17	3.18 ± 2.23	96.18 ± 2.09	0.98 ± 0.01
REG	104	5-5-1	5.25 ± 3.02	93.61 ± 3.29	95.52 ± 3.6	94.75 ± 3.02	94.23 ± 3.4	6.23 ± 3.59	93.67 ± 3.32	0.97 ± 0.02
	104	8-8-1	4.14 ± 3.86	95.36 ± 3.31	95.64 ± 3.53	95.86 ± 3.86	95.43 ± 4.2	4.9 ± 4.58	95.24 ± 3.8	0.98 ± 0.01
	104	10-10-1	2.56 ± 1.76	95.2 ± 2.03	96.41 ± 1.8	97.44 ± 1.76	97.11 ± 2.04	3.03 ± 2.12	96.05 ± 1.98	0.98 ± 0.02
ES	104	5-5-1	4.87 ± 4.26	93.4 ± 6.38	94.32 ± 4.31	95.13 ± 4.26	94.48 ± 4.77	5.54 ± 4.82	93.85 ± 4.83	0.98 ± 0.02
	104	8-8-1	1.86 ± 1.64	90.34 ± 10.6	94.5 ± 5.03	98.14 ± 1.64	97.48 ± 2.26	2.35 ± 2.17	93.49 ± 6.24	0.98 ± 0.02
	104	10-10-1	4.64 ± 2.77	89.15 ± 8.84	92.77 ± 4.71	95.36 ± 2.77	93.65 ± 4.18	5.92 ± 3.71	91.25 ± 6.36	0.97 ± 0.02
Set 03										
MSE	104	5-5-1	6.98 ± 4.27	94.79 ± 2.62	93.81 ± 3.02	93.02 ± 4.27	92.4 ± 4.27	8.41 ± 5.25	93.33 ± 3.15	0.96 ± 0.02
	104	8-8-1	2.66 ± 1.85	96.21 ± 1.47	96.83 ± 1.47	97.34 ± 1.85	96.97 ± 2.06	3.22 ± 2.23	96.45 ± 1.61	0.98 ± 0.02
	104	10-10-1	2.59 ± 1.79	95.33 ± 2.68	96.46 ± 1.99	97.41 ± 1.79	97.04 ± 1.97	3.09 ± 2.1	96.05 ± 2.17	0.98 ± 0.02
REG	104	5-5-1	6.94 ± 2.48	93.28 ± 2.86	93.16 ± 1.81	93.06 ± 2.48	92.35 ± 2.21	8.32 ± 3.03	92.51 ± 2.07	0.94 ± 0.02
	104	8-8-1	3.76 ± 0.89	95.97 ± 4.49	96.12 ± 2.39	96.24 ± 0.89	95.72 ± 1.2	4.5 ± 1.05	95.68 ± 2.83	0.98 ± 0.01
	104	10-10-1	1.63 ± 1.18	96.53 ± 1.54	97.53 ± 1.27	98.37 ± 1.18	98.04 ± 1.48	1.98 ± 1.44	97.19 ± 1.43	0.98 ± 0.01
ES	104	5-5-1	6.97 ± 3.07	89.5 ± 7.78	91.34 ± 5.12	93.03 ± 3.07	91.64 ± 3.91	8.08 ± 3.73	90.49 ± 5.65	0.96 ± 0.03
	104	8-8-1	3.36 ± 3.81	90.36 ± 7.71	94.05 ± 4.06	96.64 ± 3.81	95.58 ± 4.18	4.27 ± 4.21	92.75 ± 4.9	0.97 ± 0.02
	104	10-10-1	7.12 ± 7.06	94.64 ± 3.56	93.88 ± 3.98	92.88 ± 7.06	92.38 ± 5.16	8.05 ± 6.05	93.43 ± 3.57	0.97 ± 0.02

Table 4 shows the classification performance for the best 5 to 20 features selected with PCA, LDA and SFS with mean square error (MSE), early stop (ES) and regularization (REG) training stop criterion. Different architectures were tested in order to classify Sets 1, 2 and 3. For each set, only the best results for each feature selection technique are shown.

Table 4 – Classification results according to the architecture, where (f) the number of input characteristics and (N) the number of neurons in the hidden layer, the stop criterion (SC) of neural network and the feature selection technique (FST).

