



VANESSA GONÇALVES VIDOTTI PIMENTA

**“SENSITIVITY AND SPECIFICITY OF MACHINE LEARNING CLASSIFIERS AND
SPECTRAL DOMAIN OCT FOR THE DIAGNOSIS OF GLAUCOMA”**

**“SENSIBILIDADE E ESPECIFICIDADE DE SISTEMAS DE APRENDIZAGEM DE
MÁQUINA E DA TOMOGRAFIA DE COERÊNCIA ÓPTICA DE DOMÍNIO
ESPECTRAL PARA O DIAGNÓSTICO DE GLAUCOMA”**

CAMPINAS

2014



UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE CIÊNCIAS MÉDICAS

VANESSA GONÇALVES VIDOTTI PIMENTA

“SENSITIVITY AND SPECIFICITY OF MACHINE LEARNING CLASSIFIERS AND SPECTRAL DOMAIN OCT FOR THE DIAGNOSIS OF GLAUCOMA”

Orientador: Prof. Dr. Vital Paulino Costa

“SENSIBILIDADE E ESPECIFICIDADE DE SISTEMAS DE APRENDIZAGEM DE MÁQUINA E DA TOMOGRAFIA DE COERÊNCIA ÓPTICA DE DOMÍNIO ESPECTRAL PARA O DIAGNÓSTICO DE GLAUCOMA”

Tese de Doutorado apresentada à Pós-graduação da Faculdade de Ciências Médicas da Universidade Estadual de Campinas para obtenção do Título de **Doutora em Ciências Médicas**, área de concentração Oftalmologia.

Doctorate thesis presented to the Postgraduation Programme of the School of Medical Sciences of the University of Campinas to obtain the **Ph.D (MSc) grade in Medical Sciences**, area of concentration Ophthalmology.

ESTE EXEMPLAR CORRESPONDE A VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELA ALUNA VANESSA GONÇALVES VIDOTTI PIMENTA E
ORIENTADA PELO PROF. DR. VITAL PAULINO COSTA _____

CAMPINAS

2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Ciências Médicas
Maristella Soares dos Santos - CRB 8/8402

V669s Vidotti Pimenta, Vanessa Gonçalves, 1980-
Sensibilidade e especificidade de sistemas de aprendizagem de máquina e da tomografia de coerência óptica de domínio espectral para o diagnóstico de glaucoma / Vanessa Gonçalves Vidotti Pimenta. – Campinas, SP : [s.n.], 2014.

Orientador: Vital Paulino Costa.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Ciências Médicas.

1. Glaucoma. 2. Inteligência artificial - Classificação. 3. Tomografia de coerência óptica. 4. Sensibilidade e especificidade. I. Costa, Vital Paulino. II. Universidade Estadual de Campinas. Faculdade de Ciências Médicas. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Sensitivity and specificity of machine learning classifiers and spectral domain OCT for the diagnosis of glaucoma

Palavras-chave em inglês:

Glaucoma

Artificial Intelligence, Classification

Optical coherence tomography

Sensitivity and specificity

Área de concentração: Oftalmologia

Titulação: Doutora em Ciências Médicas

Banca examinadora:

Vital Paulino Costa [Orientador]

Ivan Maynard Tavares

Newton Kara José Junior

Keila Miriam Monteiro de Carvalho

José Paulo Cabral de Vasconcellos

Data de defesa: 27-01-2014

Programa de Pós-Graduação: Ciências Médicas

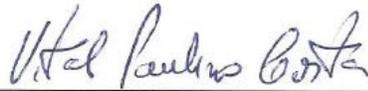
BANCA EXAMINADORA DA DEFESA DE DOUTORADO

VANESSA GONÇALVES VIDOTTI PIMENTA

Orientador (a) PROF(A). DR(A). VITAL PAULINO COSTA

MEMBROS:

1. PROF(A). DR(A). VITAL PAULINO COSTA



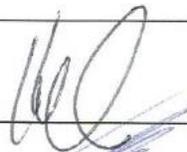
2. PROF(A). DR(A). IVAN MAYNART TAVARES



3. PROF(A). DR(A). NEWTON KARA JOSÉ JUNIOR



4. PROF(A).DR(A). KEILA MIRIAM MONTEIRO DE CARVALHO



5. PROF(A).DR(A). JOSÉ PAULO CABRAL DE VASCONCELLOS



Programa de Pós-Graduação em Ciências Médicas da Faculdade de Ciências Médicas
da Universidade Estadual de Campinas

Data: 27 de janeiro de 2014

DEDICATÓRIA

Aos meus pais, Isaura e Lédio, pelo exemplo de vida e pela constante
dedicação à minha formação.

Ao meu esposo Daniel, pela compreensão, apoio e paciência.

AGRADECIMENTOS

À minha família, pelo amor, apoio e incentivo constante.

Ao meu orientador, Prof. Dr. Vital Paulino Costa, pela dedicação na formação de seus alunos e pelo incansável incentivo na realização desse trabalho, com muita paciência e amizade.

Ao Dr. Edson Gomi e Dr. Marcelo Dias, pelo grande apoio nesse projeto e pelo auxílio na compreensão do mundo da inteligência artificial.

Aos amigos do setor de Glaucoma da UNICAMP, em especial ao Dr. Fabrício Reis Silva, à Dra. Graziela Massa Resende e à Dra. Fernanda Cremasco, pelo companheirismo e cumplicidade em toda esta trajetória.

Às técnicas, Sílvia Alves Ramos F. Silva e Márcia Aparecida F. Silva, pelo auxílio na realização dos exames de campo visual, o que tornou esse projeto possível.

A todos os professores, médicos contratados, funcionários e voluntários da clínica oftalmológica do Hospital das Clínicas da UNICAMP, pelos ensinamentos e estímulo profissional durante toda a minha especialização.

Aos pacientes, motivo para a constante busca de novos conhecimentos e que gentilmente participaram desse projeto, o tornando possível.

SUMÁRIO

	PÁG.
RESUMO	Xxxv
ABSTRACT	Xli
1. INTRODUÇÃO	47
1.1. Diagnóstico de Glaucoma.....	50
1.2. Tomografia de Coerência Óptica.....	53
1.3. Sistemas de Aprendizagem de Máquina.....	60
1.3.1. Criação e avaliação do Classificador.....	63
1.3.2. Algoritmos.....	64
1.3.2.1. Árvore de Decisão (CTREE).....	65
1.3.2.2. <i>Random Forest</i> (RAN).....	67
1.3.2.3. <i>Bagging</i> (BAG).....	69
1.3.2.4. <i>Adaptative Boosting</i> (ADA).....	71
1.3.2.5. <i>Ensemble Selection</i> (ENS).....	73
1.3.2.6. <i>Multilayer Perceptrons</i> (MLP).....	75
1.3.2.7. Rede de Funções de Base Radial (RBF).....	77
1.3.2.8. <i>Naive-Bayes</i> (NB).....	79
1.3.2.9. <i>Support Vector Machine</i> (SVM).....	80
1.4. Análise de desempenho (sensibilidade e especificidade).....	83
1.5. Justificativa.....	85
1.6. Objetivos.....	86
2. METODOLOGIA	87
2.1. População.....	89

2.1.1. Critérios de Inclusão para Indivíduos Saudáveis.....	92
2.1.2. Critérios de Inclusão para Indivíduos com Glaucoma.....	92
2.2. Tomografia de Coerência Óptica.....	93
2.3. Classificadores de Aprendizagem de Máquina.....	94
2.4. Validação Cruzada.....	95
2.5. Análise Estatística.....	96
3. RESULTADOS.....	97
4. DISCUSSÃO.....	109
5. CONCLUSÕES.....	119
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	123
7. ANEXOS.....	141
7.1. Anexo 1. Parecer de Aprovação do Comitê de Ética em Pesquisa da Faculdade de Ciências Médicas da Universidade Estadual de Campinas (FCM-UNICAMP).....	143
7.2. Anexo 2. Tabela de conversão da acuidade visual medida em escala de Snellen para escala de LogMAR.....	145
8. APÊNDICES.....	147
8.1. Apêndice 1. Quadro de dados coletados.....	149
8.2. Apêndice 2. Termo de Consentimento Livre e Esclarecido.....	153
8.3. APÊNDICE 3. Artigo publicado.....	157

LISTA DE TABELAS, FIGURAS E QUADROS

	PÁG.
Figura 1. Interferômetro de Michelson. d1 e d2 representam os comprimentos dos braços de referência e da amostra, respectivamente.....	55
Figura 2. Comparação de imagens obtidas pelo OCT - corte transversal da retina de um olho humano na região macular: a. Stratus-TD-OCT que utiliza interferômetro de domínio temporal, b. Cirrus-SD-OCT que utiliza interferômetro de domínio espectral.....	56
Figura 3. Exemplo de uma análise realizada pelo Cirrus SD-OCT. a: protocolo utilizado (OD - olho direito, OS - olho esquerdo). b: mapa topográfico da espessura da CFNR peripapilar em esquema de cores. Ao centro, dados relacionados à espessura média global, valores de espessura média nos quatro quadrantes e nas 12 horas de relógio, classificados de acordo com a escala de cores de distribuição da população normal que compõe o banco de dados do aparelho. c: mapa de desvio da espessura da CFNR peripapilar em relação ao banco de dados do aparelho. Ao centro, gráfico de perfil da espessura da CFNR peripapilar comparando os dois olhos, com análise de simetria. d: gráfico de perfil da espessura da CFNR peripapilar em comparação com o perfil de distribuição do banco de dados do aparelho em escala de cores. e: imagem de <i>B-scan</i> do corte circular ao redor do NO.....	59
Figura 4. Hierarquia do Aprendizado.....	61
Figura 5. Exemplo de Árvore de Decisão para representação do conceito JogarTênis.....	67

Figura 6. <i>Random Forest</i> - arquitetura geral. x : conjunto de treinamento, ao qual é aplicado n estruturas diferentes de árvores, cada uma obtendo um resultado (K_1-K_n). Uma votação simples é aplicada e o resultado final é determinado pela letra K	68
Figura 7. <i>Bagging</i> - arquitetura geral. São geradas amostras aleatórias a partir de um conjunto de treinamento. Em seguida, classificadores são derivados de cada amostra, criando-se, assim, um classificador composto. É realizada uma votação simples para eleger a classificação mais popular.....	71
Figura 8. <i>Adaptative Boosting</i> - arquitetura geral. Os conjuntos de treinamento são gerados de forma ponderada, aumentando a probabilidade de incluir no novo conjunto, exemplos classificados incorretamente no conjunto anterior. Após a criação de n conjuntos e, portanto, n classificadores, gerados a partir dos conjuntos, é realizada uma votação ponderada para classificar o problema em questão.....	73
Figura 9. <i>Ensemble Selection</i> - arquitetura geral.....	75
Figura 10. Representação de uma rede neural MLP. Apresenta uma camada de entrada, uma de saída e, nesse caso, duas camadas ocultas.....	76
Figura 11. Rede de Funções de Base Radial - arquitetura geral. Apresenta uma camada de entrada, uma de saída e apenas uma camada escondida.....	78
Figura 12. Exemplo de conjunto de dados de treinamento pertencentes a duas classes (triângulos e círculos), classificado segundo três diferentes hipóteses.....	81

Figura 13.	Exemplo de conjunto não linearmente separável.....	82
Figura 14.	Exemplo de criação de um espaço com maior dimensão separável por um hiperplano. (a): exemplo de conjunto não linearmente separável; (b): fronteira não linear onde é possível separar as duas classes; (c): criação de um espaço de maior dimensão possibilitando a separação entre as classes por um hiperplano.....	83
Figura 15.	Curva ROC.....	84
Figura 16.	Sistema de Classificação de Opacificação do Cristalino III.....	91
Tabela 1.	Características demográficas da população estudada.....	100
Tabela 2.	Média das medidas da espessura da camada de fibras nervosas retiniana (\pm desvio padrão) obtidas pelo SD-OCT nos dois grupos.....	101
Tabela 3.	Áreas abaixo da curva ROC (aROC) para cada parâmetro do SD-OCT e sensibilidade (%) com especificidade fixa em 80% e 90%.....	103
Tabela 4.	Áreas abaixo da curva ROC (aROC) para os classificadores de aprendizagem de máquina (MLC) e sensibilidade (%) com especificidade fixa em 80% e 90%.....	104
Tabela 5.	Áreas abaixo da curva ROC (aROC) para os classificadores de aprendizagem de máquina (MLC) treinados com menos parâmetros do SD-OCT.....	106
Figura 17.	Área abaixo da curva ROC (aROC) para o melhor classificador treinado com a quantidade de parâmetros que permitiu melhor performance (BAG = 0,818) e aROC do melhor parâmetro do SD-OCT (I – quadrante inferior = 0,813) ($p=0,932$).....	107

Quadro 1. Estudos que compararam a capacidade diagnóstica do TD-OCT e do SD-OCT.....	111
---	------------

LISTA DE ABREVIATURAS E SIGLAS

%	Porcento
µm	Micrômetro
ADA	<i>Adaptative Boosting</i>
ANN	<i>Artificial Neural Network</i>
aROC	Área sob a <i>Receiver Operating Curve</i>
AT	<i>Average thickness</i>
AV	Acuidade visual
BAG	<i>Bagging</i>
CART	<i>Classification and Regression Tree</i>
CFNR	Camada de Fibras Nervosas da Retina
CNO	Cabeça do Nervo Óptico
CTREE	Árvore de Decisão
CV	Campo Visual
D	Dioptrias
dB	Decibéis
DP	Desvio padrão
EE	Equivalente esférico
ENS	<i>Ensemble Selection</i>
FCM-UNICAMP	Faculdade de Ciências Médicas da Universidade Estadual de Campinas
FDT	<i>Frequency Doubling Technology</i>
FSM	<i>FSMickelberg</i>
GDx	<i>Scanning Laser Polarimetry</i>
GHT	<i>Glaucoma hemifield test</i>

GNU	<i>General Public License</i>
GPAA	Glaucoma Primário de Ângulo Aberto
GPAF	Glaucoma Primário de Ângulo Fechado
HC-UNICAMP	Hospital das Clínicas da Universidade Estadual de Campinas
HRT	<i>Heidelberg Retina Tomograph</i>
I	Quadrante inferior
IC	Intervalo de Confiança
ID3	<i>Inductive Decision Trees</i>
KNOMA	Laboratório de Engenharia do Conhecimento
LDA	Análise discriminante linear
MD	<i>Mean deviation</i>
MDs	Distância Mahalanobis
MLC	Classificadores de Aprendizagem de Máquina / <i>Machine Learning</i> <i>Classifiers</i>
MLP	<i>Multilayer Perceptrons</i>
mmHg	Milímetro de mercúrio
MRA	<i>Moorfield Regression Analysis</i>
n	Número absoluto
N	Quadrante nasal
NB	<i>Naive-Bayes</i>
OCT	Tomografia de Coerência Óptica
OMS	Organização Mundial de Saúde
p	Coeficiente de significância estatística

PÁG	Página
PAP	Perimetria Automatizada Padrão
PCA	<i>Principal Component Analysis</i>
PIO	Pressão Intraocular
PSD	<i>Pattern standard deviation</i>
RAN	<i>Random Forest</i>
RB	<i>R Bathija</i>
RBF	Rede de Funções de Base Radial
ROC	<i>Receiver Operating Curve</i>
RVM	<i>Relevance Vector Machine</i>
S	Quadrante superior
SD-OCT	Tomografia de Coerência Óptica de domínio espectral
SSMoG	<i>Subspace Mixture of Gaussians</i>
SVM	<i>Support Vector Machine</i>
SVMg	<i>Support Vector Machine Gaussian</i>
SVMI	<i>Support Vector Machine</i> Linear
T	Quadrante temporal
TD-OCT	Tomografia de Coerência Óptica de domínio temporal
UNICAMP	Universidade Estadual de Campinas
VC	Vapnik-Chervonenkis
Weka	<i>Waikato Environment for Knowledge Analysis</i>

RESUMO

O presente estudo teve como objetivo investigar a sensibilidade e especificidade dos classificadores de aprendizagem de máquina (MLC) e da tomografia de coerência óptica de domínio espectral (SD-OCT) para o diagnóstico de glaucoma.

Sessenta e dois pacientes com glaucoma e defeito de campo visual inicial a moderado, e 48 indivíduos saudáveis foram incluídos. Todos os participantes foram submetidos a um exame oftalmológico completo, perimetria automatizada acromática e medida da espessura da camada de fibras nervosas da retina com SD-OCT (Cirrus HD-OCT; Carl Zeiss Meditec, Inc., Dublin, Califórnia, EUA). Uma curva ROC (receiver operating curve) foi obtida para cada parâmetro do SD-OCT: espessura média, espessura nos quatro quadrantes (temporal, superior, nasal e inferior) e a espessura a cada hora de relógio. Posteriormente, foram testados os seguintes MLCs construídos a partir dos dados do OCT: *Árvore de Decisão (CTREE)*, *Random Forest (RAN)*, *Bagging (BAG)*, *Adaptative Boosting (ADA)*, *Ensemble Selection (ENS)*, *Multilayer Perceptrons (MLP)*, Rede de Funções de Base Radial (RBF), *Naive-Bayes (NB)*, *Support Vector Machine - Linear (SVMl)* e *Support Vector Machine - Gaussian (SVMg)*. Áreas sob as curvas ROC (aROCs) foram obtidas para cada parâmetro do SD-OCT e para cada MLC.

A média de idade foi de $57,0 \pm 9,2$ anos para os indivíduos saudáveis e de $59,9 \pm 9,0$ anos para os pacientes com diagnóstico de glaucoma ($p=0,103$). Valores do “mean deviation” (MD) foram de $-4,1 \pm 2,4$ dB para pacientes com glaucoma e $-1,5 \pm 1,6$ dB para indivíduos normais ($p<0,001$). Os parâmetros do SD-OCT com as maiores aROCs foram: quadrante inferior (0,813), espessura média global (0,807), 7 horas (0,765) e 6 horas de relógio (0,754). As aROCs dos classificadores variaram de 0,785 (ADA) a 0,818 (BAG). A aROC obtida com o

BAG não foi significativamente diferente da aROC obtida com o melhor parâmetro isolado do SD-OCT ($p=0,932$).

O SD-OCT mostrou boa acurácia diagnóstica em um grupo de pacientes com glaucoma inicial a moderado. Os MLCs não melhoraram a sensibilidade e a especificidade do SD-OCT para o diagnóstico do glaucoma.

Descritores: Glaucoma, classificadores de aprendizagem de máquina, tomografia de coerência óptica, sensibilidade, especificidade, domínio espectral.

ABSTRACT

This study aimed to investigate the sensitivity and specificity of machine learning classifiers (MLC) and spectral domain optical coherence tomography (SD-OCT) for the diagnosis of glaucoma.

Sixty-two patients with early to moderate glaucomatous visual field damage and 48 healthy individuals were included. All subjects underwent a complete ophthalmologic examination, achromatic standard automated perimetry, and RNFL imaging with SD-OCT (Cirrus HD-OCT; Carl Zeiss Meditec, Inc., Dublin, California, USA). Receiver operating characteristic (ROC) curves were obtained for all SD-OCT parameters: average thickness, four quadrants (temporal, superior, nasal and inferior) and 12 clock hour measurements. Subsequently, the following MLCs were tested: Classification Tree (CTREE), Random Forest (RAN), Bagging (BAG), AdaBoost M1 (ADA), Ensemble Selection (ENS), Multilayer Perceptron (MLP), Radial Basis Function (RBF), Naive-Bayes (NB), and Support Vector Machine (SVM). Areas under the ROC curves (aROCs) were obtained for each parameter and each MLC.

The mean age was 57.0 ± 9.2 years for healthy individuals and 59.9 ± 9.0 years for glaucoma patients ($p=0.103$). Mean deviation values were -4.1 ± 2.4 dB for glaucoma patients and -1.5 ± 1.6 dB for healthy individuals ($p<0.001$). The SD-OCT parameters with the greater aROCs were inferior quadrant (0.813), average thickness (0.807), 7 o'clock position (0.765), and 6 o'clock position (0.754). The aROCs obtained with classifiers varied from 0.785 (ADA) to 0.818 (BAG). The aROC obtained with BAG was not significantly different from the aROC obtained with the best single SD-OCT parameter ($p=0.932$).

SD-OCT showed good diagnostic accuracy in a group of patients with early to moderate glaucoma. In this series, MLCs did not improve the sensitivity and specificity of SD-OCT for the diagnosis of glaucoma.

Key Words Glaucoma, machine learning classifiers, optical coherence tomography, sensitivity, specificity, spectral domain.

INTRODUÇÃO

1. INTRODUÇÃO

O glaucoma é uma neuropatia óptica progressiva caracterizada por alterações estruturais na cabeça do nervo óptico (CNO) e na camada de fibras nervosas da retina (CFNR), com alterações funcionais correspondentes, caracterizadas por perda de campo visual (CV)(1). A pressão intraocular (PIO) elevada é o fator de risco mais importante para o desenvolvimento dessa doença (1,2) e o único passível de tratamento (1–3). Outros fatores de risco incluem idade avançada, descendência africana, miopia, história familiar da doença, baixa espessura corneana central (4–6) e diabetes (7). Como o dano glaucomatoso é irreversível, o diagnóstico e a intervenção precoces são fundamentais para evitar o aparecimento de alterações funcionais.

A Organização Mundial de Saúde (OMS) indica o glaucoma como a segunda maior causa de cegueira do mundo (12,3%), superada apenas pela catarata, responsável por 47,8% (8). Segundo Quigley et al (9), haverá 79,6 milhões de pessoas com glaucoma de ângulos aberto e fechado em 2020, o que corresponde a 2,65% da população mundial acima dos 40 anos, com uma estimativa de que, entre esses, 5,9 milhões apresentarão cegueira bilateral por glaucoma primário de ângulo aberto (GPAA), e 5,3 milhões por glaucoma primário de ângulo fechado (GPAF).

A prevalência de glaucoma aumenta com a idade e varia conforme a população avaliada, com estudos revelando valores de 1 a 3% na Europa, 1 a 4 % na Ásia e 2 a 3% na Austrália (9,10). Nos Estados Unidos, a prevalência varia de 1 a 5% dependendo do grupo populacional avaliado. Em norte-americanos de origem europeia ela foi de 1 a 2% (11,12), nos

de descendência africana de 4% (11) e de 2 a 5% nos latinos provenientes do México (13,14). Taxas maiores são encontradas na África, variando aproximadamente entre 1% (15) na Nigéria e 8% em Gana (16).

Os dados de prevalência no Brasil ainda são escassos. Sakata et al (17) avaliaram 1636 indivíduos com mais de 40 anos na região sul do país. A prevalência de glaucoma nesse grupo foi de 3,4% (IC 95% 2,5-4,3), sendo 2,4% de GPAA e 0,7% de GPAF. Além disso, os autores observaram que a prevalência de glaucoma aumentou com a idade e foi maior em indivíduos que se definiram como não pertencendo à raça branca.

1.1. Diagnóstico de Glaucoma

O glaucoma é uma doença caracterizada pela perda de células ganglionares da retina, resultando em alterações estruturais da CFNR e do NO (1). Esse dano estrutural leva a alterações funcionais correspondentes, determinada pela perda progressiva e irreversível de CV (18,19). Dessa maneira, o diagnóstico do glaucoma é baseado em métodos que identificam e quantificam o padrão de defeito visual (dano funcional) e a lesão do nervo óptico e da CFNR (dano estrutural) (20,21).

O defeito visual, caracterizado por perda de CV, é o que acarreta impacto negativo na qualidade de vida dos portadores dessa doença (22,23). Portanto, o diagnóstico precoce e o início imediato do tratamento são essenciais para retardar e até prevenir a progressão da lesão neurológica e a perda de CV (24–26). No entanto, o diagnóstico precoce é um grande desafio,

já que a doença é, nas fases iniciais, assintomática; não existe um teste de triagem efetivo; e a população tem baixa consciência da importância de um exame preventivo e pouca informação sobre os principais fatores de risco da doença (11–14,27,28). No Brasil, Sakata et al (17) observaram que, ao final do estudo para avaliação da prevalência de glaucoma, aproximadamente 90% dos indivíduos detectados como sendo portadores de glaucoma primário desconheciam o seu diagnóstico.

Assim, na tentativa de entender melhor os danos causados pela doença, alguns autores começaram a estudar a relação entre o dano estrutural e o dano funcional observado em pacientes com glaucoma. Quigley et al (18), em 1982, apresentaram um dos primeiros estudos que avaliou essa relação e observaram que a perda definitiva dos axônios ocorre antes do aparecimento de defeitos reprodutíveis no CV em alguns pacientes com suspeita de glaucoma. Outro grupo, também liderado por Quigley (29), realizou uma análise histológica em 6 olhos humanos com glaucoma e observou que, nos 30° centrais da retina, as regiões com diminuição de 5 e 10 dB de sensibilidade no CV estavam associadas à redução de cerca de 20% e 40% da quantidade total de células ganglionares, respectivamente. Estudos mais recentes também demonstraram que defeitos estruturais na CNO e na CFNR peripapilar podem preceder o desenvolvimento de perda detectável na PAP em indivíduos com glaucoma (30,31).

Harwerth et al (32) observaram, por meio de um modelo experimental de glaucoma em macacos *rhesus*, que 30-50% das células ganglionares da retina são perdidas antes que apareçam defeitos no CV demonstrados pela perimetria automatizada padrão (PAP).

Dessa maneira, há um crescente interesse em determinar os sinais precoces de glaucoma na CNO e na CFNR peripapilar. A perda de rima no sentido vertical, a assimetria

entre as escavações de um mesmo paciente, e a presença de hemorragias peripapilares são sinais precoces de dano glaucomatoso no nervo óptico, todavia demonstrou-se que esses achados são pouco prevalentes em amostras populacionais com glaucoma (33,34). A análise da CFNR também pode propiciar o diagnóstico precoce dessa doença, como mostrado pelo estudo de Sommer et al (30), que revelaram que 60% dos pacientes que evoluíram para glaucoma tinham lesão da CFNR seis anos antes do aparecimento de defeito de CV. Esta informação foi confirmada logo após por Quigley et al (35), que observaram que, em pacientes hipertensos oculares, a avaliação da relação escavação disco ou da aparência da CFNR em fotografias, obtidas 5 anos antes da perda de campo visual, eram igualmente preditivas de dano futuro. Da mesma maneira, Kraus et al (36), num estudo prospectivo, demonstraram que 50% dos pacientes hipertensos oculares que desenvolveram dano detectável no campo visual após um acompanhamento de 3,5 anos, apresentavam defeito na CFNR observado por meio de fotografia no início do estudo. Entretanto, a avaliação do nervo óptico e da CFNR, seja por observação direta ou por técnicas fotográficas, é subjetiva e requer examinador experiente (30,37).

Ao longo dos anos, com o avanço tecnológico, foram surgindo métodos que tornaram a avaliação do nervo óptico e da CFNR mais objetiva. São instrumentos que, através da criação de imagens computacionais, fornecem informações objetivas, em tempo real, da análise da CNO e da CFNR peripapilar (38–44). O desafio dos médicos é saber como utilizar a informação que esses instrumentos fornecem para auxiliar na tomada de decisões no manejo do glaucoma.

Assim, diversas técnicas foram introduzidas nos últimos anos visando a detecção de alterações glaucomatosas morfológicas e funcionais mais precoces do que os métodos

convencionais. Dentre as mais empregadas para avaliação funcional, destacam-se as técnicas de perimetria seletiva como o FDT (*Frequency Doubling Technology*) (45,46) e a perimetria azul-amarelo (47,48). Para avaliação anatômica, dispõe-se atualmente do GDx (*Scanning laser polarimetry*) (43), do HRT (*Heidelberg Retina Tomograph*) (49) e do OCT (*Optical Coherence Tomograph*) (50,51).

Além disso, desde 1990 (Goldbaum MH, et. al. IOVS 1990; 31: ARVO Abstract 503), métodos computadorizados que utilizam classificadores de aprendizagem de máquina / *machine learning classifiers* (MLCs) têm sido aplicados em imagens do NO, da CFNR e em medidas de função visual para melhorar a detecção de glaucoma, com resultados sugerindo que estes métodos são tão bons quanto (52–55) ou melhores do que (56–71) as técnicas atualmente disponíveis para classificar os olhos como glaucomatosos ou saudáveis.

Dessa maneira, esforços têm sido feitos para que possamos detectar o glaucoma o mais precocemente possível e para que a análise dos dados que dispomos tenha alta sensibilidade e especificidade no diagnóstico da doença.

1.2. Tomografia de Coerência Óptica

A tomografia de coerência óptica (OCT), descrita pela primeira vez em 1991 por Huang et al (72), é uma técnica de imagem de alta resolução, sem contato, que utiliza um interferômetro de digitalização de baixa coerência, em que a luz é emitida em fase contínua durante um intervalo de tempo muito curto, para produzir imagens seccionais da retina e da

CFNR peripapilar *in vivo* (72,73). Essa tecnologia tornou-se comercialmente disponível em 2002 com o OCT 3 (Stratus OCT, Carl Zeiss Meditec, Inc., Dublin, CA) que utilizava o interferômetro de domínio temporal (TD-OCT), tecnologia patenteada e transferida para a Carl Zeiss Meditec (Dublin,CA) (74).

Esse sistema consiste em um interferômetro (tipicamente o Interferômetro de Michelson – Figura 1) (75) com uma fonte de luz que emite um feixe infravermelho (830nm de comprimento de onda) de fraca potência (até 750 mW). A luz é dividida em duas partes, referência e amostra, e depois é recombinada no detector. Assim, quando dois feixes de luz se encontram no espaço, os seus campos eletromagnéticos se superpõem e o campo resultante é determinado pela soma vetorial dos campos de cada onda. Essa superposição de ondas é chamada de interferência. Quando as ondas se superpõem em fase (crista com crista ou vale com vale) produzem uma onda resultante com amplitude máxima, criando um ponto brilhante. Se os dois feixes se encontram fora de fase (crista com vale), irão cancelar-se, produzindo um ponto escuro (76).

A técnica do OCT é baseada na detecção e análise do sinal de interferência produzido entre a reflexão de referência e o sinal refletido, ou retro-espalhado, vindo de vários pontos do objeto que está sendo analisado. A imagem do OCT utiliza essa informação para representar variações na reflectividade óptica através da profundidade do tecido ao longo de um ponto, criando o que é conhecido como *A-scan*. Estes *scans* individuais axiais podem ser reunidos de forma linear através do tecido, criando uma imagem de corte transversal, conhecida como *B-scan*. Uma coleção de *B-scans* paralelos pode formar uma imagem tridimensional (74).

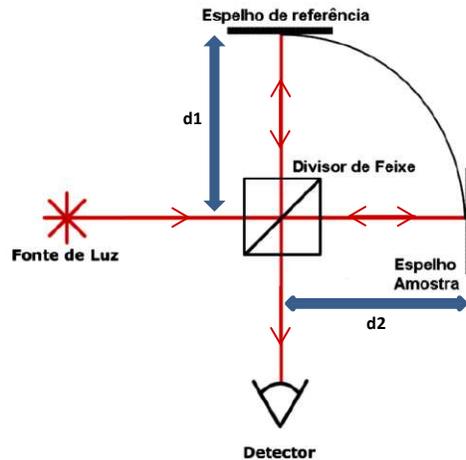


Figura 1. Interferômetro de Michelson. **d1** e **d2** representam os comprimentos dos braços de referência e da amostra, respectivamente (modificado de Michelson AA, 1995) (75).

Na tecnologia de domínio temporal, o espelho de referência deve ser móvel, porque o padrão de interferência só se dá quando o tempo de vôo dos fótons retro-refletidos pelo espelho de referência for o mesmo dos fótons retro-espalhados pela amostra. Desta forma, é possível, através do deslocamento do espelho de referência e conseqüentemente da alteração do tempo de vôo dos fótons, selecionar a profundidade que se deseja observar (74).

Posteriormente, desenvolveu-se a técnica de OCT baseada no domínio espectral, na qual o padrão de interferência é dado não no espaço direto, mas sim no espaço das frequências, ou espaço recíproco, também conhecido como domínio de Fourier. Assim, com uma análise matemática conhecida como Transformada de Fourier (77), foi possível manter o espelho de referência imóvel, o que tornou o exame cerca de 70 vezes mais rápido.

Quando o TD-OCT Stratus começou a ser utilizado na prática clínica, diversos estudos demonstraram altas sensibilidade e especificidade para o diagnóstico de glaucoma, além de uma boa correlação com os resultados de CV detectados com a PAP (51,78,79).

Recentemente, várias empresas desenvolveram novas versões de OCT empregando a tecnologia de domínio espectral (SD-OCT), já que o princípio dessa técnica se baseia mais na aplicação da fórmula de Fourier do que na tecnologia anteriormente patenteada. Os SD-OCTs têm maior resolução axial e velocidade de varredura do que o TD-OCT. O Cirrus-SD-OCT (Carl Zeiss Meditec, Inc.) tem uma resolução axial de cinco μm e uma velocidade de digitalização de 27.000 *A-scans* por segundo, enquanto o Stratus-TD-OCT tem uma resolução axial de oito a dez μm e uma velocidade de digitalização de 400 *A-scans* por segundo. Portanto, os novos OCTs permitem que mais dados sejam coletados num menor espaço de tempo (Figura 2).

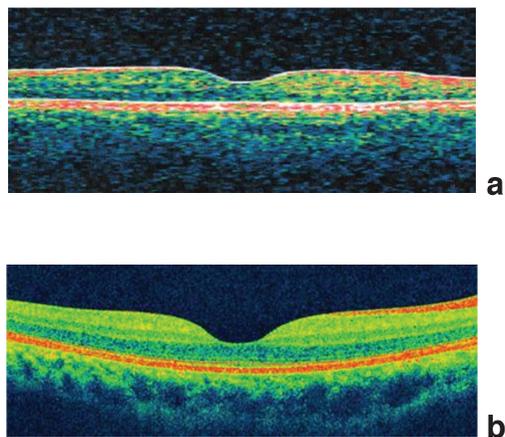


Figura 2. Comparação de imagens obtidas pelo OCT - corte transversal da retina de um olho humano na região macular: **a.** Stratus-TD-OCT que utiliza interferômetro de domínio temporal, **b.** Cirrus-SD-OCT que utiliza interferômetro de domínio espectral (80,81).

Estudos comparando essas tecnologias têm demonstrado que a sensibilidade e especificidade dos vários parâmetros da CFNR analisados pelo Cirrus-SD-OCT são excelentes e equivalentes ao Stratus-TD-OCT (82–88). Além disso, o Cirrus-SD-OCT é um aparelho que apresenta boa reprodutibilidade (89,90), não havendo necessidade de dilatação pupilar (91) para realização do teste. Isso torna o exame seguro e confortável para o paciente.

Nesse estudo utilizamos o Cirrus-SD-OCT (software versão 3.0, Carl Zeiss Meditec, Inc.) que, empregando a tecnologia de domínio espectral, estima a espessura da camada de fibras nervosas da retina peripapilar. O modo cubo 200x200 do disco óptico consiste em 200x200 *A-scans* que cobrem uma área de seis mm² centrada no disco óptico. Depois de criar um mapa da espessura da CFNR em três dimensões formando um cubo, o software determina o centro do disco e extrai um círculo ao redor da papila (1,73 mm de raio) para a medida da espessura da CFNR. Ao final da captura da imagem, o Cirrus-SD-OCT fornece um índice sobre a intensidade do sinal que pode variar de zero a dez, sendo dez o sinal máximo. Quando os valores são inferiores a cinco, o índice aparece em vermelho, mostrando que o sinal está abaixo de um padrão aceitável; quando o valor é de cinco ou mais, o índice aparece em verde, determinando um sinal aceitável. O Cirrus-SD-OCT fornece em seguida mapas da CFNR com 17 parâmetros: espessura média, espessura nos quatro quadrantes (temporal, superior, nasal e inferior) e a espessura a cada hora de relógio.

Para cada um dos 17 parâmetros avaliados, o aparelho fornece uma análise estatística comparando o resultado encontrado para cada paciente com o banco de dados do aparelho, que é formado por 284 indivíduos normais (43% brancos, 24% asiáticos, 18% africanos, 12% hispânicos, 1% indiano, 6% misto), 134 homens e 150 mulheres com idade variando de 18 a 84 anos (média = 46,5 anos) (92). A comparação, que leva em consideração a idade e o sexo

do indivíduo, permite classificar o resultado segundo os percentis de distribuição normal da população: $p < 1\%$ em vermelho, $1\% \leq p < 5\%$ em amarelo, $5\% \leq p \leq 95\%$ em verde e $p > 95\%$ em branco (Figura 3).

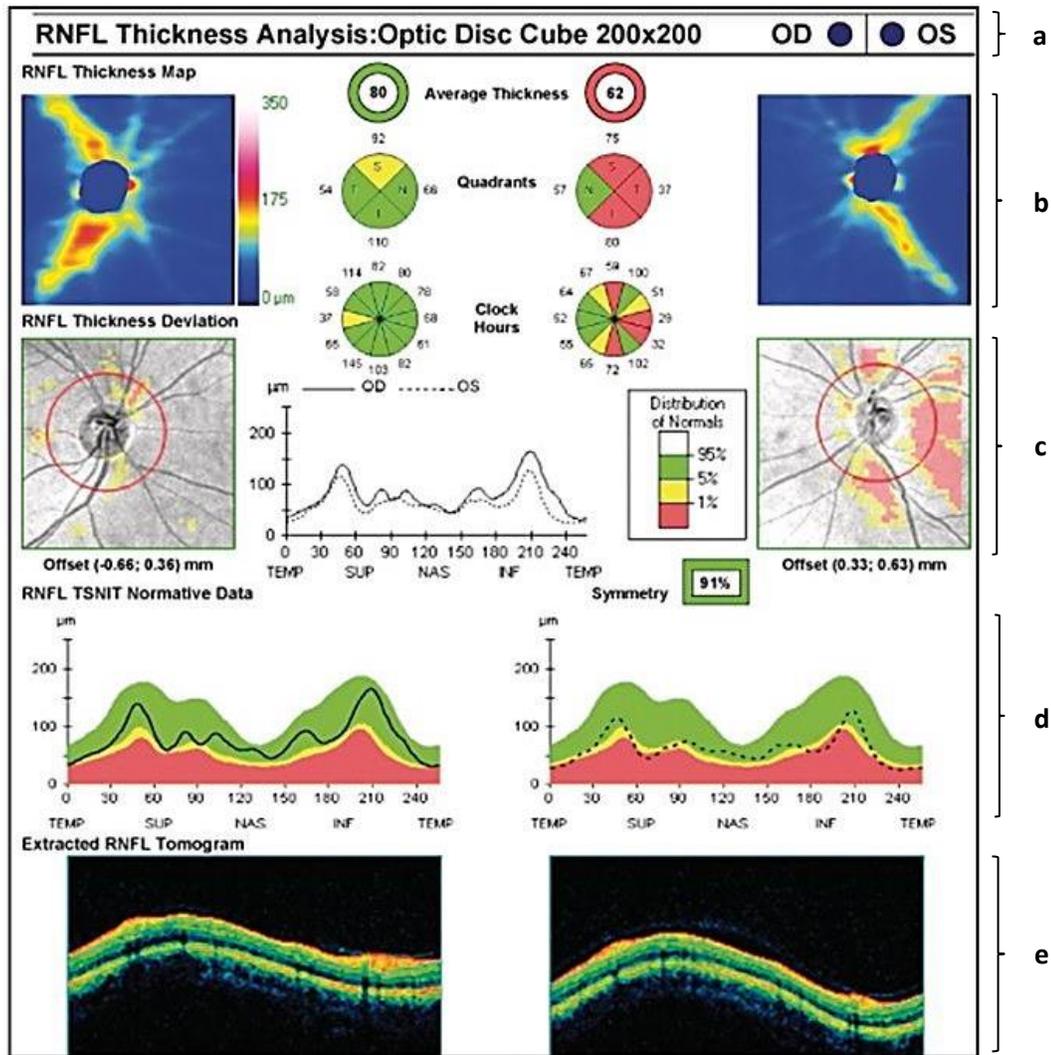


Figura 3. Exemplo de uma análise realizada pelo Cirrus-SD-OCT. **a:** protocolo utilizado (OD - olho direito, OS - olho esquerdo). **b:** mapa topográfico da espessura da CFNR peripapilar em esquema de cores. Ao centro, espessura média global, valores de espessura média nos quatro quadrantes e nas 12 horas de relógio, classificados de acordo com a escala de cores de distribuição da população normal que compõe o banco de dados do aparelho. **c:** mapa de desvio da espessura da CFNR peripapilar em relação ao banco de dados do aparelho. Ao centro, gráfico de perfil da espessura da CFNR peripapilar comparando os dois olhos, para análise de simetria. **d:** gráfico de perfil da espessura da CFNR peripapilar em comparação com o perfil de distribuição do banco de dados do aparelho em escala de cores. **e:** imagem de *B-scan* do corte circular ao redor do NO. (Fonte: Serviço de Glaucoma da Unicamp).