SC	f	N (n-n-1)	FST	FPR (%)	TPR (%)	ACC (%)	SPE (%)	PRE (%)	FA (%)	F (%)	AUC
Set 01											
MSE	16	10	PCA	2.85 ± 2.41	94.83 ± 2.94	96.07 ± 2.61	97.15 ± 2.41	96.63 ± 2.89	3.31 ± 2.82	95.66 ± 2.9	0.97 ± 0.02
	20	10	LDA	3.12 ± 2.34	92.48 ± 4.3	94.85 ± 2.98	96.88 ± 2.34	96.44 ± 2.59	3.59 ± 2.77	94.22 ± 3.43	0.97 ± 0.02
	20	10	SFS	2.94 ± 2.59	96.29 ± 1.92	96.69 ± 2	97.06 ± 2.59	96.78 ± 2.73	3.42 ± 3.03	96.44 ± 2.08	0.98 ± 0.01
REG	16	10	PCA	2.27 ± 1.71	95.92 ± 2.67	96.89 ± 2.02	97.73 ± 1.71	97.45 ± 1.89	2.63 ± 2.03	96.56 ± 2.27	0.98 ± 0.02
	16	10	LDA	4.42 ± 2.27	94.53 ± 1.9	95.1 ± 1.94	95.58 ± 2.27	95.08 ± 2.46	5.09 ± 2.63	94.66 ± 2.03	0.97 ± 0.01
	20	10	SFS	1.75 ± 2.68	95.33 ± 4.47	96.89 ± 3.5	98.25 ± 2.68	97.89 ± 3.25	2.04 ± 3.14	96.48 ± 3.95	0.98 ± 0.04
ES	5	8	PCA	2.24 ± 2.06	86.48 ± 8.88	92.08 ± 5.31	97.76 ± 2.06	97.57 ± 2.21	2.11 ± 1.82	91.52 ± 5.5	0.96 ± 0.03
	8	8	LDA	3.29 ± 2.83	91.69 ± 3.66	94.4 ± 2.92	96.71 ± 2.83	96.55 ± 2.7	3.28 ± 2.6	94.04 ± 2.87	0.98 ± 0.02
	11	8	SFS	3.78 ± 3.15	92.3 ± 7.88	94.42 ± 5.32	96.22 ± 3.15	95.51 ± 3.61	4.24 ± 3.18	93.81 ± 5.62	0.98 ± 0.02
Set 02											
MSE	18	10	PCA	3.9 ± 2.08	93.75 ± 1.43	95.03 ± 1.69	96.1 ± 2.08	95.51 ± 2.2	4.61 ± 2.42	94.51 ± 1.78	0.97 ± 0.02
	20	10	LDA	4.21 ± 4.09	93.98 ± 3.02	94.96 ± 3.54	95.79 ± 4.09	95.3 ± 4.54	4.98 ± 4.85	94.5 ± 3.78	0.96 ± 0.03
	18	10	SFS	5.08 ± 4.28	95.94 ± 1.21	95.38 ± 2.5	94.92 ± 4.28	94.68 ± 4.28	6.03 ± 5.1	95.11 ± 2.47	0.97 ± 0.02
REG	20	10	PCA	4.94 ± 2.53	91.05 ± 3.54	93.23 ± 2.85	95.06 ± 2.53	94.14 ± 2.83	5.82 ± 2.97	92.36 ± 3.13	0.96 ± 0.02