1.3. Sistemas de Aprendizagem de Máquina

Na medicina moderna temos à disposição uma grande quantidade de conhecimento e, diariamente, enfrentamos o desafio de adquirir, analisar e aplicar toda essa informação para resolver problemas clínicos complexos. Transformar um conjunto de dados em informações e descobrir suas inter-relações para produzir conhecimento útil e aplicá-lo à solução de problemas, são tarefas típicas da computação. Uma área da computação normalmente empregada na resolução de problemas da área médica é a inteligência artificial, sendo que uma subárea de especial relevância para questões de diagnóstico médico é a aprendizagem de máquina. Dessa forma, sistemas computacionais capazes de aprender regras ao analisarem registros médicos disponíveis em vários meios como bancos de dados, textos não estruturados, imagens, filmes e sons, podem auxiliar significativamente o diagnóstico e tratamento de doenças (93).

Portanto, o processo de aprendizagem de máquina é um subcampo da inteligência artificial, que procura desenvolver algoritmos que possam dotar os computadores da capacidade de aprendizagem de forma a elevar o seu desempenho automaticamente através da experiência (94). Ou seja, o processo de aprendizagem inclui mudanças adaptáveis no sistema, o que o permite fazer, da próxima vez, a mesma tarefa com mais eficiência.

O aprendizado utiliza o princípio da indução (inferência lógica) com o objetivo de obter conclusões genéricas a partir de um conjunto de exemplos. Assim, um conceito é aprendido efetuando-se inferência indutiva sobre os exemplos apresentados. Para a indução derivar conhecimento novo representativo, os exemplos têm que estar bem definidos e em

quantidade suficiente, obtendo assim hipóteses úteis para um determinado tipo de problema. Quanto mais exemplos relevantes forem selecionados para treinamento no indutor, mais bem classificado será o novo conjunto de dados. O intuito final é construir um classificador que possa determinar a classe a que um exemplo não rotulado pertence (95).

O aprendizado indutivo pode ser dividido em supervisionado e não supervisionado. O aprendizado supervisionado é utilizado para classificação de exemplos em classes pré-definidas, resolvendo problemas preditivos. Para dados categóricos (não numéricos), esse problema é conhecido como classificação e para valores contínuos (numéricos), como regressão. O aprendizado não-supervisionado é utilizado para agrupamento, reunindo exemplos semelhantes e resolvendo, dessa maneira, problemas descritivos. Após os agrupamentos, normalmente é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema que está sendo analisado (Figura 4) (96).

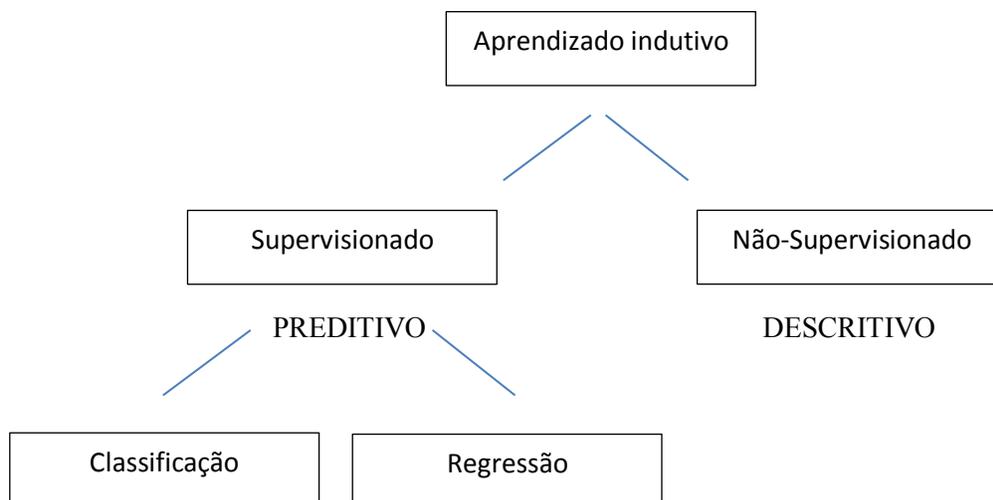


Figura 4. Hierarquia do aprendizado.

Nos classificadores baseados em algoritmos de aprendizagem de máquina, geralmente emprega-se uma forma de aprendizagem supervisionada, onde o programa aprende a partir de exemplos positivos e negativos para a doença, que representam os casos em que, por exemplo, há sinais de glaucoma (exemplos positivos), ou não (exemplos negativos). O treinamento é repetido diversas vezes com o fornecimento de dados de treinamento diferentes, com a classificação positiva ou negativa previamente realizada por um oftalmologista, até que o conceito (identificação de sinais de glaucoma em um teste) possa ter sido devidamente aprendido pelo sistema. Passa-se então a utilizar o classificador com novos dados de exames, porém não previamente classificados por um médico especialista, para que o sistema então faça sua classificação automaticamente (97).

Atualmente existem diversos *softwares* desenvolvidos para criar classificadores. Esses *softwares* são ferramentas de mineração de dados (*data mining*), que consiste no processo automático de descoberta de conhecimento em bases de dados muito volumosas. Os primeiros *softwares* para mineração de dados começaram a ser desenvolvidos em meados da década de 90, ainda em ambiente acadêmico. Dentre as ferramentas disponíveis para *data mining*, a mais simples e largamente utilizada e aceita é a ferramenta Weka (Waikato Environment for Knowledge Analysis) (97). O sistema Weka é um *software* livre (de código aberto) para mineração de dados, desenvolvido em Java, dentro das especificações da GNU (*General Public License*). As suas características, bem como as técnicas nele utilizadas, foram descritas de forma detalhada por Witten e Frank (97,98).

1.3.1. Criação e avaliação do classificador

A criação e avaliação dos classificadores de aprendizagem de máquina supervisionados seguem algumas etapas (99):

- 1) Definição do problema e seleção dos atributos – primeiramente é necessário definir o problema a ser solucionado pela máquina (por exemplo, diagnóstico de glaucoma ou diagnóstico de normalidade) e então coletar os dados, também chamados de atributos ou parâmetros, que sejam relevantes (como dados do OCT). Se há um profissional experiente à disposição, ele pode sugerir quais dados são mais informativos. Se não há este profissional, é possível medir tudo o que está disponível na esperança de que os parâmetros relevantes possam ser isolados. Uma característica importante que esse conjunto de treinamento deve ter é a capacidade de representar adequadamente a população em geral, pois se foram apresentados apenas dados de pacientes normais ao mecanismo de aprendizagem durante o seu treinamento, o sistema vai falhar futuramente ao tentar reconhecer dados de indivíduos com glaucoma.
- 2) Pré-processamento dos dados – o objetivo dessa etapa é reduzir a quantidade de dados realizando uma seleção que identifica e remove atributos irrelevantes e redundantes, tornando o processo de resolução do problema mais rápido e efetivo. A seleção de atributos não é usada apenas para lidar com ruídos (dados imperfeitos ou classes rotuladas incorretamente), mas também para lidar com uma amostra de dados muito grande.
- 3) Seleção do algoritmo – não existe um único algoritmo que apresente o melhor desempenho para todos os problemas, portanto a sua escolha é um passo crítico. Os algoritmos disponíveis variam de acordo com a aplicação e a estrutura de dados

escolhida. Dessa forma, é preciso testar alguns algoritmos para definir aquele que tem o melhor resultado para o problema em questão.

- 4) Treinamento e teste – nessa etapa, são apresentados ao programa dados de treinamento positivos e negativos (por exemplo, OCT de indivíduo com glaucoma e OCT de indivíduo normal) previamente classificados por profissional qualificado. Idealmente, uma população é utilizada para treinamento e outra população é então utilizada para testar a capacidade diagnóstica do classificador. Para maximizar o emprego dos dados coletados excluindo a necessidade de uma nova população para teste, é possível empregar o método da validação cruzada para treinar e testar o classificador (100). Esse método consiste em dividir aleatoriamente a amostra de pacientes pré-classificados por um profissional experiente, em 10 subgrupos, cada um contendo aproximadamente a mesma proporção de indivíduos saudáveis e doentes. Nove subgrupos são utilizados como dados de treino, enquanto que o subconjunto remanescente é utilizado como dados de teste. Este processo é realizado 10 vezes, até que cada um dos 10 subgrupos seja utilizado como teste.

1.3.2. Algoritmos

Os classificadores de aprendizagem de máquina podem ser desenvolvidos utilizando inúmeros algoritmos. Em seguida estão apresentados os mais comumente empregados em medicina: *Árvore de Decisão* (CTREE), *Random Forest* (RAN), *Bagging* (BAG), *Adaptative Boosting* (ADA), *Ensemble Selection* (ENS), *Multilayer Perceptrons* (MLP), Rede de Funções de Base Radial (RBF), *Naive-Bayes* (NB), *Support Vector Machine - Linear* (SVMl) e *Support Vector Machine - Gaussian* (SVMg).

1.3.2.1. Árvore de Decisão (CTREE)

O algoritmo de árvores de decisão tem sido muito utilizado pelo fato de possuir uma representação intuitiva, que torna o modelo de classificação fácil de ser interpretado. Assim, é gerada uma estrutura de árvore que ajuda na classificação e predição das amostras desconhecidas. Com base nos registros dos dados de treinamento, uma árvore é montada como um conjunto de nós que são conectados por ramificações. Basicamente existem três tipos de nós: o nó raiz, que inicia a árvore; os nós comuns, que dividem um determinado atributo e geram ramificações; e os nós folha, que contém as informações de classificação do algoritmo. Ao montar a árvore, cada nó indica o teste de um atributo, cada ramo corresponde a um valor do atributo e cada folha representa uma classe (101).

Existem diversos algoritmos na literatura utilizados para a construção de árvores de decisão, sendo o *ID3 (inductive decision trees)* o método clássico. *ID3* é um algoritmo recursivo que cria a árvore desde a raiz até as folhas. A meta é a construção de árvores que possuam o menor tamanho e a maior acurácia possíveis. Uma questão chave para a construção de uma árvore de decisão consiste na estratégia para a escolha dos atributos que estarão mais próximos da raiz da árvore, ou seja, definir os atributos que são inicialmente avaliados para determinar a classe a qual uma observação pertence. Geralmente são utilizadas medidas baseadas na entropia da informação, que identifica o grau de incerteza de uma informação, para escolher o melhor atributo em cada etapa de formação. Assim, é calculado o ganho de informação de cada atributo do conjunto e aquele que apresentar o maior valor será selecionado como nó raiz da árvore e seus possíveis valores serão os ramos que o conectará aos demais nós (102).

Um exemplo clássico de CTREE é o conceito JogarTênis (103), no qual é avaliada a possibilidade de prática do esporte dependendo das condições climáticas. A Figura 5 representa uma árvore onde cada nó (retângulo) é um teste sobre um atributo (Previsão, Umidade e Vento) e os ramos que seguem imediatamente o nó são os possíveis valores que este atributo pode assumir. O primeiro nó é a raiz e os últimos são as folhas. O processo de classificação inicia-se com a apresentação de uma instância (exemplo) de valores de atributos (parâmetros) para a árvore, e o primeiro teste do nó raiz é efetuado (Previsão). Em seguida, o ramo correspondente ao valor do atributo na instância (Ensolarado, Nublado ou Chuvoso) conduzirá o fluxo de avaliação para o nó seguinte, e assim sucessivamente até que um valor (sim ou não) para o conceito (JogarTênis) seja obtido. Assim, uma instância com os seguintes valores de atributos: Previsão=Ensolarado, Temperatura=Quente, Umidade=Alta, Vento=Forte; produzirá como classificação JogarTênis=Não, indicando que a instância é negativa.

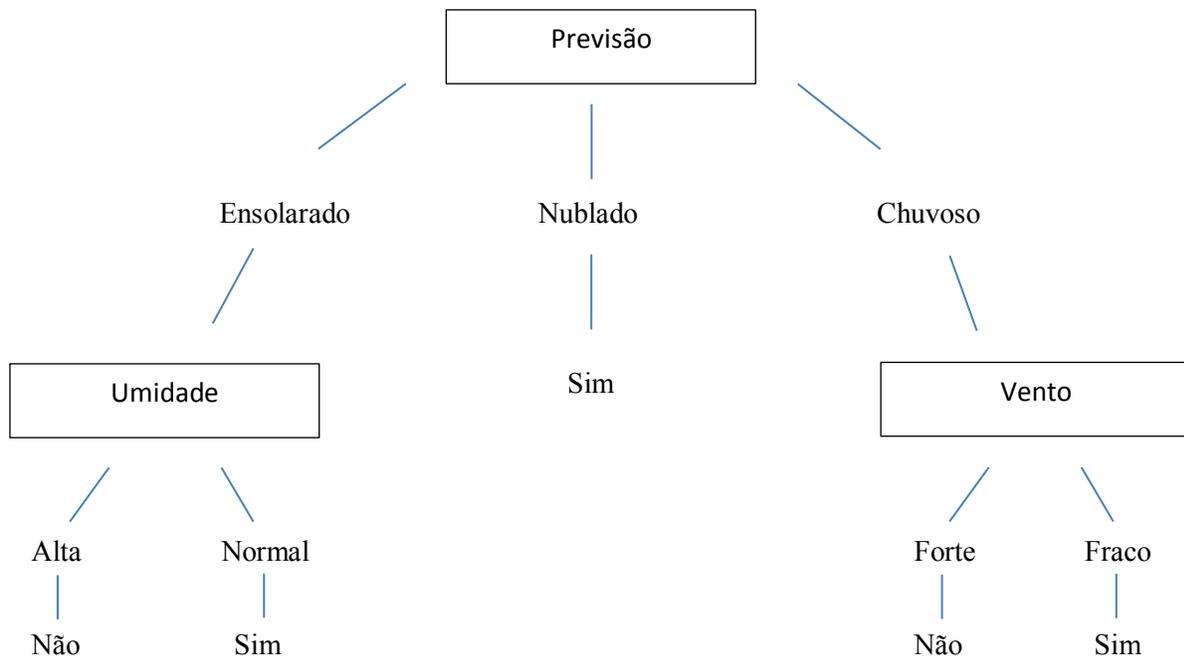


Figura 5. Exemplo de árvore de decisão para representação do conceito JogarTênis (Adaptado de Mitchell,1997)(103).

As árvores de decisão permitem saber por onde percorreu a informação e o algoritmo trabalha bem com valores discretos, pois, caso contrário, a árvore pode se tornar imensa e de difícil compreensão.

1.3.2.2. *Random Forest* (RAN)

RAN, ou Floresta Randômica, é um classificador proposto por Breiman em 2001 (104) e funciona com base em voto de um conjunto de árvores de decisão. Em adição à construção de cada árvore usando uma amostra aleatória de dados, RAN muda como as árvores classificadoras são construídas. Nas árvores padrão, cada nó é dividido usando a melhor

separação entre todas as variáveis. No RAN, cada nó é dividido usando o melhor entre um subgrupo de preditores escolhidos randomicamente naquele nó. Cada árvore de decisão é construída com profundidade máxima, isto é, sem poda. Isso confere uma melhora no desempenho quando comparados a outros classificadores.

O algoritmo *Random Forest* é de fácil manuseio, pois o usuário manipula somente dois parâmetros: o número de variáveis (características) do subgrupo randomizado em cada nó e o número de árvores da floresta. O algoritmo RAN funciona da seguinte maneira: 1) desenha n amostras do conjunto original de dados; 2) para cada amostra, cresce uma árvore de classificação sem poda com a seguinte modificação: em cada nó, ao invés de escolher a melhor separação entre todos os preditores, uma amostra aleatória dos preditores é feita e a melhor separação entre essas variáveis é escolhida; 3) a predição de novos dados é feita por meio da combinação de todas as árvores, através do voto majoritário (Figura 6).

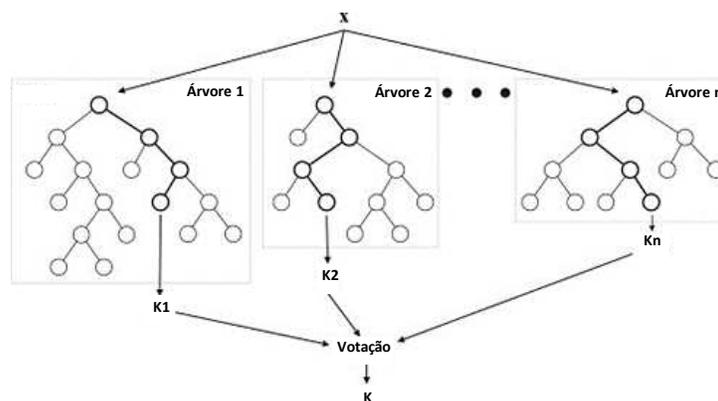


Figura 6. *Random Forest* - arquitetura geral. **x**: conjunto de treinamento, ao qual é aplicado n estruturas diferentes de árvores, cada uma obtendo um resultado (K_1 - K_n). Uma votação simples é aplicada e o resultado final é determinado pela letra **K**. (modificado de Verikas et al, 2011)(105).

Assim, essencialmente, a técnica de RAN dá um passo além na classificação realizada pela CTREE, utilizando métodos como:

- 1) *Bagging*: gerar, através de reamostragem (*bootstrapping*), um conjunto de, por exemplo, 500 amostras retiradas da amostra original através de seleção aleatória com reposição;
- 2) *Boosting*: desenvolver em cada amostra, uma árvore onde, para cada observação incorretamente classificada numa determinada amostra do *bootstrapping*, sua ponderação seja aumentada quando houver a criação de uma nova árvore em outra amostra do mesmo *bootstrapping*;
- 3) *Randomizing*: utilizar um conjunto diferente de variáveis, escolhido de maneira aleatória, em cada método/técnica empregada.

O pacote RAN produz opcionalmente uma informação adicional: a medida da importância de uma variável preditora. No geral este é um conceito de difícil definição, porque a importância de uma variável pode ser secundária à sua interação com outras variáveis. O algoritmo RAN estima a importância de uma variável verificando o quanto o erro preditor aumenta quando os dados dessa variável são modificados, enquanto os outros dados são mantidos (106).

1.3.2.3. *Bagging* (BAG)

BAG, ou ***Bootstrap Aggregating***, é um algoritmo que se utiliza da combinação de classificadores, com o objetivo de obter uma taxa de acerto melhor do que a obtida pela aplicação de um classificador único (107).

Dessa maneira, a partir de uma base de dados de treinamento, o algoritmo é construído em quatro etapas:

- 1) Geração de diferentes amostras, que devem ter tamanhos iguais, obtidas de forma aleatória e com distribuição uniforme. Isso é conseguido por meio de reposição, ou seja, para as amostras terem tamanhos iguais algumas instâncias se repetirão, enquanto outras serão omitidas;
- 2) Obtenção de um classificador para cada amostra;
- 3) Obtenção de um classificador composto, que inclui todos os classificadores individuais gerados na fase anterior;
- 4) Votação simples para eleger a classificação mais popular dentre os classificadores individuais.

Portanto, o método BAG gera um classificador composto, no qual o resultado final é um conjunto de classificadores que são utilizados de forma integrada para classificar uma nova instância, onde a resposta será aquela escolhida pela maioria dos classificadores (Figura 7).

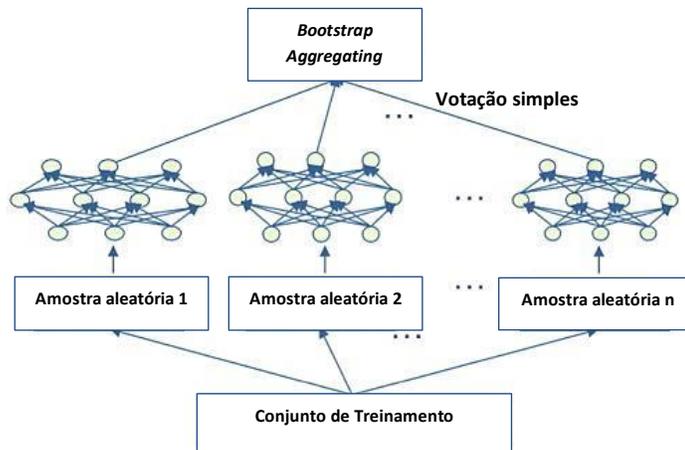


Figura 7. *Bagging* – arquitetura geral. São geradas amostras aleatórias a partir de um conjunto de treinamento. Em seguida, classificadores são derivados de cada amostra, criando-se, assim, um classificador composto. É realizada uma votação simples para eleger a classificação mais popular (modificado de Wang et al, 2012) (108).

Esse algoritmo é muito útil quando o conjunto de treinamento apresenta dados que possuem comportamento instável, ou seja, pequenas mudanças no conjunto de treinamento acarretam mudanças de classificação. Nestes casos, um classificador único não é capaz de oferecer resposta confiável para todas as situações, mas um conjunto de classificadores pode ter maior chance de acerto (95).

1.3.2.4. *Adaptative Boosting* (ADA)

O método *Boosting* é outro método de combinação de classificadores, que foi desenvolvido com o intuito de oferecer classificadores mais eficientes. A partir daí foram criados vários algoritmos, sendo que o mais conhecido é o ADA (109,110).

Neste método, como no método BAG, são geradas amostras que dão origem a classificadores que são utilizados de forma integrada. Entretanto, existem duas distinções entre os métodos *Bagging* e *Boosting*: a maneira como são geradas as amostras e a maneira como são combinados os resultados dos classificadores.

A maneira como as amostras são geradas no *Boosting* difere do *Bagging*, pois não são obtidas através de uma distribuição uniforme, mas sim através da variação do peso associado a cada um dos exemplos para a geração das próximas amostras. Assim, a primeira amostra é criada assumindo a distribuição uniforme, em que todas as instâncias da base têm a mesma probabilidade de serem incluídas nessa primeira amostra. De acordo com o classificador obtido nesse primeiro passo, é possível diminuir a probabilidade de serem incluídas na próxima amostra instâncias que foram corretamente classificadas e aumentar a probabilidade das instâncias que foram incorretamente classificadas. Dessa forma, atribui-se peso maior para exemplos classificados incorretamente na próxima interação.

A segunda diferença está na forma como são combinados os classificadores. No método *Bagging*, a classificação final é feita por meio de uma votação simples entre as respostas fornecidas por cada classificador, enquanto no método *Boosting* a votação é ponderada segundo um índice de importância entre os classificadores gerados (Figura 8).

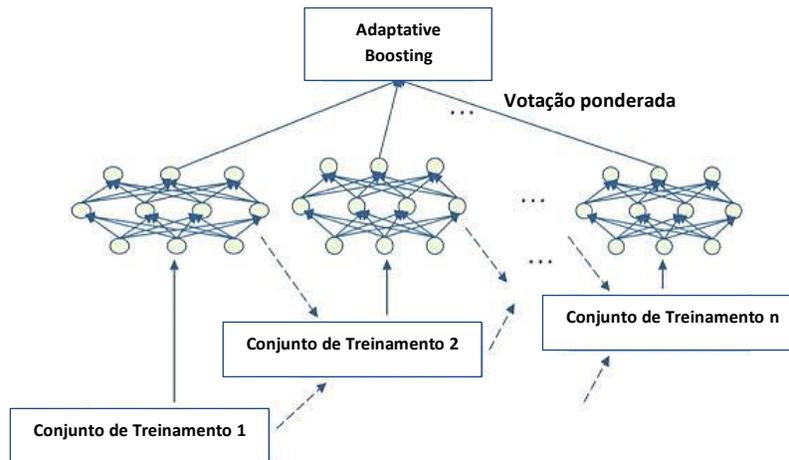


Figura 8. *Adaptative Boosting* - arquitetura geral. Os conjuntos de treinamento são gerados de forma ponderada, aumentando a probabilidade de incluir no novo conjunto, exemplos classificados incorretamente no conjunto anterior. Após a criação de n conjuntos e, portanto, n classificadores, gerados a partir dos conjuntos, é realizada uma votação ponderada para classificar o problema em questão (modificado de Wang et al, 2012)(108).

1.3.2.5. *Ensemble Selection* (ENS)

Em 2004, Caruana e colaboradores (111) criaram o algoritmo ENS (seleção de conjuntos), que é um classificador que utiliza um conjunto de classificadores a fim de melhorar a sua performance. O algoritmo treina um grande número de classificadores com a suposição de que pelo menos alguns terão bom desempenho. O melhor conjunto de classificadores é então selecionado.

Nenhum algoritmo é superior aos outros em todas as tarefas, pois o desempenho depende das características dos dados que estão sendo utilizados. Na prática, diferentes algoritmos são capazes de obter vantagens em diferentes características em relação a um

determinado banco de dados. Intuitivamente, o objetivo do ENS é, automaticamente, detectar e combinar as forças desses algoritmos únicos para criar uma soma que é melhor do que as partes.

O ENS é estruturado da seguinte maneira:

- 1) Seleção de uma variedade de algoritmos de aprendizagem;
- 2) Escolha de um grande número de parâmetros para cada algoritmo;
- 3) Divisão dos dados entre treinamento, afinamento e teste;
- 4) Construção da “biblioteca”;
- 5) Seleção do objetivo métrico apropriado para o problema;
- 6) Seleção dos classificadores com os melhores desempenhos para o objetivo métrico selecionado;
- 7) Predição através da média dos preditores constituintes da biblioteca.

Dessa maneira, o ENS baseia-se na construção de uma “biblioteca”, que é um amplo e diversificado conjunto de classificadores que utilizam diferentes parâmetros. Após a construção desta biblioteca, inicia-se a combinação dos modelos. É escolhido o melhor modelo de toda a biblioteca, que é aquele que faz o melhor trabalho de acordo com o conjunto de validação. Então, outros modelos são adicionados, um de cada vez, com o objetivo de obter o melhor desempenho do conjunto de classificadores. Esse processo é realizado de forma automática pelo algoritmo (Figura 9).

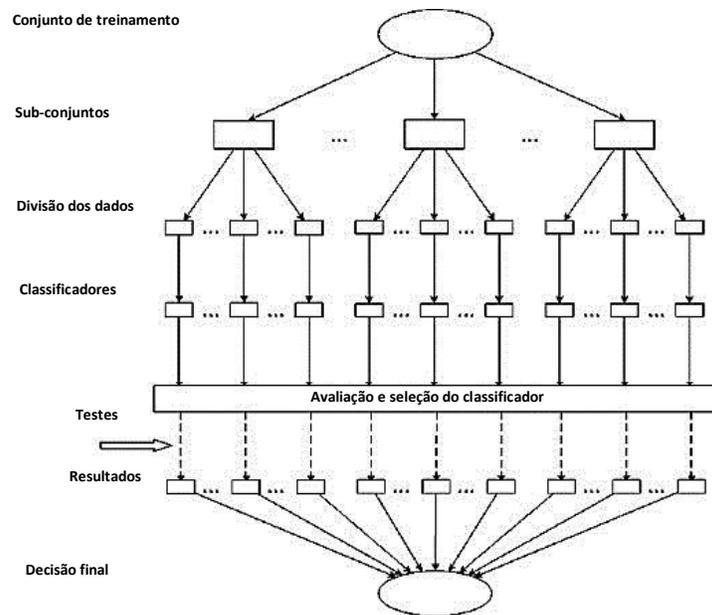


Figura 9. *Ensemble Selection* - arquitetura geral (modificado de Liu et al, 2009)(112).

O ENS é muito útil quando existe um grande banco de dados, já que a maioria dos algoritmos que utilizam aprendizado indutivo é computacionalmente complexa e sofre com problemas de memória quando aplicada para essas condições. A solução do ENS para esse problema é a divisão horizontal do banco de dados em partes menores, treinamento do modelo preditivo em cada parte menor e combinação desses modelos (113).

1.3.2.6. *Multilayer Perceptrons*(MLP)

MLP, ou redes neurais multicamadas, são redes neurais artificiais inspiradas na maneira como se organiza e como funciona o cérebro humano. Consiste, portanto, de um grupo interligado de neurônios artificiais que processa a informação utilizando uma abordagem computacional (114). Na maioria dos casos, uma rede neural artificial é um

sistema adaptativo que muda a sua estrutura interna ou externa baseado na informação que flui através da rede, durante a fase de aprendizagem (115).

Nessa estrutura em rede, destacam-se as unidades (neurônios) e suas conexões (sinapses). São obtidas por simulação em computadores e, a cada conexão, associa-se um peso sináptico (valor numérico) que caracteriza a força da conexão entre dois neurônios. O aprendizado em uma rede neural artificial é realizado por um processo de adaptação de seus pesos sinápticos.

A rede é segregada em três camadas conforme observamos na Figura 10: camada de entrada (*input layer*), que recebe a informação a ser processada; camada de saída (*output layer*), onde os resultados do processamento são fornecidos; e unidades intermediárias entre as camadas de entrada e saída conhecidas como camadas ocultas ou escondidas (*hidden layers*). O sentido de propagação da informação é único, da camada de entrada em direção à camada de saída, percorrendo as camadas escondidas, o que forma uma rede multicamadas.

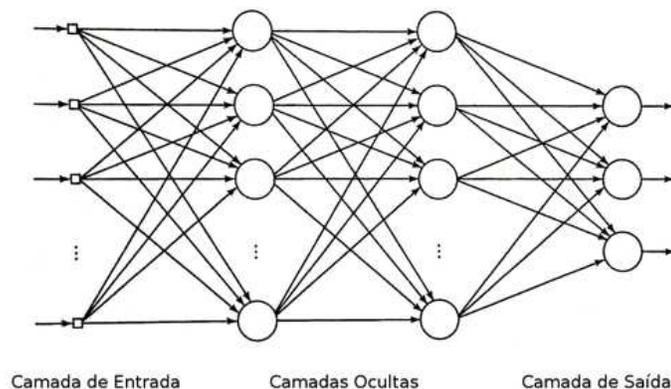


Figura 10. Representação de uma rede neural MLP. Apresenta uma camada de entrada, uma de saída e, nesse caso, duas camadas ocultas (116).

Quando os dados a serem classificados podem ser separados linearmente, ou seja, uma linha ou plano pode ser desenhado para separar a instância de entrada entre suas categorias de conexão, então uma rede neural de camada única pode ser treinada e classificar corretamente estes dados. Mas, se as instâncias não podem ser linearmente separadas, é preciso construir uma rede multicamada (MLP), que consiste em um grande número de unidades (neurônios) agrupado num padrão de conexão em que a camada de entrada e a de saída são separadas por uma ou mais camadas escondidas. Assim, a função que separa as instâncias deixa de ser linear e passa, na maioria das vezes, a função sigmóide.

Sabe-se que nas redes neurais de camada única, o aprendizado é realizado facilmente devido à simplicidade no cálculo do erro de comparação entre a saída esperada e a saída da rede. Contudo, nas redes multicamadas não é possível calcular diretamente o erro para as unidades escondidas, uma vez que não são definidas as saídas destas. Assim, foi desenvolvida uma técnica de aprendizado para as redes multicamadas denominada como algoritmo de *back-propagation* (117,118). Este algoritmo tem como idéia principal propagar o erro da saída de volta para as camadas anteriores, de forma que todas as unidades da rede possam aprender com o erro. Este ciclo é repetido até que a taxa de erro seja aceitavelmente baixa em comparação com os exemplos marcados no conjunto de ensino.

1.3.2.7. Rede de Funções de Base Radial (RBF)

A RBF é uma rede de arquitetura neural que tem praticamente as mesmas aplicações que a rede neural multicamada (MLP), ou seja, aproximação de mapeamento (função) não linear e classificação de padrões (119,120).

Diferentemente da rede MLP, que pode apresentar uma ou mais camadas de neurônios ocultos, a rede RBF, em sua concepção básica, apresenta apenas uma camada oculta, além de uma camada de saída. Além disso, usualmente, a camada oculta tem mais unidades escondidas. Os neurônios da camada oculta, chamados de funções de base radial, possuem função de ativação não linear, que é bem diferente das funções sigmóides que a rede MLP usa. Por outro lado, os neurônios da camada de saída possuem, em geral, função de ativação linear (Figura 11).

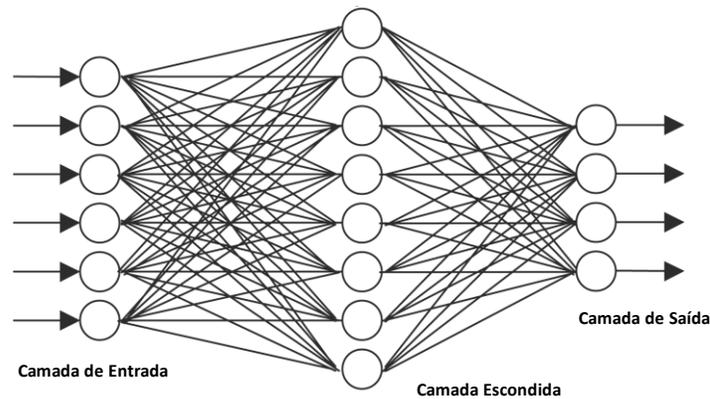


Figura 11. Rede de Funções de Base Radial - arquitetura geral. Apresenta uma camada de entrada, uma de saída e apenas uma camada escondida (121).

O funcionamento da rede RBF envolve basicamente a especificação do número de funções de base, determinação dos seus respectivos parâmetros e determinação dos pesos dos neurônios de saída. A seleção do número apropriado de funções de base é um aspecto crítico das RBF, pois é o que controla a complexidade e a habilidade de generalização da rede. Quando há poucas funções de base, a rede pode não se adaptar aos dados de treinamento adequadamente por uma flexibilidade limitada. Por outro lado, aquelas com muitas funções de

base podem ter pouca habilidade de generalização, pois se tornam muito flexíveis e erroneamente se adaptam aos ruídos existentes nos dados de treinamento. A principal desvantagem da RBF é a complexidade de compreensão da maneira como se obtém os resultados nas camadas de saída (99).

1.3.2.8. *Naive-Bayes*(NB)

O classificador NB é um exemplo de algoritmo de aprendizagem estatístico, pois é completamente baseado em probabilidades. Contrariamente às redes neurais artificiais, ele apresenta um modelo estatístico explícito, que prevê a probabilidade de uma instância pertencer a uma determinada classe, ao invés de simplesmente classificá-la (99).

O classificador é considerado ingênuo (*naive*) por assumir que os atributos são condicionalmente independentes, ou seja, a informação de um evento não se relaciona a nenhum outro, o que simplifica os cálculos envolvidos.

Além disso, ele obtém melhores resultados quando os dados são categóricos ao invés de contínuos (numéricos). Se lidarmos com dados contínuos, é possível usar distribuição normal para calcular as probabilidades, ou devemos estratificar os dados durante o processo de pré-processamento (122).

O objetivo principal do classificador NB é calcular a probabilidade de uma amostra desconhecida pertencer a uma determinada classe entre as classes possíveis, ou seja, determinar a classe mais provável. Ele necessita de um conjunto de treinamento que já foi

classificado previamente e quando recebe uma amostra sem classificação, ele define como saída a classe mais provável para esta amostra de acordo com os cálculos probabilísticos.

Isso é realizado basicamente seguindo três estágios:

- 1) Cálculo das probabilidades das classes – cada classe do conjunto de treinamento (como, glaucoma e normal) possui sua probabilidade calculada. Isso é realizado dividindo-se o número de instâncias de determinada classe (por exemplo, o número de pacientes com glaucoma) pelo número total de instâncias (pacientes) do conjunto de treinamento;
- 2) Cálculo das probabilidades da amostra desconhecida – cada valor de cada atributo da amostra desconhecida possui sua probabilidade calculada para cada possível classe;
- 3) Com as probabilidades de cada classe calculadas, verifica-se qual é a classe que possui maior probabilidade para a amostra desconhecida. Com isso, o algoritmo apresenta como saída a classe que possui maior probabilidade de conter a amostra desconhecida.

Uma diferença importante entre o método NB e outros classificadores é que não há nenhuma pesquisa explícita através do espaço de possíveis modelos para o classificador.

1.3.2.9. *Support Vector Machine* - Linear (SVMl) e *Gaussian* (SVMg).

As máquinas de vetores de suporte - SVM - são embasadas na teoria de aprendizado estatístico, desenvolvida por Vapnik (123,124) a partir de estudos iniciados em 1971. A teoria de aprendizado estatístico estabelece condições matemáticas que auxiliam na escolha de um classificador com bom desempenho para os conjuntos de treinamento e teste. Ou seja, busca-

se uma função capaz de classificar os dados de treinamento da forma mais correta possível, sem dar atenção exacerbada a pontos individuais.

O objetivo do processo de aprendizagem estatística é, então, obter uma função indicadora que minimize o risco funcional, por meio de relações extraídas deste conjunto. O risco funcional é a probabilidade de a saída desejada ser diferente da saída obtida pela máquina após a escolha de uma função indicadora. Para que isto ocorra, busca-se um compromisso entre a minimização do erro de treinamento e a complexidade da função escolhida para a separação dos dados em classes. Essas condições são alcançadas através do controle da dimensão Vapnik-Chervonenkis (dimensão VC) do conjunto de funções do qual o classificador deve ser extraído. A dimensão VC não tem conexão com a noção geométrica de dimensão, é uma medida da capacidade ou poder de expressão de um conjunto de funções, ou seja, mede a complexidade das hipóteses examinadas pelo algoritmo na busca de uma solução.