	17	10	LDA	4.85 ± 4.23	93.83 ± 1.84	94.55 ± 2.77	95.15 ± 4.23	94.53 ± 4.11	5.78 ± 5.05	94.03 ± 2.76	0.96 ± 0.02
	19	10	SFS	3.17 ± 2.37	95.79 ± 1.41	96.34 ± 1.84	96.83 ± 2.37	96.35 ± 2.69	3.77 ± 2.76	95.97 ± 1.98	0.97 ± 0.01
ES	13	5	PCA	3.61 ± 3.71	90.76 ± 5.71	93.81 ± 3.63	96.39 ± 3.71	95.8 ± 3.95	4 ± 3.87	93.16 ± 4.4	0.96 ± 0.03
	7	10	LDA	3.56 ± 1.99	91.22 ± 5.88	94.15 ± 2.96	96.44 ± 1.99	95.61 ± 2.58	4.15 ± 2.34	93.31 ± 4.02	0.98 ± 0.01
	17	10	SFS	4.45 ± 2.85	93.57 ± 4.45	94.66 ± 3.09	95.55 ± 2.85	93.8 ± 4.13	6.24 ± 4.22	93.65 ± 3.83	0.98 ± 0.02
Set 03											
MSE	20	10	PCA	3.53 ± 2.92	92.47 ± 3.65	94.66 ± 2.96	96.47 ± 2.92	95.79 ± 3.33	4.22 ± 3.47	93.9 ± 3.37	0.95 ± 0.03
	18	10	LDA	4.75 ± 2.78	93.64 ± 3.22	94.52 ± 2.62	95.25 ± 2.78	94.52 ± 3.12	5.69 ± 3.3	93.92 ± 2.94	0.96 ± 0.03
	18	10	SFS	3.52 ± 2.5	94.31 ± 2.33	95.48 ± 1.94	96.48 ± 2.5	96.1 ± 2.5	4.22 ± 2.97	94.98 ± 2.09	0.96 ± 0.02
REG	18	8	PCA	3.66 ± 3.18	92.29 ± 3.31	94.5 ± 2.56	96.34 ± 3.18	95.52 ± 3.8	4.4 ± 3.82	93.67 ± 2.92	0.96 ± 0.01
	19	10	LDA	6.03 ± 3	92.14 ± 3.95	93.13 ± 3.35	93.97 ± 3	92.78 ± 3.79	7.24 ± 3.61	92.35 ± 3.82	0.95 ± 0.03
	20	10	SFS	3.75 ± 1.61	94.68 ± 3.53	95.53 ± 2.35	96.25 ± 1.61	95.78 ± 1.85	4.52 ± 1.93	95.05 ± 2.67	0.97 ± 0.03
ES	19	8	PCA	5.24 ± 5.17	91.25 ± 5.42	93.21 ± 4.3	94.76 ± 5.17	93.61 ± 6.29	6.49 ± 6.42	92.31 ± 4.8	0.97 ± 0.02
	18	10	LDA	2.84 ± 2.93	90.02 ± 7.87	93.88 ± 4.92	97.16 ± 2.93	96.34 ± 3.71	3.33 ± 3.3	93 ± 5.71	0.97 ± 0.03
	7	8	SFS	3.45 ± 2.27	89.11 ± 8.07	93.03 ± 4.5	96.55 ± 2.27	95.6 ± 2.56	4.12 ± 2.53	92.08 ± 4.85	0.98 ± 0.02