SVM lineares podem ser utilizadas para classificação de conjuntos de treinamento linearmente separáveis. Assim, é possível separar os padrões de classes diferentes contidos no mesmo conjunto por pelo menos um hiperplano (Figura 12) (101).

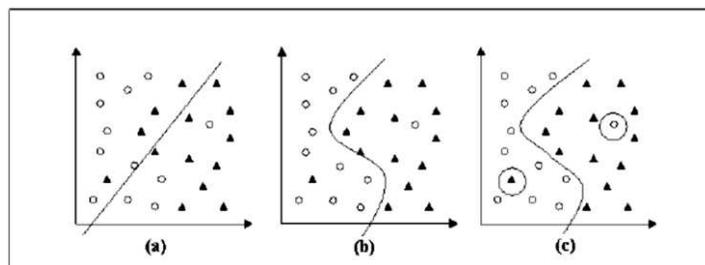


Figura 12. Exemplo de conjunto de dados de treinamento pertencentes a duas classes (triângulos e círculos), classificado segundo três diferentes hipóteses (modificado de Lorena AC, 2007) (125).

Em algumas situações, é difícil encontrar aplicações cujos conjuntos sejam linearmente separáveis. Isso se deve a diversos fatores, dentre eles a presença de ruídos nos dados ou a própria natureza do problema, que pode ser não-linear (Figura 13).

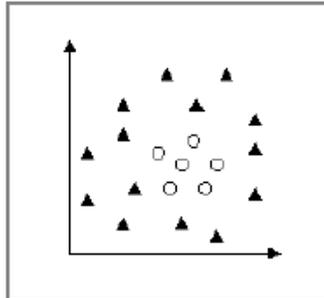


Figura 13. Exemplo de conjunto não linearmente separável (modificado de Lorena AC, 2007) (125).

As SVMs lineares podem ser generalizadas de forma a realizar classificações não lineares. Para isso, utiliza-se o artifício de levar os dados de treinamento para um espaço de maior dimensão. Os dados neste espaço se tornam linearmente separáveis (Figura 14). Pode-se então aplicar as SVMs lineares sobre os dados mapeados neste espaço, determinando um hiperplano de margem maximizada. A SVM não linear incorpora este conceito por meio da utilização de funções Kernel, que permitem o acesso a espaços complexos (em alguns casos infinitos) de forma simplificada. Alguns dos Kernels mais utilizados na prática são os polinomiais, os Gaussianos e os sigmoidais (126,127).

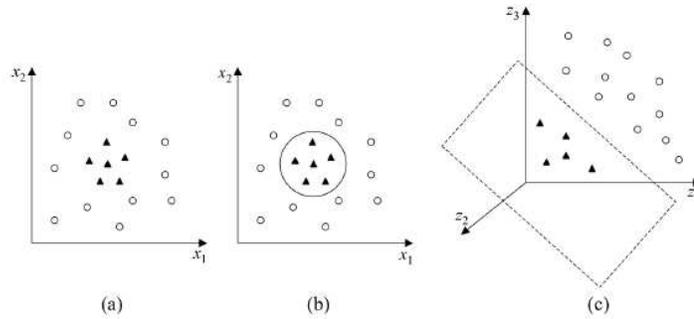


Figura 14. Exemplo de criação de um espaço com maior dimensão separável por um hiperplano. (a): exemplo de conjunto não linearmente separável; (b): fronteira não linear onde é possível separar as duas classes; (c): criação de um espaço de maior dimensão possibilitando a separação entre as classes por um hiperplano (modificado de Lorena AC, 2007) (125).

1.4. Análise de Desempenho (sensibilidade e especificidade)

Métodos estatísticos aplicados à medicina diagnóstica estão voltados ao problema de classificar indivíduos em grupos que indicam a presença ou ausência de uma determinada doença. A probabilidade de um teste diagnóstico produzir um resultado positivo, dado que o indivíduo é portador da doença, é chamada sensibilidade; e a probabilidade do teste produzir um resultado negativo, dado que o indivíduo não é portador da doença, é chamada especificidade (128,129).

A relação entre a sensibilidade e a especificidade evidencia o desempenho de um teste e pode ser demonstrada por meio da curva ROC (*receiver operating characteristic*). Esta análise é realizada por meio de um método gráfico simples e robusto que permite estudar a variação da sensibilidade e especificidade para diferentes valores de corte (Figura 15) (130).

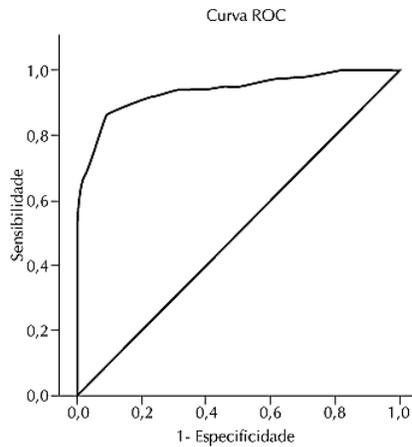


Figura 15. Curva ROC.

As curvas ROC foram desenvolvidas no campo das comunicações como uma forma de demonstrar as relações entre sinal e ruído (131). Interpretando o sinal como os verdadeiros positivos (sensibilidade) e o ruído como os falsos positivos (1-especificidade), podemos entender como esse conceito pode ser aplicado em medicina.

A linha diagonal, que pode ser vista na figura 15, corresponde a um teste que é positivo ou negativo aleatoriamente. A curva ROC permite evidenciar os valores para os quais existe maior otimização da sensibilidade em função da especificidade, que corresponde ao ponto em que se encontra mais próxima do canto superior esquerdo do diagrama, uma vez que o índice de positivos verdadeiros é um e o de falsos positivos é zero (132).

O conhecimento da área sob a curva ROC (aROC) possibilita quantificar a acurácia de um teste diagnóstico (proporcional à área sob a curva), além de permitir a comparação com outros testes. Dessa maneira, a aROC constitui um dos índices mais usados para sintetizar a qualidade da curva, sendo um resumo estatístico útil para a determinação da acurácia do teste.

Assim, um método totalmente incapaz de discriminar indivíduos doentes e não doentes, teria uma área sob a curva de 0,50 (seria a hipótese nula). Quanto maior a capacidade do teste em discriminar os indivíduos, mais a curva se aproximaria do canto superior esquerdo do gráfico, e a área sob a curva seria próxima de 1,0. Considera-se $aROC > 0,7$ como indicadora de um desempenho satisfatório (133).

Dessa maneira, a aROC pode ser aplicada para avaliar o desempenho diagnóstico para cada parâmetro do OCT e também para os MLCs (132).

1.5. Justificativa

O glaucoma é uma neuropatia óptica progressiva que causa danos irreversíveis ao nervo óptico (1). O exame de campo visual acromático revela defeitos após a lesão de pelo menos 30-50% das fibras nervosas da retina (32). Dessa maneira, o diagnóstico precoce é essencial para evitar a progressão da doença num momento em que a perda de CV não é pronunciada a ponto de afetar a qualidade de vida do paciente.

A análise da CFNR possibilita o diagnóstico precoce da doença e pode ser avaliada por métodos objetivos. Assim, diversas tecnologias, dentre elas o SD-OCT, têm sido desenvolvidas para auxiliar o oftalmologista nesta tarefa, mas a quantidade de informação gerada é muito grande (38–44,50,51). Por isso, torna-se necessário o conhecimento dos dados mais relevantes para o diagnóstico da doença. Além disso, por meio da computação, é possível resumir o complexo número de dados disponíveis em uma única saída que represente a probabilidade ou possibilidade de glaucoma. Isso é realizado pela inteligência artificial por meio do processo de aprendizagem de máquina.

Dentro do nosso conhecimento, não há nenhum estudo na literatura avaliando o uso de classificadores de aprendizagem de máquina com os dados obtidos pelo SD-OCT para o diagnóstico de glaucoma, o que justifica a realização desse trabalho.

1.6. Objetivos

O presente estudo tem como objetivos:

- 1) Determinar a sensibilidade e especificidade dos parâmetros fornecidos pelo SD-OCT para o diagnóstico de glaucoma;
- 2) Determinar a sensibilidade e especificidade dos classificadores de aprendizagem de máquina treinados com dados do SD-OCT para o diagnóstico de glaucoma;
- 3) Comparar a capacidade diagnóstica das análises descritas acima.

METODOLOGIA

2. METODOLOGIA

2.1. População

Desenvolveu-se um estudo prospectivo observacional transversal que incluiu 110 olhos de 110 participantes (62 pacientes com glaucoma e 48 indivíduos saudáveis) com mais de 40 anos. Estes indivíduos foram examinados no Ambulatório de Glaucoma do Hospital das Clínicas da Universidade Estadual de Campinas (HC-UNICAMP), entre agosto de 2008 e novembro de 2010. Obteve-se, por escrito, consentimento de todos os participantes, após o esclarecimento dos mesmos quanto aos aspectos éticos e clínicos do estudo (apêndice 2). Todos os procedimentos foram realizados conforme a Declaração de Helsinque e a resolução 466/12 do Conselho Nacional de Saúde (196/96 revogada). O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências Médicas da Universidade Estadual de Campinas, parecer nº 406/2005 (anexo 1).

Todos os indivíduos incluídos no estudo tiveram registrados os seus dados demográficos, tais como, idade, sexo, raça e antecedentes oftalmológicos. A definição de raça foi baseada no registro hospitalar do HC-UNICAMP ou através da autodenominação. Foram considerados caucasianos os registrados como brancos, e afrodescendentes os negros, mulatos, cafuzos ou pardos.

Subsequentemente, cada participante foi submetido a uma avaliação oftalmológica completa, incluindo:

- 1) Revisão do histórico médico;

- 2) Acuidade visual com a melhor correção utilizando a Tabela de Snellen. Posteriormente, foi realizada conversão da medida decimal em LogMAR (*logarithm of the minimum angle of resolution*) para permitir as análises comparativas (anexo 2) (134);
- 3) Biomicroscopia à lâmpada de fenda Haag-Streit modelo 900 (Haag-Streit, Koeniz, Suíça). Neste procedimento, analisaram-se as características da câmara anterior, do cristalino e o aspecto do nervo óptico com uma lente de 78 D, após midríase medicamentosa obtida com uma gota de tropicamida 1%;
- 4) Pressão intraocular (PIO) medida com tonômetro de aplanção de Goldmann (modelo R900, Haag-Streit, Koeniz, Suíça), após anestesia tópica com uma gota de oxibuprocaína 0,4% e instilação de uma gota de fluoresceína sódica 1%;
- 5) Gonioscopia com lente de Posner, após anestesia tópica com uma gota de oxibuprocaína 0,4%;
- 6) PAP com estratégia SITA Standard 24-2 (Humphrey Field Analyzer II, Carl Zeiss Meditec, Inc., Dublin, CA);
- 7) Imagem da CFNR peripapilar utilizando o Cirrus SD-OCT (Cirrus HD-OCT; Carl Zeiss Meditec Inc., Dublin, Califórnia).

Para serem incluídos no estudo, os participantes deveriam apresentar uma acuidade visual corrigida melhor ou igual a 20/40, erro refracional esférico de $\pm 5,0$ D, erro refracional cilíndrico dentro de $\pm 3,0$ D, ângulo aberto à gonioscopia e PAP confiável. A confiabilidade do exame de perimetria computadorizada foi definida como: erros falso positivos $< 20\%$, erros falso negativos $< 20\%$ e perdas de fixação $< 20\%$.

Foram excluídos indivíduos com doenças retinianas concomitantes, uveíte, neuropatia óptica não glaucomatosa, pseudofacia ou afacia, cirurgia refrativa, história familiar de glaucoma para os indivíduos saudáveis e catarata significativa de acordo com os critérios do Sistema de Classificação de Opacificação do Cristalino III (135) (Figura 16), definida como a máxima opacidade nuclear (NC3, NO3), cortical (C3) e subcapsular (P3).

Para o grupo normal, um olho foi selecionado aleatoriamente, se ambos os olhos fossem elegíveis. Já no grupo com glaucoma, se os dois olhos pudessem ser incluídos no estudo, selecionava-se o olho com dano mais precoce no campo visual baseado no valor de MD (*mean deviation*).

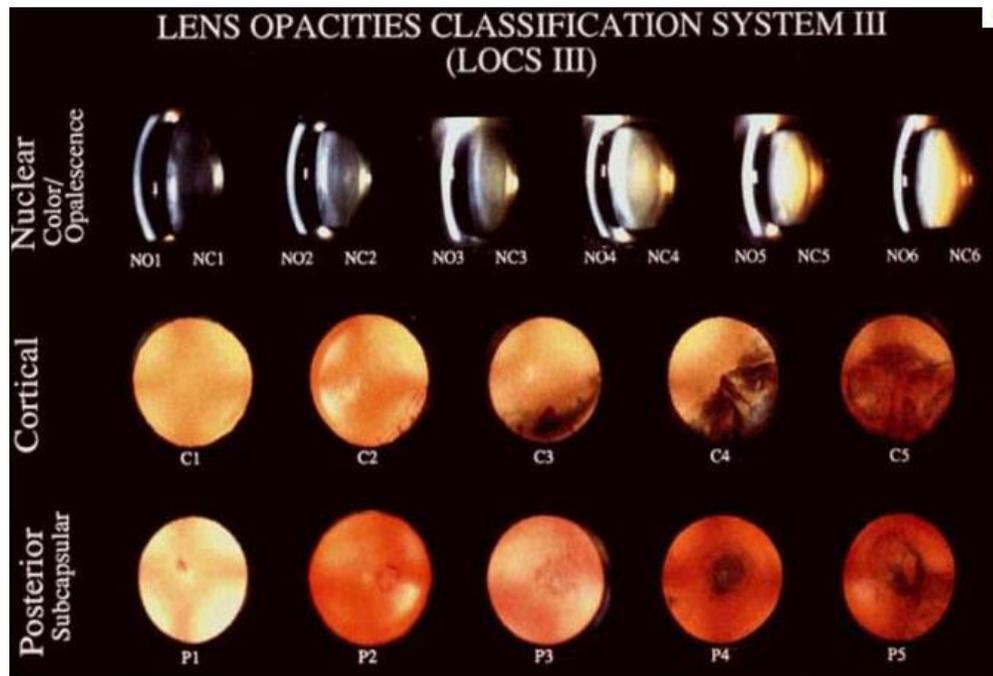


Figura 16. Sistema de Classificação de Opacificação do Cristalino III (*Lens Opacification Classification System* III – LOCS III, modificado de Chylack et al, 1993) (135).

2.1.1. Critérios de Inclusão para Indivíduos Saudáveis

Os critérios de inclusão para indivíduos saudáveis foram: PIO \leq 21 mmHg, sem história de pressão intraocular elevada, CNO de aspecto normal e dois CVs consecutivos confiáveis e normais.

Este grupo foi formado por acompanhantes de pacientes que compareciam para avaliação no serviço de oftalmologia do Hospital das Clínicas da FCM-UNICAMP ou por funcionários do Hospital das Clínicas.

2.1.2. Critérios de Inclusão para Indivíduos com Glaucoma

Olhos glaucomatosos foram definidos como aqueles com história de duas ou mais medidas de PIO $>$ 21mmHg, alterações glaucomatosas na CNO, e dois exames de CV confiáveis mostrando defeito glaucomatoso, com o exame mais recente dentro de 3 meses da data de inclusão. Os pacientes que formaram esse grupo eram acompanhados no setor de glaucoma do serviço de oftalmologia do Hospital das Clínicas da FCM-UNICAMP e, após preencherem os critérios de inclusão, aceitaram participar do estudo.

Classificamos como alterações glaucomatosas do NO a presença de pelo menos dois dos seguintes critérios:

- 1) Assimetria da relação escavação/disco maior que 0,2;
- 2) Atrofia focal da rima neural (*notch* ou *notching*);
- 3) Hemorragia peripapilar;

4) Relação escavação/disco maior que 0,6.

Defeito de CV típico de glaucoma foi definido como aquele que reunisse dois dos seguintes critérios:

1) Grupo de três pontos não periféricos no gráfico *Pattern Deviation* com probabilidade < 5% em um único hemisfério, com pelo menos 1 dos pontos com probabilidade < 1 %;

2) *Pattern standard deviation* (PSD) fora do limite normal de 95%.

3) *Glaucoma hemifield test* (GHT) fora do limite normal de 99%;

A gravidade do dano glaucomatoso avaliado pelo CV foi determinada de acordo com o valor do índice global *mean deviation* (MD), baseado no sistema de estadiamento de Hodapp-Parrish-Anderson (136):

- a) Dano inicial: $MD \geq -6\text{dB}$;
- b) Dano moderado: $-6\text{dB} > MD \geq -12\text{dB}$;
- c) Dano avançado: $MD < -12\text{dB}$.

Apenas os pacientes com dano inicial ou moderado foram incluídos no estudo, pois a inclusão de indivíduos com doença avançada aumentaria artificialmente a capacidade diagnóstica do teste que queremos avaliar.

2.2. Tomografia de Coerência Óptica

Obteve-se uma imagem da CFNR peripapilar de todos os participantes com o Cirrus SD-OCT (software versão 3.0, Carl Zeiss Meditec, Inc.), que utiliza tecnologia de domínio

espectral. Utilizou-se para análise o modo cubo 200x200 do disco óptico. Todas as imagens foram obtidas com olhos não dilatados, por um único oftalmologista bem treinado (VGV), que era mascarado para o diagnóstico. O Cirrus SD-OCT fornece mapas da CFNR com 17 parâmetros: espessura média, espessura média dos quatro quadrantes (superior, inferior, nasal e temporal) e espessura média a cada hora de relógio. Todos os dados foram alinhados de acordo com a orientação do olho direito. Portanto, 9 horas de relógio representou o lado temporal do disco óptico para ambos os olhos. Somente exames bem centrados e com uma intensidade de sinal ≥ 7 foram incluídos.

2.3. Classificadores de Aprendizagem de Máquina

Com base nos dados obtidos com o Cirrus SD-OCT, os classificadores de aprendizagem de máquina foram desenvolvidos utilizando os seguintes algoritmos já descritos anteriormente: Árvore de Decisão (CTREE), *Random Forest* (RAN), *Bagging* (BAG), *Adaptative Boosting* (ADA), *Ensemble Selection* (ENS), *Multilayer Perceptrons* (MLP), Rede de Funções de Base Radial (RBF), *Naive-Bayes* (NB), *Support Vector Machine - Linear* (SVMl) e *Support Vector Machine - Gaussian* (SVMg).

Inicialmente, os MLCs foram treinados de maneira supervisionada utilizando os 17 parâmetros do SD-OCT. Posteriormente, uma seleção retrógrada foi utilizada para encontrar o número mínimo de parâmetros que resultasse no melhor desempenho para cada classificador.

Assim, a análise começou com a avaliação de desempenho do classificador sobre o conjunto completo de recursos contendo os 17 atributos do SD-OCT. Sequencialmente, o atributo que apresentou a menor aROC, calculada sobre os dados do SD-OCT isoladamente,

foi retirado e a precisão do classificador foi novamente computada. Este processo de redução de dimensão, com base em áreas sob as curvas ROC dos parâmetros do SD-OCT classificados em ordem decrescente, foi realizado 16 vezes. O processo começou com a exclusão do parâmetro com a menor aROC e foi interrompido quando o parâmetro com a maior aROC foi usado.

O critério para avaliar os algoritmos envolveu a análise de curvas ROC geradas a partir dos resultados de diversas tarefas de classificação para cada classificador.

Os dados foram transferidos para o servidor localizado no Laboratório de Engenharia do Conhecimento (KNOMA) do Programa de Computação e Sistemas Digitais da Universidade de São Paulo.