DISCUSSION

Two state-of-the-art classifiers were considered in this work: neural networks and SVM. Three methods for improving neural networks generalization were used: mean square error, regularization and early stopping.

Regarding SVM classifiers, a comparison between the results presented in Tables 1 and 2 shows that the classification performance with the whole set of features is better than the performance obtained using reduced sets of features. Whilst the higher TPR using 104 features is equal to 95.59%, using feature selection techniques the higher TPR is 88.46%. The same can be noticed with other metrics, as the accuracy (down from 91.96% to 90.97%). The same behavior is noticed with the neural networks classifiers. The TPR values falls from 97.36% using the whole set of features to 96.29% using a reduced set of features, the Accuracy falls from 98.15% to 96.89%.

Although dimensionality reduction techniques are useful concerning the redundancy removal of the input vector and noise filtering, in this work, the complete set of features showed the best performance compared to those obtained with reduced sets of characteristics.

In general, Neural Networks classifiers performed better than SVM due to the methods used for improving neural networks generalization performance. The best accuracy (98.15%) and true positive ratio (97.36%) values were obtained with the mean square error criterion. Figures 3 to 5 show a comparison between TPR values obtained with neural networks and SVM classifiers for Set 1, Set 2 and Set 3, using full set of features as input.

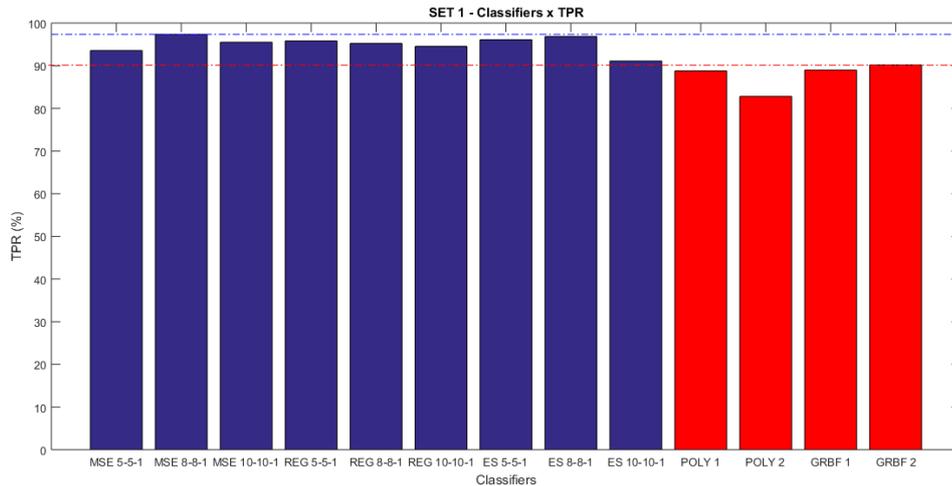


Figure 3. Set 1: TPR values obtained for each classifier using the whole 104 features as input. Blue bars are related to neural networks and red bars to SVM classifiers. The blue line indicates the maximum TPR value for neural network classifiers (97.36%) and the red lines the maximum TPR value for SVM (90.18%) classifiers.

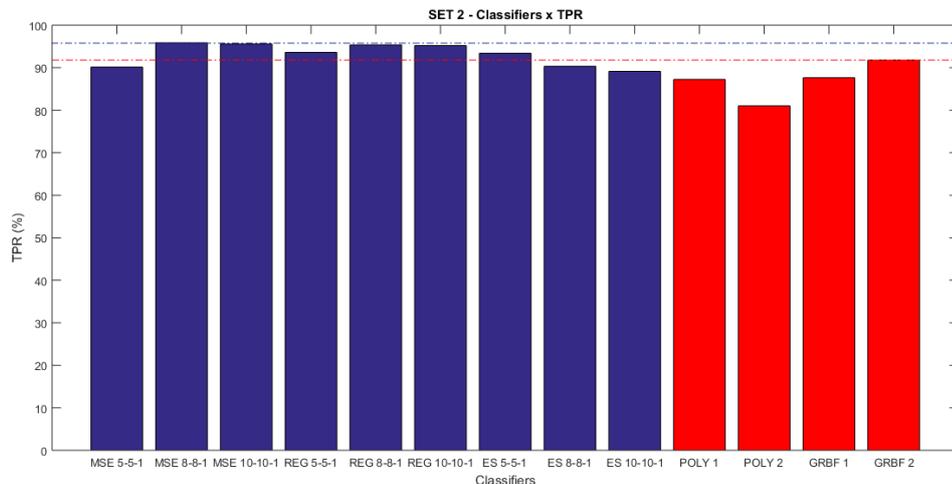


Figure 4. Set 2: TPR values obtained for each classifier using the whole 104 features as input. Blue bars are related to neural networks and red bars to SVM classifiers. The blue line indicates the maximum TPR value for neural network classifiers (95.88%) and the red lines the maximum TPR value for SVM (91.8%) classifiers.