2.4. Validação Cruzada

Para maximizar a utilização dos nossos dados coletados, o método de validação cruzada foi empregado (100). Esse método substitui a necessidade de treinar os classificadores com uma população e, depois, testar a sua performance em uma população diferente.

Assim, todos os dados (110 indivíduos incluídos no estudo) foram aleatoriamente divididos em 10 subgrupos, cada um contendo aproximadamente a mesma proporção de medições de espessura da CFNR de indivíduos saudáveis e glaucomatosos. Nove subgrupos foram utilizados como dados de treino, enquanto o subconjunto remanescente foi utilizado como teste. Este processo foi realizado 10 vezes, até que cada um dos 10 subgrupos tivesse

sido utilizado como um conjunto de teste. Os resultados dos testes foram calculados para estimar a precisão do classificador.

2.5. Análise Estatística

Todas as análises foram realizadas utilizando o *software* MedCalc versão 11.0.1 e 12.7.4 (MedCalc Software, Mariakerke, Bélgica). Após confirmar a distribuição normal da amostra por meio do teste D'Agostino-Pearson, aplicamos o teste t de *Student* para comparar as variáveis contínuas, enquanto as variáveis categóricas foram analisadas utilizando o teste qui-quadrado ou o teste exato de Fisher.

A curva ROC foi obtida para todos os parâmetros do SD-OCT (espessura média, quatro quadrantes e 12 horas de relógio) e para cada classificador de aprendizagem de máquina, antes e depois da otimização. As áreas abaixo das curvas foram comparadas utilizando o teste z (razão crítica z, Hanley e McNeil, 1983).

O nível de significância estatística foi determinado como $p < 0,05$.

RESULTADOS

3. RESULTADOS

Cento e dez olhos de 110 indivíduos foram incluídos neste estudo. Entre os 110 indivíduos, 62 pacientes tinham diagnóstico de glaucoma e 48 eram indivíduos saudáveis. A Tabela 1 mostra as características demográficas da população estudada. A média de idade foi de $57,0 \pm 9,2$ anos para os indivíduos saudáveis e de $59,9 \pm 9,0$ anos para pacientes com glaucoma ($p=0,103$). Não houve diferença significativa entre os grupos controle e glaucoma em relação à pressão intraocular (PIO) ($14,8 \pm 2,8$ mmHg e $13,8 \pm 2,5$ mmHg, respectivamente) ($p=0,062$), embora os pacientes com glaucoma estivessem utilizando um número médio de $2,2 \pm 1,2$ medicamentos para reduzir a PIO. Valores médios de MD foram $-4,1 \pm 2,4$ dB para pacientes com glaucoma e $-1,5 \pm 1,6$ dB para indivíduos saudáveis ($p<0,001$).

No grupo de pacientes com glaucoma, 82,3% ($n = 51$) foram classificados como tendo dano inicial no CV e 17,7% ($n = 11$) como tendo dano moderado.

Tabela 1. Características demográficas da população estudada.

	Normal (n=48)	Glaucoma (n=62)	p
Idade (anos; média±DP)	57,0 ± 9,2	59,9 ± 9,0	0,103
Idade (anos; variação)	45 – 82	43 – 78	-
Sexo (Masc.[%]:Fem.[%])	23[47,9] : 25[52,1]	31[50] : 31[50]	0,830
Raça (Caucasianos[%]:Afro-descendentes[%])	37[77,1] : 11[22,9]	46[74,2] : 16[25,8]	0,461
AV LogMAR (média±DP)	0,04 ± 0,09	0,1 ± 0,1	0,003
EE (D; média±DP)	1,53 ± 2,15	1,2 ± 1,9	0,467
PIO (mmHg; média±DP)	14,8 ± 2,8	13,8 ± 2,5	0,062
Medicações (média±DP)	0	2,2 ± 1,2	<0,001
MD (dB; média±DP)	-1,5 ± 1,6	-4,1 ± 2,4	<0,001
PSD (dB; média±DP)	1,8 ± 0,8	4,3 ± 2,4	<0,001

DP – desvio padrão; LogMAR – *logarithm of the minimum angle of resolution*; AV - acuidade visual; EE - equivalente esférico; D - dioptrias; dB - decibéis; MD - *mean deviation*; PSD - *pattern standard deviation*.

A Tabela 2 compara os valores médios da espessura da CFNR em ambos os grupos obtidos pelo Cirrus-SD-OCT. A média da espessura da CFNR foi significativamente menor no grupo de glaucoma, exceto para as posições de 3, 8 e 9 horas.

Tabela 2. Média das medidas da espessura da camada de fibras nervosas retiniana (\pm desvio padrão) obtidas pelo Cirrus-SD-OCT nos dois grupos.

OCT	Normal (n=48)	Glaucoma (n=62)	P
Espessura média global (μm)	93,5 \pm 9,9	80,6 \pm 10,8	<0,001
<u>Quadrantes</u>			
Temporal (μm)	63,1 \pm 9,7	56,6 \pm 12,9	0,004
Superior (μm)	115,1 \pm 21,5	100,9 \pm 17,4	<0,001
Nasal (μm)	72,7 \pm 10,7	65,8 \pm 9,8	<0,001
Inferior (μm)	121,0 \pm 17,1	99,4 \pm 17,5	<0,001
<u>Horas de relógio</u>			
1 (μm)	104,0 \pm 18,5	91,4 \pm 19,1	<0,001
2 (μm)	90,7 \pm 16,9	77,6 \pm 13,3	<0,001
3 (μm)	61,2 \pm 10,2	58,6 \pm 10,4	0,188
4 (μm)	66,1 \pm 12,5	61,1 \pm 13,0	0,047
5 (μm)	99,6 \pm 20,4	87,4 \pm 16,8	<0,001
6 (μm)	133,3 \pm 25,2	109,4 \pm 24,4	<0,001
7 (μm)	130,1 \pm 23,7	101,4 \pm 31,5	<0,001
8 (μm)	63,6 \pm 12,1	57,9 \pm 16,5	0,050
9 (μm)	50,9 \pm 10,7	46,5 \pm 14,0	0,074
10 (μm)	75,0 \pm 13,9	65,2 \pm 14,7	<0,001
11 (μm)	124,3 \pm 22,1	103,5 \pm 24,9	<0,001
12 (μm)	123,3 \pm 26,1	107,6 \pm 24,8	0,001

Os parâmetros do SD-OCT com os maiores aROCs foram: quadrante inferior (0,813 - IC 95% 0,727-0,881), espessura média (0,807 - IC 95% 0,721-0,876), 7 horas de relógio (0,765 - IC 95% 0,674-0,840) e 6 horas de relógio (0,754 - IC 95% 0,663-0,831) (Tabela 3). Para obter uma especificidade fixa de 80%, as melhores sensibilidades foram observadas com os seguintes parâmetros: posição de 7 horas (64,5%), espessura média (62,1%), quadrante inferior (61,3%) e quadrante superior (56,4%) (Tabela 3).

Tabela 3. Áreas abaixo da curva ROC (aROC) para cada parâmetro do Cirrus-SD-OCT e sensibilidade (%) com especificidade fixa em 80% e 90%.

OCT	aROC (IC 95%)	Especificidade 80%	Especificidade 90%
Espessura média global	0,807 (0,721-0,876)	62,1	54,0
<u>Quadrantes</u>			
Temporal	0,675 (0,579-0,761)	50,0	33,1
Superior	0,737 (0,645-0,816)	56,4	46,7
Nasal	0,685 (0,590-0,771)	49,2	26,6
Inferior	0,813 (0,727-0,881)	61,3	53,2
<u>Horas de relógio</u>			
1	0,703 (0,608-0,786)	45,2	29,8
2	0,723 (0,630-0,804)	51,6	34,7
3	0,574 (0,476-0,668)	29,8	22,6
4	0,605 (0,507-0,696)	29,0	11,3
5	0,671 (0,575-0,757)	43,5	27,4
6	0,754 (0,663-0,831)	47,6	33,9
7	0,765 (0,674-0,840)	64,5	38,7
8	0,631 (0,534-0,721)	42,7	27,4
9	0,625 (0,527-0,715)	44,3	36,3
10	0,699 (0,604-0,783)	50,8	42,0
11	0,740 (0,648-0,819)	40,3	33,9
12	0,672 (0,576-0,759)	38,7	21,0

IC – intervalo de confiança

As aROCs dos classificadores de aprendizagem de máquina variaram de 0,569 (CTREE, IC 95% 0,462-0,676) a 0,795 (SVMg, IC 95% 0,713-0,877) (Tabela 4), quando treinados com todos os parâmetros do SD-OCT. Para obter uma especificidade fixa de 80%, as melhores sensibilidades foram observadas com o ENS (73,4%), o NB (66,1%) e o BAG (64,5%) (Tabela 4).

Tabela 4. Áreas abaixo da curva ROC (aROC) para os classificadores de aprendizagem de máquina (MLC) e sensibilidade (%) com especificidade fixa em 80% e 90%.

MLC	aROC (IC 95%)	Especificidade 80%	Especificidade 90%
BAG	0,752 (0,662 – 0,842)	64,5	40,3
NB	0,788 (0,704 – 0,872)	66,1	56,4
SVMI	0,788 (0,704 – 0,872)	62,9	56,4
SVMg	0,795 (0,713 – 0,877)	62,9	56,4
MLP	0,591 (0,485 – 0,697)	59,7	51,6
RBF	0,672 (0,572 – 0,772)	57,2	50,0
RAN	0,758 (0,669 – 0,847)	61,3	46,8
ENS	0,721 (0,627 – 0,815)	73,4	46,8
CTREE	0,569 (0,462 – 0,676)	59,7	52,4
ADA	0,744 (0,653 – 0,835)	62,1	50,0

IC – intervalo de confiança; BAG – *Bagging*; NB – *Naive-Bayes*; SVMI – *Support Vector Machine*- Linear; SVMg - *Support Vector Machine*- *Gaussian*; MLP – *Multilayer Perceptron*; RBF – Rede de Funções de Base Radial; RAN – *Random Forest*; ENS – *Ensemble Selection*; CTREE – Árvore de Decisão; ADA – *Adaptative Boosting*.

Todos os classificadores apresentaram melhor desempenho quando treinados com um número menor de parâmetros. Assim, quando menos parâmetros foram utilizados, as aROCs variaram de 0,785 (ADA, IC 95% 0,701-0,869) a 0,818 (BAG, IC 95% 0,740-0,896) (Tabela 5). No entanto, essa melhora só foi estatisticamente significativa para os classificadores MLP ($p=0,002$) e CTREE ($p=0,001$).

O SVMg foi o classificador que obteve a melhor aROC (0,795) quando utilizamos todos os parâmetros como dados de treinamento. Ao selecionar um número menor de dados para treinamento, o BAG foi o classificador que obteve a melhor aROC (0,818). Apesar dos valores sugerirem uma pequena melhora, ela não foi estatisticamente significativa ($p=0,690$).

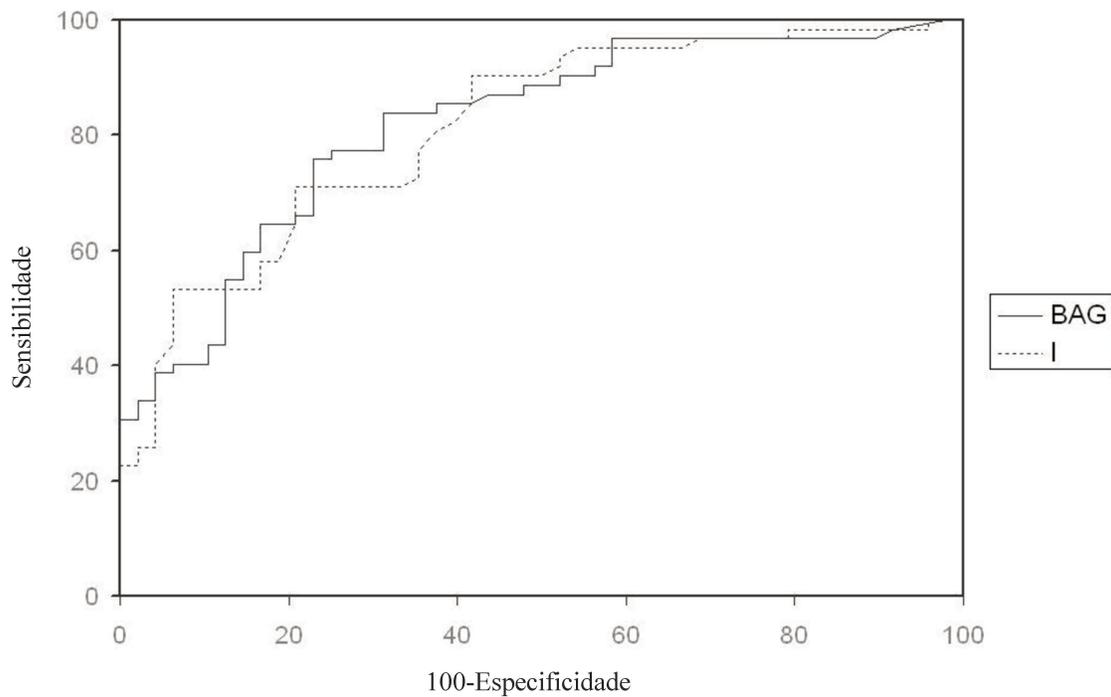
Tabela 5. Áreas abaixo da curva ROC (aROC) para os classificadores de aprendizagem de máquina (MLC) treinados com menos parâmetros do Cirrus-SD-OCT.

MLC	aROC (nº parâmetros)	95% IC
BAG	0,818 (2)	0,740 – 0,896
NB	0,817 (2)	0,739 – 0,895
SVMI	0,815 (2)	0,737 – 0,893
SVMg	0,813 (2)	0,734 – 0,892
MLP	0,804 (1)	0,723 – 0,885
RBF	0,801 (1)	0,720 – 0,882
RAN	0,797 (7)	0,715 – 0,879
ENS	0,797 (1)	0,715 – 0,879
CTREE	0,789 (1)	0,706 – 0,872
ADA	0,785 (2)	0,701 – 0,869

IC – intervalo de confiança; BAG – *Bagging*; NB – *Naive-Bayes*; SVMI – *Support Vector Machine* Linear; SVMg - *Support Vector Machine* Gaussian; MLP – *Multilayer Perceptron*; RBF – Rede de Funções de Base Radial; RAN – *Random Forest*; ENS – *Ensemble Selection*; CTREE – Árvore de Decisão; ADA – *Adaptive Boosting*.

A melhor aROC obtida pelos classificadores foi com o BAG quando treinado com dois parâmetros (0,818). Comparando esse resultado com a melhor aROC obtida com o parâmetro do SD-OCT isolado (quadrante inferior - 0,813), não houve diferença estatisticamente significativa ($p=0,932$) (Figura 17).

Figura 17. Área abaixo da curva ROC (aROC) para o melhor classificador treinado com a quantidade de parâmetros que permitiu melhor performance (BAG = 0,818) e aROC do melhor parâmetro do SD-OCT (I – quadrante inferior = 0,813) ($p=0,932$).



DISCUSSÃO

4. DISCUSSÃO

Muitos estudos têm sido publicados avaliando a sensibilidade e especificidade do TD-OCT, e alguns compararam a capacidade de diagnóstico do TD-OCT e do SD-OCT, demonstrando que ambos apresentam um desempenho bom e semelhante (82–88), conforme o Quadro 1.

Quadro 1. Estudos que compararam a capacidade diagnóstica do Stratus-TD-OCT e do Cirrus-SD-OCT.

Autores	População	Resultados: aROC da espessura média global	p
Chang et al (82)	54 glaucomatosos 50 normais	Stratus – 0,889 Cirrus – 0,904	p>0,05
Jeoung et al (83)	55 glaucoma pré-perimétrico 55 normais	Stratus – 0,697 Cirrus – 0,716	p=0,469
Park et al (84)	100 glaucomatosos 74 normais	Stratus – 0,934 Cirrus – 0,953	p=0,15
Moreno-Montañés et al (85)	86 glaucomatosos 130 normais	Stratus – 0,829 Cirrus – 0,837	p=0,706
Leung et al (86)	83 glaucomatosos 97 normais	Stratus – 0,956 Cirrus – 0,962	p>0,05
Huang et al (87)	30 glaucoma pré-perimétrico 47 glaucoma inicial 62 normais	Glaucoma pré-perimétrico: Stratus – 0,881 Cirrus – 0,951 Glaucoma inicial	p=0,006 p>0,05
Hong et al (88)	129 glaucomatosos 59 normais	Stratus – 0,764 Cirrus – 0,795 (quadrante superior)	p=0,281

aROC - área abaixo da curva ROC; TD – *time domain*; SD – *spectral domain*.

Técnicas de MLC têm sido empregadas com diversas tecnologias, criadas para realizar a avaliação estrutural e funcional no glaucoma, incluindo TD-OCT (53,57,59,62,137), HRT (58,60,61,64,68,69,137–139), GDx (56,70), e o campo visual (52,54,55,59–61,63–67,71,140,141).

Embora quatro grupos (53,57,62,137) tenham empregado MLCs a partir de dados obtidos com o TD-OCT, não há nenhum estudo que avalie o uso de MLCs em conjunto com os dados do SD-OCT.

Bizios et al (57) utilizaram dados fornecidos pelo TD-OCT para treinar dois MLCs (MLP e SVM) com o objetivo de comparar a capacidade diagnóstica entre os classificadores. Além dos parâmetros fornecidos pelo aparelho (espessura média, 4 quadrantes, 12 horas de relógio, espessura superior máxima, espessura inferior máxima, e a diferença entre os pontos de maior e menor espessura), os autores calcularam oito novos parâmetros por considerá-los mais sensíveis no diagnóstico da doença. Assim, foram calculados: espessura superior mínima, espessura inferior mínima, percentil 10 e 90 dos 64 *A-scans* do quadrante superior e inferior, diferença entre os percentis 10 e 90 da espessura média, e a transformação dos 256 *A-scans* fornecidos pelo aparelho por um algoritmo de redução de dimensionalidade, em que o valor não corresponde a nenhum *A-scan* particular, mas reflete todas as 256 medidas fornecidas pelo TD-OCT (Bizios et al. 2007; Invest Ophthalmol Vis Sci 48: ARVO E-Abstract 525). Os autores observaram que, apesar das diferenças na forma como MLP e SVM executam tarefas de classificação, os dois métodos foram bem sucedidos no reconhecimento de padrões pertinentes nos dados de treinamento e na aplicação desse conhecimento para classificar novos dados, alcançando desempenho semelhante. As melhores aROCs foram 0,982 para MLP (IC 95% 0,966-0,999) e 0,989 para SVM (IC 95% 0,979-1,0) quando treinados com valores de

espessura transformados a partir dos 256 *A-scans*. Quando os classificadores foram treinados com os parâmetros convencionais, o melhor resultado foi obtido ao utilizar os 17 parâmetros em conjunto (espessura média, quatro quadrantes e 12 horas de relógio), obtendo aROCs de 0,977 para MLP (IC 95% 0,958-0,995) e 0,977 para SVM (IC 95% 0,959-0,999). Neste estudo, foram incluídos os pacientes com diagnóstico de glaucoma de ângulo aberto com dano no campo visual inicial ou moderado ($MD \geq -12\text{dB}$, $PSD < 5\%$ e GHT fora dos limites de normalidade) correspondente às alterações glaucomatosas na CNO e/ou na CFNR julgado por exame de fotografia.

Diferentemente da nossa avaliação que priorizou a análise da comparação entre a eficácia dos MLCs e dos dados fornecidos pelo OCT isolados, Bizios et al (57) comparam a performance entre dois classificadores e concluíram que o tipo de dados fornecidos a eles para treinamento são possivelmente mais importantes do que o método de classificação em si. Isso porque houve uma tendência de melhora na análise quando foram alterados os dados de treinamento.

Burgansky-Eliash et al (53) também utilizaram o Stratus-OCT e demonstraram que, entre os 38 parâmetros fornecidos pela máquina, que incluíam dados da mácula, do nervo óptico e da CFNR peripapilar, a maior aROC foi obtida com a área da rima do NO (0,969), seguida pela espessura média da CFNR peripapilar (0,938). O melhor classificador foi o SVM usando oito parâmetros para treinamento (aROC = 0,981). Semelhante aos nossos resultados, demonstraram que a redução do número de parâmetros fornecidos pelo OCT e utilizados para treinamento dos MLC pode melhorar a performance de classificação. No caso do SVM, eles obtiveram aROC de 0,948 quando treinado com todos os 38 parâmetros, melhorando para 0,981 quando foram escolhidos os parâmetros que mais se correlacionavam com o MD obtido

no exame de CV (oito parâmetros). Apesar dessa melhora ter sido observada, ela não foi estatisticamente significativa. No nosso caso, a melhora só foi significativa em 2 dos 10 classificadores testados (MLP: aROC de 0,591 para 0,804 - $p=0,002$; CTREE: aROC de 0,569 para 0,789 - $p=0,001$). Além disso, a relação sensibilidade e especificidade obtida com o melhor parâmetro individual e com os classificadores não foram significativamente diferentes ($p=0,05$), um achado que também observamos em nossa série.

No presente estudo (apêndice 3), não utilizamos dados do nervo óptico como no estudo anteriormente citado porque, no momento da aquisição das imagens, o aparelho utilizado (Cirrus-SD-OCT, software versão 3.0, Carl Zeiss Meditec, Inc.) ainda não possuía o *software* de avaliação de dados do nervo óptico. No estudo de Burgansky-Eliash et al (53) o melhor parâmetro entre as medidas de CFNR peripapilar, foi a espessura média global (0,938), conforme citado acima. No nosso estudo, este foi o segundo dado de melhor performance diagnóstica (0,807). Os autores compararam esse dado com o melhor resultado obtido com o MLC (SVM treinado com oito parâmetros = 0,981) e não encontraram diferença estatisticamente significativa ($p=0,05$), assim como no nosso caso.

Em nosso conhecimento, o presente estudo foi o primeiro a empregar MLCs com os dados obtidos a partir de SD-OCT. Nós comparamos o desempenho dos parâmetros do SD-OCT e dos MLCs em discriminar entre os olhos saudáveis e glaucomatosos. O melhor parâmetro individual foi o quadrante inferior (aROC = 0,813), o melhor classificador treinado com todos os parâmetros do OCT foi o SVMg (aROC = 0,795) e o melhor classificador treinado com alguns parâmetros do SD-OCT foi o BAG (aROC = 0,818). Embora a aROC tenha sido um pouco maior para este classificador, a diferença não foi estatisticamente significativa quando comparada ao quadrante inferior ($p=0,932$). Assim, não houve nenhuma

diferença significativa na capacidade de diagnóstico dos MLCs e dos parâmetros individuais do OCT. Classificadores de máquinas complexos que usam parâmetros de entrada numerosos tendem a ter um melhor desempenho em bancos de dados maiores (53). Por isso, é possível que o tamanho da nossa amostra seja pequeno para permitir a utilização de todos os parâmetros de SD-OCT.

As comparações entre os diversos estudos já publicados são muito difíceis, devido a diferenças nos dados demográficos da população, na definição e na gravidade do glaucoma, e porque poucos estudos utilizaram MLCs com parâmetros do OCT, e nenhum deles com SD-OCT. A capacidade de diagnóstico dos parâmetros testados neste estudo foi inferior quando comparado com alguns estudos publicados anteriormente, o que pode ser atribuído a várias razões. Primeiro, uma grande proporção (82,3%) dos nossos pacientes tiveram dano glaucomatoso inicial no CV, o que tende a dificultar o diagnóstico do glaucoma. Em segundo lugar, as diferenças de média de idade entre os grupos também podem influenciar a precisão de MLC. Embora a idade não tenha sido incluída na formação dos MLCs utilizados por Bizios et al (57) e Burganski Eliash et al (53), sabe-se que uma redução na espessura da CFNR ocorre com a idade, favorecendo o diagnóstico de glaucoma em pacientes mais velhos.

De fato, o estudo de Huang et al (62), em que os pacientes tinham lesão glaucomatosa precoce ($MD > -6$ dB) e onde não houve diferença significativa de idade entre o grupo glaucoma e o grupo controle ($p=0,07$), mostrou aROCs semelhantes às obtidas em nosso estudo. A espessura do quadrante inferior medido com o Stratus-OCT também foi o melhor parâmetro individual para diferenciar entre olhos normais e glaucomatosos (aROC = 0,832), seguido da espessura global média da CFNR peripapilar (aROC = 0,825). Os autores descreveram aROCs de 0,824 para a análise discriminante linear (LDA), 0,849 para a distância

de Mahalanobis (MD), e 0,821 para MLP. Os autores aplicaram o PCA (*Principal Component Analysis*) na tentativa de melhorar a performance dos classificadores. O PCA é um método matemático que mapeia os atributos do espaço de coordenadas original para outro espaço transformado, onde os dados são representados por combinações algébricas, reduzindo a sua dimensão (142). Os dados reprocessados foram reaplicados para os algoritmos LDA, MD e MLP, gerando uma melhoria no desempenho de todos os classificadores, aumentando as aROCs para 0,915, 0,991 e 0,874, respectivamente. Nesse estudo, os autores conseguiram demonstrar que uma melhora significativa na capacidade diagnóstica pode ser obtida quando empregamos MLCs aos dados do OCT. O classificador que mais se destacou foi o MD, que não utilizamos em nossa análise e, por isso, torna difícil a comparação com os nossos dados, além de não termos utilizado o PCA como técnica de redução de dados redundantes ou indutores de confusão.

Naithani et al (137) comparou a capacidade diagnóstica de MLCs treinados com dados do Stratus-OCT (LDA; *Artificial Neural Network* – ANN; e *Classification and Regression Tree* - CART) com classificadores já presentes no aparelho do HRT II (*Moorfield Regression Analysis* - MRA; *Linear discriminant function* desenvolvido por *R Bathija* – RB *function*; e *Linear discriminant function* desenvolvido *FS Mickelberg* – FSM *function*). Os autores observaram que o parâmetro do OCT que melhor diferenciou olhos normais de glaucomatosos iniciais foi a espessura global média da CFNR peripapilar (aROC = 0,937). Além disso, os classificadores treinados com uma seleção dos dados fornecidos pelo OCT (LDA – 0,982; CART – 0,979; ANN – 0,938) tiveram uma performance melhor do que os fornecidos pelo HRT (FSM *function* – 0,859; RB *function* – 0,842; MRA – 0,767). Nesse estudo, diferentemente do nosso, não foi o objetivo dos autores comparar a capacidade diagnóstica dos

dados fornecidos pelo aparelho com os MLCs, mas sim comparar a aplicabilidade dos MLCs com dados do OCT e com dados do HRT.

Dentre as limitações que podem ser destacadas no nosso estudo temos o tamanho da amostra e a ausência de uma nova população para testar os MLCs. Quanto maior o número de dados disponíveis para treinar os classificadores, mais representativa é a amostra e maior a chance de obtermos um bom aprendizado, principalmente quando a complexidade do classificador é grande. Além disso, apesar de termos utilizado a técnica da validação cruzada, o ideal seria aplicar os MLCs, já treinados, em uma nova população, para testar a sua capacidade em discriminar casos positivos de casos negativos para a doença. Essa nova etapa do estudo já está em andamento no Setor de Glaucoma do Serviço de Oftalmologia do Hospital das Clínicas da UNICAMP.

Na prática clínica, os oftalmologistas especialistas em glaucoma reúnem diversas informações para concluir se há ou não doença glaucomatosa. Dentre essas informações podemos destacar a idade, a pressão intraocular, as características estruturais da CFNR e do NO, os dados funcionais do campo visual, além da raça e história familiar. Dessa forma, estudos têm sido conduzidos com o objetivo de treinar os MLCs com um maior número de dados, associando, assim, informações funcionais e estruturais, visando melhorar sua capacidade diagnóstica.

Bowd et al (59) combinou medidas funcionais, obtidas com a PAP, e estruturais, fornecidas pelo Stratus-OCT, para treinar dois MLCs (*Relevance Vector Machine* – RVM; e *Subspace Mixture of Gaussians* – SSMoG). As aROCs para o RVM treinado com somente dados do OCT, somente dados do CV e com dados do OCT e CV combinados foram,

respectivamente, de 0,809, 0,815 e 0,845. As aROCs para o SSMoG treinado com somente dados do OCT, somente dados do CV e com dados do OCT e CV combinados foram de 0,817, 0,841 e 0,869, respectivamente. Assim, os autores observaram que o uso de técnicas combinadas melhorou significativamente a capacidade diagnóstica dos classificadores quando comparados àqueles treinados somente com dados do OCT, mas não quando comparamos os mesmos classificadores treinados somente com dados do CV.

Seguindo essa mesma linha, recentemente, Silva et al (71), na tentativa de melhorar os resultados obtidos na nossa série, incluíram, além dos parâmetros do OCT, dados do CV (PAP) para treinamento dos mesmos MLCs utilizados em nosso estudo. O melhor desempenho foi obtido com o classificador RAN (aROC = 0,946) treinado com 4 parâmetros do OCT+CV. Esse resultado foi melhor do que o obtido com o melhor parâmetro isolado do OCT (quadrante inferior aROC = 0,813, $p < 0,05$), mas não foi estatisticamente diferente do melhor parâmetro isolado do CV (PSD aROC = 0,915, $p = 0,37$).

CONCLUSÕES

5. CONCLUSÕES

Em conclusão, o presente estudo demonstrou que:

1) O SD-OCT mostrou boa sensibilidade e especificidade para o diagnóstico de glaucoma em um grupo de pacientes com glaucoma inicial a moderado, sendo o melhor parâmetro a espessura da camada de fibras nervosas do quadrante inferior (aROC = 0,813);

2) Os MLCs também mostraram boa sensibilidade e especificidade para o diagnóstico de glaucoma nesta série quando treinados com dados do SD-OCT, sendo o melhor desempenho alcançado com o BAG (aROC = 0,818), dentre os 10 testados;

3) Os MLCs não melhoraram a sensibilidade e especificidade dos parâmetros do SD-OCT para o diagnóstico de glaucoma nesta amostra.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Weinreb RN, Khaw PT. Primary open-angle glaucoma. *Lancet*. 2004; 363(9422):1711–20.
2. Heijl A, Leske MC, Bengtsson B, Hyman L, Bengtsson B, Hussein M. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol*. 2002; 120(10):1268–79.
3. Kass MA, Heuer DK, Higginbotham EJ, Johnson CA, Keltner JL, Miller JP, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002; 120(6):701–713; discussion 829–830.
4. Bengtsson B, Heijl A. A long-term prospective study of risk factors for glaucomatous visual field loss in patients with ocular hypertension. *J Glaucoma*. 2005; 14(2):135–8.
5. Gordon MO, Beiser JA, Brandt JD, Heuer DK, Higginbotham EJ, Johnson CA, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002; 120(6):714–720; discussion 829–830.
6. Medeiros FA, Sample PA, Zangwill LM, Bowd C, Aihara M, Weinreb RN. Corneal thickness as a risk factor for visual field loss in patients with preperimetric glaucomatous optic neuropathy. *Am J Ophthalmol*. 2003; 136(5):805–13.
7. Bonovas S, Peponis V, Filioussi K. Diabetes mellitus as a risk factor for primary open-angle glaucoma: a meta-analysis. *Diabet Med*. 2004; 21(6):609–14.
8. Resnikoff S, Pascolini D, Etya'ale D, Kocur I, Pararajasegaram R, Pokharel GP, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ*. 2004; 82(11):844–51.
9. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006; 90(3):262–7.

10. Rudnicka AR, Mt-Isa S, Owen CG, Cook DG, Ashby D. Variations in primary open-angle glaucoma prevalence by age, gender, and race: a Bayesian meta-analysis. *Invest Ophthalmol Vis Sci.* 2006; 47(10):4254–61.
11. Tielsch JM, Sommer A, Katz J, Royall RM, Quigley HA, Javitt J. Racial variations in the prevalence of primary open-angle glaucoma. The Baltimore Eye Survey. *JAMA.* 1991; 266(3):369–74.
12. Klein BE, Klein R, Sponsel WE, Franke T, Cantor LB, Martone J, et al. Prevalence of glaucoma. The Beaver Dam Eye Study. *Ophthalmology.* 1992; 99(10):1499–504.
13. Quigley HA, West SK, Rodriguez J, Munoz B, Klein R, Snyder R. The prevalence of glaucoma in a population-based study of Hispanic subjects: Proyecto VER. *Arch Ophthalmol.* 2001; 119(12):1819–26.
14. Varma R, Ying-Lai M, Francis BA, Nguyen BB-T, Deneen J, Wilson MR, et al. Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye Study. *Ophthalmology.* 2004; 111(8):1439–48.
15. Murdoch IE, Cousens SN, Babalola OE, Yang YF, Abiose A, Jones BR. Glaucoma prevalence may not be uniformly high in all “black” populations. *Afr J Med Med Sci.* 2001; 30(4):337–9.
16. Ntim-Amponsah CT, Amoaku WMK, Ofosu-Amaah S, Ewusi RK, Idirisuriya-Khair R, Nyatepe-Coo E, et al. Prevalence of glaucoma in an African population. *Eye (Lond).* 2004; 18(5):491–7.
17. Sakata K, Sakata LM, Sakata VM, Santini C, Hopker LM, Bernardes R, et al. Prevalence of glaucoma in a South Brazilian population: Projeto Glaucoma. *Invest Ophthalmol Vis Sci.* 2007; 48(11):4974–9.

18. Quigley HA, Addicks EM, Green WR. Optic nerve damage in human glaucoma. III. Quantitative correlation of nerve fiber loss and visual field defect in glaucoma, ischemic neuropathy, papilledema, and toxic neuropathy. *Arch Ophthalmol*. 1982; 100(1):135–46.
19. Kerrigan-Baumrind LA, Quigley HA, Pease ME, Kerrigan DF, Mitchell RS. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Invest Ophthalmol Vis Sci*. 2000; 41(3):741–8.
20. Harwerth RS, Wheat JL, Fredette MJ, Anderson DR. Linking structure and function in glaucoma. *Prog Retin Eye Res*. 2010; 29(4):249–71.
21. Weinreb RN, Kaufman PL. The glaucoma research community and FDA look to the future: a report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci*. 2009; 50(4):1497–505.
22. Freeman EE, Muñoz B, West SK, Jampel HD, Friedman DS. Glaucoma and quality of life: the Salisbury Eye Evaluation. *Ophthalmology*. 2008; 115(2):233–8.
23. McKean-Cowdin R, Wang Y, Wu J, Azen SP, Varma R, Los Angeles Latino Eye Study Group. Impact of visual field loss on health-related quality of life in glaucoma: the Los Angeles Latino Eye Study. *Ophthalmology*. 2008; 115(6):941–8.
24. Comparison of glaucomatous progression between untreated patients with normal-tension glaucoma and patients with therapeutically reduced intraocular pressures. Collaborative Normal-Tension Glaucoma Study Group. *Am J Ophthalmol*. 1998; 126(4):487–97.
25. The effectiveness of intraocular pressure reduction in the treatment of normal-tension glaucoma. Collaborative Normal-Tension Glaucoma Study Group. *Am J Ophthalmol*. 1998; 126(4):498–505.
26. Leske MC, Heijl A, Hussein M, Bengtsson B, Hyman L, Komaroff E, et al. Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial. *Arch Ophthalmol*. 2003; 121(1):48–56.

27. De Voogd S, Ikram MK, Wolfs RCW, Jansonius NM, Hofman A, de Jong PTVM. Incidence of open-angle glaucoma in a general elderly population: the Rotterdam Study. *Ophthalmology*. 2005; 112(9):1487–93.
28. Mitchell P, Smith W, Attebo K, Healey PR. Prevalence of open-angle glaucoma in Australia. The Blue Mountains Eye Study. *Ophthalmology*. 1996; 103(10):1661-9.
29. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol*. 1989; 107(5):453–64.
30. Sommer A, Katz J, Quigley HA, Miller NR, Robin AL, Richter RC, et al. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol*. 1991; 109(1):77–83.
31. Johnson CA, Cioffi GA, Liebmann JR, Sample PA, Zangwill LM, Weinreb RN. The relationship between structural and functional alterations in glaucoma: a review. *Semin Ophthalmol*. 2000; 15(4):221–33.
32. Harwerth RS, Carter-Dawson L, Shen F, Smith EL 3rd, Crawford ML. Ganglion cell losses underlying visual field defects from experimental glaucoma. *Invest Ophthalmol Vis Sci*. 1999; 40(10):2242–50.
33. Quigley HA, Hohman RM, Addicks EM, Massof RW, Green WR. Morphologic changes in the lamina cribrosa correlated with neural loss in open-angle glaucoma. *Am J Ophthalmol*. 1983; 95(5):673–91.
34. Quigley HA, Dunkelberger GR, Green WR. Chronic human glaucoma causing selectively greater loss of large optic nerve fibers. *Ophthalmology*. 1988; 95(3):357–63.
35. Quigley HA, Katz J, Derick RJ, Gilbert D, Sommer A. An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. *Ophthalmology*. 1992; 99(1):19–28.

36. Kraus H, Bartosová L, Hycl J, Kondrová J, Moravcová Z, Stránská L. The retinal nerve fiber layer in glaucoma. II. The status of the nerve fiber layer and development of changes in the visual field. Prospective study. *Cesk Slov Oftalmol*. 2000; 56(3):149-53.
37. Heijl A, Leske MC, Bengtsson B, Bengtsson B, Hussein M. Measuring visual field progression in the Early Manifest Glaucoma Trial. *Acta Ophthalmol Scand*. 2003; 81(3):286–93.
38. Mohammadi K, Bowd C, Weinreb RN, Medeiros FA, Sample PA, Zangwill LM. Retinal nerve fiber layer thickness measurements with scanning laser polarimetry predict glaucomatous visual field loss. *Am J Ophthalmol*. 2004; 138(4):592–601.
39. Reus NJ, Lemij HG. Diagnostic accuracy of the GDx VCC for glaucoma. *Ophthalmology*. 2004; 111(10):1860–5.
40. Lalezary M, Medeiros FA, Weinreb RN, Bowd C, Sample PA, Tavares IM, et al. Baseline optical coherence tomography predicts the development of glaucomatous change in glaucoma suspects. *Am J Ophthalmol*. 2006; 142(4):576–82.
41. Shah NN, Bowd C, Medeiros FA, Weinreb RN, Sample PA, Hoffmann EM, et al. Combining structural and functional testing for detection of glaucoma. *Ophthalmology*. 2006; 113(9):1593–602.
42. Lemij HG, Reus NJ. New developments in scanning laser polarimetry for glaucoma. *Curr Opin Ophthalmol*. 2008; 19(2):136–40.
43. Da Pozzo S, Marchesan R, Ravalico G. Scanning laser polarimetry - a review. *Clin Experiment Ophthalmol*. 2009; 37(1):68–80.
44. Zangwill L, Bowd C, Medeiros FA. Optic disc imaging. In: Shaaraway TM, Sherwood MB, Hitchings RA et al, eds. *Glaucoma Vol. 1: Medical Diagnosis and Therapy*. Philadelphia: Saunders Elsevier, 2009.