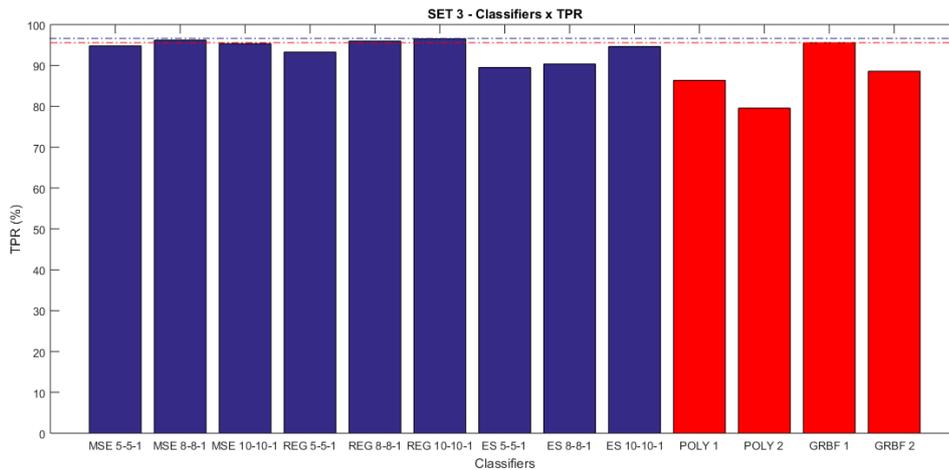


Figure 5. Set 3: TPR values obtained for each classifier using the whole 104 features as input. Blue bars are related to neural networks and red bars to SVM classifiers. The blue line indicates the maximum TPR value for neural network classifiers (96.53%) and the red lines the maximum TPR value for SVM (95.59%) classifiers.

From Figures 3 to 5, one can see that in all three sets, neural networks performed better than SVM in terms of TPR. We performed a statistical significance test of the area under the ROC curve²¹ to compare the results obtained with Neural Networks and SVM, leading to a $p = 0.001 < 0.05$, (i.e., the neural network classifier with early stop criterion is superior to SVM classifier, at the 5% significance level).

The neural network configuration that achieved better classification performance was the f-8-8-1, with mean square error criteria.

From Table 5 we conclude that our classification results are comparable to other results in the literature. The classification with neural networks show a better performance in terms of accuracy than any other method suggested in literature.

Table 5 – Bifurcation detection: comparison with other methods in literature in terms of true positive rate (TPR) and accuracy (ACC).

Article	Year	Image Type	Technique	TPR	ACC
Alberti <i>et al.</i>	2011	IVUS	AdaBoost	75,09± 13.7%	94±4.5%
Baboiu and Hamarneh	2012	Retinal angiography	Bifurcationness filter	95%±0.4	95% ~97%
Wang <i>et al.</i>	2014	IVOCT	Distance computing	94%	Not presented
Macedo <i>et al.</i>	2015	IVOCT	Adaboost	98,89±0,67%	95.58±2.08%
This work	2016	IVOCT	Neural Networks - MSE	97.36±0.81%	98.15 ± 0.79

The best AUC value obtained in this study, 0.99, is better than the value of 0.96 obtained by Macedo *et al.* (2015) using samples of a same population. Assessing the difference between the areas that lie under these two ROC curves, we found that $p = 0.011 < 0.05$, the null hypothesis should be rejected (i.e., the AUC obtained in this work is superior to the value of AUC obtained in the work of Macedo *et al.* (2015), at the 5% level).

CONCLUSION

In this paper, we presented different models for the task of bifurcation classification in IV-OCT sequences based in neural networks and SVM. These models are a combination of supervised machine learning and features selection using PCA, LDA and scalar feature selection techniques. We trained and tested the classifiers with three different sets of human IV-OCT frames. Neural networks classifiers showed a better classification performance than SVM classifiers. With neural networks in association with the mean square error criterion, we achieved the best TPR and accuracy values than any other method suggested in literature. Our methods appear to offer a robust and reliable automated classification of OCT frames tool that might assist physicians indicating potential frames to analyze.

ACKNOWLEDGEMENT

The authors would like to thank FAPEAM (process no. 062.00938/2015- PECTI/AM-PG) and CAPES for their financial support.

REFERENCES

- WHO, "Cardiovascular diseases (CVDs)" World Health Organization, September 2016, <<http://www.who.int/mediacentre/factsheets/fs317/en/>> (21 November 2016).
- G. K. Hansson, "Inflammation, atherosclerosis, and coronary artery disease," *New England Journal of Medicine*, vol. 352 1685-1695 (2005).
- C. Zarins, D. Giddens, B. Bharadvaj, V. Sottiurai, R. Mabon, and S. Glagov, "Carotid bifurcation atherosclerosis. Quantitative correlation of plaque localization with flow velocity profiles and wall shear stress," *Circulation Research*. Paper 53, 502-514 (1983).
- A. Wahle, G. P. M. Prause, C. Von Birgelen, R. Erbel, and M. Sonka, "Fusion of angiography and intravascular ultrasound in vivo: establishing the absolute 3-D frame orientation," *Biomedical Engineering, IEEE Transactions on*, vol. 46, 1176-1180 (1999).
- H. G. Bezerra, M. A. Costa, G. Guagliumi, A. M. Rollins, and D. I. Simon, "Intracoronary Optical Coherence Tomography: A Comprehensive Review Clinical and Research Applications," *JACC: Cardiovascular Interventions*, vol. 2, 1035-1046 (2009).
- A. Wang, J. Eggermont, N. Dekker, P. J. H. de Koning, J. H. C. Reiber, and J. Dijkstra, "3D assessment of stent cell size and side branch access in intravascular optical coherence tomographic pullback runs" *Computerized Medical Imaging and Graphics*, vol. 38 113-122, (2014).
- A. Elbasiony and H. Levkowitz, "Characterization of Atherosclerosis Plaque in OCT Images Using Texture Analysis and Parametric Equations," *Information Visualisation (IV)*, 2011 15th International Conference on. 237-240 (2011).
- S. Tu, L. Xu, J. Ligthart, B. Xu, K. Witberg, Z. Sun, et al., "In vivo comparison of arterial lumen dimensions assessed by co-registered three-dimensional (3D) quantitative coronary angiography, intravascular ultrasound and optical coherence tomography," *The International Journal of Cardiovascular Imaging*, vol. 28, 1315-1327 (2012).
- M. Alberti, S. Balocco, C. Gatta, F. Ciompi, O. Pujol, J. Silva, et al., "Automatic Bifurcation Detection in Coronary IVUS Sequences," *IEEE Transactions on Biomedical Engineering* 59, 1022-1031 (2012).
- Z. Mengliu, G. Hamarneh. "Bifurcation detection in 3D vascular images using novel features and random forest." *IEEE 11th International Symposium on Biomedical Engineering*, 421-424 (2014)
- A. Wang, J. Eggermont, J. H. Reiber, and J. Dijkstra, "Fully automated side branch detection in intravascular optical coherence tomography pullback runs," *Biomedical optics express* 5, 3160-3173 (2014).
- M. M. G. Macedo, W. V. N. Guimarães, M. Z. Galon, C. K. Takimura, P. A. Lemos, and M. A. Gutierrez, "A bifurcation identifier for IV-OCT using orthogonal least squares and supervised machine learning," *Computerized Medical Imaging and Graphics* 46, 237-248 (2015).
- Kohavi R. "A study of cross-validation and bootstrap for accuracy estimation and model selection." *International Joint Conference on Artificial Intelligence*, 1137-43 (1995)
- S. Theodoridis and K. Koutroumbas, [Pattern Recognition], Academic Press (2008).
- S. Fengxi, M. Dayong, and L. Hongfeng, "Feature Selection Based on Linear Discriminant Analysis," *International Conference on Intelligent System Design and Engineering Application (ISDEA)*, 746-749 (2010).
- S. Fengxi, G. Zhongwei, and M. Dayong, "Feature Selection Using Principal Component Analysis," *International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM)*, 27-30 (2010).
- Duda RO, Hart PE, Stork DG. [Pattern classification], Wiley-Interscience (2000).
- Moré J. "The levenberg-marquardt algorithm: implementation and theory." Springer (1978).
- C. D. Doan and S.-Y. Liong, "Generalization for multilayer neural network: Bayesian regularization or early stopping," *Proceedings of the 2nd Conference of the Asia Pacific Association of Hydrology and Water Resources*. (2004).
- H. Demuth, M. Beale, and M. Hagan. "Improve Neural Network Generalization and Avoid Overfitting." 2013. <<http://www.mathworks.com/help/nnet/ug/improve-neural-network-generalization-and-avoid-overfitting.html>> (12 February 2016).
- Hanley JA, McNeil BJ. "The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve." *Radiology* 143(1), 29-36. (1982).

D.M.Baboiu, G. Hamarneh, "Mathematical Methods in Biomedical Image Analysis", Vascular bifurcation detection in scale-space, 41-46 (2012)