45. Cello KE, Nelson-Quigg JM, Johnson CA. Frequency doubling technology perimetry for detection of glaucomatous visual field loss. *Am J Ophthalmol*. 2000; 129(3):314–22.
46. Brusini P, Busatto P. Frequency doubling perimetry in glaucoma early diagnosis. *Acta Ophthalmol Scand Suppl*. 1998; (227):23–4.
47. Johnson CA, Adams AJ, Casson EJ, Brandt JD. Blue-on-yellow perimetry can predict the development of glaucomatous visual field loss. *Arch Ophthalmol*. 1993; 111(5):645–50.
48. Demirel S, Johnson CA. Short wavelength automated perimetry (SWAP) in ophthalmic practice. *J Am Optom Assoc*. 1996; 67(8):451–6.
49. Hoffmann EM, Lamparter J, Schmidt T, Schulze A. Glaucoma diagnosis and follow-up using the Heidelberg Retina Tomograph. *Ophthalmologe*. 2009; 106(8):687–8, 690–5.
50. Grewal DS, Tanna AP. Diagnosis of glaucoma and detection of glaucoma progression using spectral domain optical coherence tomography. *Curr Opin Ophthalmol*. 2013; 24(2):150-61.
51. Zangwill LM, Bowd C, Berry CC, Williams J, Blumenthal EZ, Sánchez-Galeana CA, et al. Discriminating between normal and glaucomatous eyes using the Heidelberg Retina Tomograph, GDx Nerve Fiber Analyzer, and Optical Coherence Tomograph. *Arch Ophthalmol*. 2001; 119(7):985–93.
52. Bizios D, Heijl A, Bengtsson B. Trained artificial neural network for glaucoma diagnosis using visual field data: a comparison with conventional algorithms. *J Glaucoma*. 2007; 16(1):20–8.
53. Burgansky-Eliash Z, Wollstein G, Chu T, Ramsey JD, Glymour C, Noecker RJ, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Invest Ophthalmol Vis Sci*. 2005; 46(11):4147–52.

54. Boden C, Chan K, Sample PA, Hao J, Lee T-W, Zangwill LM, et al. Assessing visual field clustering schemes using machine learning classifiers in standard perimetry. *Invest Ophthalmol Vis Sci.* 2007; 48(12):5582–90.
55. Tucker A, Vinciotti V, Liu X, Garway-Heath D. A spatio-temporal Bayesian network classifier for understanding visual field deterioration. *Artif Intell Med.* 2005; 34(2):163–77.
56. Bowd C, Medeiros FA, Zhang Z, Zangwill LM, Hao J, Lee T-W, et al. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci.* 2005; 46(4):1322–9.
57. Bizios D, Heijl A, Hougaard JL, Bengtsson B. Machine learning classifiers for glaucoma diagnosis based on classification of retinal nerve fibre layer thickness parameters measured by Stratus OCT. *Acta Ophthalmol.* 2010; 88(1):44–52.
58. Townsend KA, Wollstein G, Danks D, Sung KR, Ishikawa H, Kagemann L, et al. Heidelberg Retina Tomograph 3 machine learning classifiers for glaucoma detection. *Br J Ophthalmol.* 2008; 92(6):814–8.
59. Bowd C, Hao J, Tavares IM, Medeiros FA, Zangwill LM, Lee T-W, et al. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2008; 49(3):945–53.
60. Horn FK, Lämmer R, Mardin CY, Jünemann AG, Michelson G, Lausen B, et al. Combined evaluation of frequency doubling technology perimetry and scanning laser ophthalmoscopy for glaucoma detection using automated classification. *J Glaucoma.* 2012; 21(1):27–34.
61. Mardin CY, Peters A, Horn F, Jünemann AG, Lausen B. Improving glaucoma diagnosis by the combination of perimetry and HRT measurements. *J Glaucoma.* 2006; 15(4):299–305.

62. Huang M-L, Chen H-Y. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2005; 46(11):4121–9.
63. Horn FK, Brenning A, Jünemann AG, Lausen B. Glaucoma detection with frequency doubling perimetry and short-wavelength perimetry. *J Glaucoma.* 2007; 16(4):363–71.
64. Bowd C, Lee I, Goldbaum MH, Balasubramanian M, Medeiros FA, Zangwill LM, et al. Predicting glaucomatous progression in glaucoma suspect eyes using relevance vector machine classifiers for combined structural and functional measurements. *Invest Ophthalmol Vis Sci.* 2012; 53(4):2382-9.
65. Chan K, Lee T-W, Sample PA, Goldbaum MH, Weinreb RN, Sejnowski TJ. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Trans Biomed Eng.* 2002; 49(9):963–74.
66. Sample PA, Goldbaum MH, Chan K, Boden C, Lee T-W, Vasile C, et al. Using machine learning classifiers to identify glaucomatous change earlier in standard visual fields. *Invest Ophthalmol Vis Sci.* 2002; 43(8):2660–5.
67. Goldbaum MH, Sample PA, Chan K, Williams J, Lee T-W, Blumenthal E, et al. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci.* 2002; 43(1):162–9.
68. Bowd C, Chan K, Zangwill LM, Goldbaum MH, Lee T-W, Sejnowski TJ, et al. Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser ophthalmoscopy of the optic disc. *Invest Ophthalmol Vis Sci.* 2002; 43(11):3444-54.
69. Mardin CY, Hothorn T, Peters A, Jünemann AG, Nguyen NX, Lausen B. New glaucoma classification method based on standard Heidelberg Retina Tomograph parameters by bagging classification trees. *J Glaucoma.* 2003; 12(4):340-46.

70. Lauande-Pimentel R, Carvalho RA, Oliveira HC, Gonçalves DC, Silva LM, Costa VP. Discrimination between normal and glaucomatous eyes with visual field and scanning laser polarimetry measurements. *Br J Ophthalmol*. 2001; 85(5):586-91.
71. Silva FR, Vidotti VG, Cremasco F, Dias M, Gomi ES, Costa VP. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry. *Arq Bras Oftalmol*. 2013; 76(3):170-4.
72. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, et al. Optical coherence tomography. *Science*. 1991; 254(5035):1178–81.
73. Schuman JS, Hee MR, Puliafito CA, Wong C, Pedut-Kloizman T, Lin CP, et al. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. *Arch Ophthalmol*. 1995; 113(5):586–96.
74. Schuman JS. Spectral domain optical coherence tomography for glaucoma (an AOS thesis). *Trans Am Ophthalmol Soc*. 2008; 106:426–58.
75. Michelson AA. *Studies in optics*. New York: Dover Publications; 1995.
76. Youngquist RC, Carr S, Davies DE. Optical coherence-domain reflectometry: a new optical evaluation technique. *Opt Lett*. 1987; 12(3):158–60.
77. Xu C, Vinegoni C, Ralston TS, Luo W, Tan W, Boppart SA. Spectroscopic spectral-domain optical coherence microscopy. *Opt Lett*. 2006; 31(8):1079–81.
78. Harwerth RS, Vilupuru AS, Rangaswamy NV, Smith EL 3rd. The relationship between nerve fiber layer and perimetry measurements. *Invest Ophthalmol Vis Sci*. 2007; 48(2):763–73.
79. Guedes V, Schuman JS, Hertzmark E, Wollstein G, Correnti A, Mancini R, et al. Optical coherence tomography measurement of macular and nerve fiber layer thickness in normal and glaucomatous human eyes. *Ophthalmology*. 2003; 110(1):177–89.

80. 1_13733_2.gif (imagem GIF, 600×711 pixels) [Internet]. [citado 14 de outubro de 2013]. Recuperado de: http://www.revophth.com/CMSImagesContent/2008/2/1_13733_2.gif
81. hd_oct.jpg (imagem JPEG, 400×261 pixels) [Internet]. [citado 14 de outubro de 2013]. Recuperado de: http://www.eyeretinasurgeons.com/images/hd_oct.jpg
82. Chang RT, Knight OJ, Feuer WJ, Budenz DL. Sensitivity and specificity of time-domain versus spectral-domain optical coherence tomography in diagnosing early to moderate glaucoma. *Ophthalmology*. 2009; 116(12):2294–9.
83. Jeoung JW, Park KH. Comparison of Cirrus OCT and Stratus OCT on the ability to detect localized retinal nerve fiber layer defects in preperimetric glaucoma. *Invest Ophthalmol Vis Sci*. 2010; 51(2):938–45.
84. Park SB, Sung KR, Kang SY, Kim KR, Kook MS. Comparison of glaucoma diagnostic Capabilities of Cirrus HD and Stratus optical coherence tomography. *Arch Ophthalmol*. 2009; 127(12):1603–9.
85. Moreno-Montañés J, Olmo N, Alvarez A, García N, Zarranz-Ventura J. Cirrus high-definition optical coherence tomography compared with Stratus optical coherence tomography in glaucoma diagnosis. *Invest Ophthalmol Vis Sci*. 2010; 51(1):335–43.
86. Leung CK, Cheung CY, Weinreb RN, Qiu Q, Liu S, Li H, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology*. 2009; 116(7):1257–63.
87. Huang L, Fan N, Shen X, He J. Comparison of the Diagnostic Ability of Retinal Nerve Fiber Layer Thickness Measured using Time Domain and Spectral Domain Optical Coherence Tomography in Primary Open Angle Glaucoma. *Eye Sci*. 2011; 26(3):132-7.

88. Hong S, Seong GJ, Kim SS, Kang SY, Kim CY. Comparison of peripapillary retinal nerve fiber layer thickness measured by spectral vs. time domain optical coherence tomography. *Curr Eye Res.* 2011; 36(2):125–34.
89. Cremasco F, Massa G, Gonçalves Vidotti V, Pedroso de Carvalho Lupinacci Á, Costa VP. Intrasession, intersession, and interexaminer variabilities of retinal nerve fiber layer measurements with spectral-domain OCT. *Eur J Ophthalmol.* 2011; 21(3):264–70.
90. Toscano DA, de Ávila MP, Chalita MRC. Reproducibility of peripapillary retinal nerve fiber layer thickness measurements using spectral domain OCT in Brazilian patients. *Arq Bras Oftalmol.* 2012; 75(5):320–3.
91. Massa GC, Vidotti VG, Cremasco F, Lupinacci APC, Costa VP. Influence of pupil dilation on retinal nerve fibre layer measurements with spectral domain OCT. *Eye (Lond).* 2010; 24(9):1498–502.
92. Kiernan DF, Hariprasad SM. Normative Databases in SD-OCT: A Status Report. *Retinal Physician.* 2010; 7(3): 16-8, 20-6, 28-9.
93. Mitchell TM. *Machine Learning.* 1^o ed. McGraw-Hill Science/Engineering/Math; 1997.
94. Reed R, II RJM. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks.* A Bradford Book; 1999.
95. Rezende SO. *Sistemas Inteligentes – Fundamentos e Aplicações.* São Paulo: Editora Manole; 2003.
96. Han J, Kamber M, Pei J. *Data mining: concepts and techniques.* Burlington, MA: Elsevier; 2012.
97. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition.* 2^o ed. Morgan Kaufmann; 2005.

98. Gewehr JE, Szugat M, Zimmer R. BioWeka—extending the Weka framework for bioinformatics. *Bioinformatics*. 2007; 23(5):651–3.
99. Kotsiantis SB. Supervised Machine Learning: A Review of Classification Techniques. *Informatica*. 2007; 31:249–268. Recuperado de: <http://citeseer.uark.edu:8080/citeseerx/viewdoc/summary?doi=10.1.1.95.9683>
100. Diamantidis NA, Karlis D, Giakoumakis EA. Unsupervised stratification of cross-validation for accuracy estimation. *Artif Intell*. 2000;116(1–2):1–16.
101. Russell SJ, Norvig P, Davis E. *Artificial intelligence: a modern approach*. Upper Saddle River, NJ: Prentice Hall; 2010.
102. Rokach L, Maimon OZ. *Data mining with decision trees theory and applications* [Internet]. Singapore: World Scientific; 2008 [citado 17 de junho de 2013]. Recuperado de: <http://site.ebrary.com/id/10255816>
103. Mitchell TM. *Machine Learning*. New York: McGraw-Hill; 1997.
104. Breiman L. Random Forests. *Machine Learning*. 2001; 45(1):5–32.
105. Verikas A, Gelzinis A, Bacauskiene M. Mining data with random forests: A survey and results of new tests. *Pattern Recognition*. 2011; 44(2):330–49.
106. Liaw A, Wiener M. Classification and Regression by Random Forest. *R News*. 2002; 2(3):18-22.
107. Breiman L. Bagging predictors. *Mach Learn*. 1996; 24(2):123–40.
108. Wang Y-R, Yu C-Y, Chan H-H. Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. *International Journal of Project Management*. 2012; 30(4):470–8.
109. Schapire RE. The Strength of Weak Learnability. *Mach Learn*. 1990; 5(2):197–227.

110. Freund Y. Boosting a weak learning algorithm by majority. *Inf Comput.* 1995; 121(2):256–85.
111. Caruana R, Niculescu-Mizil A, Crew G, Ksikes A. Ensemble selection from libraries of models. *Proceedings of the twenty-first international conference on Machine learning* [Internet]. New York, NY, USA: ACM; 2004 [citado 28 de julho de 2013]. Recuperado de: <http://doi.acm.org/10.1145/1015330.1015432>.
112. Liu K-H, Li B, Zhang J, Du J-X. Ensemble component selection for improving ICA based microarray data prediction models. *Pattern Recognition.* 2009; 42(7):1274–83.
113. Tsoumakas G, Partalas I, Vlahavas I. A taxonomy and short review of ensemble selection. In *ECAI 2008 Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications*, SUEMA. 2008; 41-6.
114. Rosenblatt F. Neurocomputing: foundations of research. In: Anderson JA, Rosenfeld E, organizadores. Cambridge, MA, USA: MIT Press; 1988 [citado 28 de julho de 2013]. p. 89–114. Recuperado de: <http://dl.acm.org/citation.cfm?id=65669.104386>.
115. Rosenblatt F. Principles of neurodynamics: perceptrons and the theory of brain mechanisms. Spartan Books; 1962.
116. Rede Neural Artificial - Perceptron de Múltiplas Camadas. Borges Solut. 2012 [citado 12 de novembro de 2013]. Recuperado de: <http://www.borges-solutions.com/rede-neural-artificial-perceptron-de-multiplas-camadas>.
117. Duda RO, Hart PE, Stork DG. *Pattern Classification*, New York: John Wiley & Sons; 2001.
118. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* 1986; 323(6088):533–6.
119. Haykin S. *Neural Networks: A Comprehensive Foundation*. 2^o ed. Prentice Hall; 1998.
120. Hassoun MH. *Fundamentals of Artificial Neural Networks*. MIT Press; 1995.

121. 0680140209001.png (imagem PNG, 1986×1160 pixels) [Internet]. [citado 12 de novembro de 2013]. Recuperado de: <http://www.emeraldinsight.com/fig/0680140209001.png>
122. On Why Discretization Works for Naive-Bayes Classifiers - Springer [Internet]. [citado 28 de julho de 2013]. Recuperado de: http://link.springer.com/chapter/10.1007%2F978-3-540-24581-0_37
123. Vapnik VN. Statistical learning theory. New York : Wiley; 1998.
124. Vapnik VN. The nature of statistical learning theory. New York: Springer; 2000.
125. Lorena AC, Carvalho ACPLF de. Uma Introdução às Support Vector Machines. Rev Informática Teórica E Apl. 2007; 14(2):43–67.
126. Genton MG. Classes of kernels for machine learning: a statistics perspective. J Mach Learn Res. 2002; 2:299–312.
127. Schölkopf B, Smola AJ. Learning with kernels support vector machines, regularization, optimization, and beyond [Internet]. Cambridge, Mass.: MIT Press; 2002 [citado 29 de julho de 2013]. Recuperado de: <http://site.ebrary.com/id/10229601>
128. Altman DG. Practical Statistics For Medical Research. Chapman and Hall; 1991.
129. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994; 308(6943):1552.
130. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ. 1994; 309(6948):188.
131. Reiser B, Faraggi D. Confidence Intervals for the Generalized ROC Criterion. Biometrics. 1997; 53(2):644–52.
132. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997; 30(7):1145–59.

133. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*. 1993; 39(4):561–77.
134. Messias A, Jorge R, Cruz AAV e. Logarithmic visual acuity charts: reasons to use and how to design it. *Arquivos Brasileiros de Oftalmologia*. 2010; 73(1):96-100.
135. Chylack LT Jr, Wolfe JK, Singer DM, Leske MC, Bullimore MA, Bailey IL, et al. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Arch Ophthalmol*. 1993; 111(6):831–6.
136. Hodapp E, Parrish RK, Anderson DR. *Clinical decisions in glaucoma*. St. Louis, Mo.: Mosby; 1993. pp. 52-61.
137. Naithani P, Sihota R, Sony P, Dada T, Gupta V, Kondal D, et al. Evaluation of optical coherence tomography and heidelberg retinal tomography parameters in detecting early and moderate glaucoma. *Invest Ophthalmol Vis Sci*. 2007; 48(7):3138–45.
138. Adler W, Peters A, Lausen B. Comparison of classifiers applied to confocal scanning laser ophthalmoscopy data. *Methods Inf Med*. 2008; 47(1):38–46.
139. Zangwill LM, Chan K, Bowd C, Hao J, Lee T-W, Weinreb RN, et al. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci*. 2004; 45(9):3144–51.
140. Goldbaum MH. Unsupervised learning with independent component analysis can identify patterns of glaucomatous visual field defects. *Trans Am Ophthalmol Soc*. 2005; 103:270–80.
141. Goldbaum MH, Sample PA, Zhang Z, Chan K, Hao J, Lee T-W, et al. Using unsupervised learning with independent component analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci*. 2005; 46(10):3676-83.
142. Jolliffe I. *Principal Component Analysis*. *Encycl Stat Behav Sci*. John Wiley & Sons, Ltd; 2005.

ANEXOS

7.1. ANEXO 1. Parecer de Aprovação do Comitê de Ética em Pesquisa da Faculdade de Ciências Médicas da Universidade Estadual de Campinas (FCM - UNICAMP)

FACULDADE DE CIÊNCIAS MÉDICAS COMITÊ DE ÉTICA EM PESQUISA

Caixa Postal 6111, 13083-970 Campinas, SP

(0_19) 3788-8936

FAX (0_19) 3788-8925

www.fcm.unicamp.br/pesquisa/etica/index.html

cep@fcm.unicamp.br

CEP, 23/08/05.
(Grupo III)

PARECER PROJETO: Nº 406/2005

CAAE: 0204.0.146.000-05

I-IDENTIFICAÇÃO:

PROJETO: “DIAGNÓSTICO PRECOCE E DE PROGRESSÃO DO GLAUCOMA BASEADOS EM SISTEMA DE APRENDIZAGEM DE CLASSIFICADORES HÍBRIDOS”

PESQUISADOR RESPONSÁVEL: Vital Paulino Costa

INSTITUIÇÃO: Ambulatório de Glaucoma/HC/UNICAMP

APRESENTAÇÃO AO CEP: 01/08/2005

APRESENTAR RELATÓRIO EM: 23/08/06

II - OBJETIVOS

1. Desenvolver sistema de aprendizagem de classificadores híbridos para integrar dados anatômicos, epidemiológicos e funcionais para aumentar a sensibilidade e especificidade do diagnóstico de glaucoma. 2. Desenvolver sistema de aprendizagem de classificadores híbridos que integrará dados anatômicos e funcionais para avaliação longitudinal de pacientes, facilitando o diagnóstico precoce de glaucoma. 3. Desenvolver sistema de aprendizagem de classificadores híbridos que integrará dados anatômicos e funcionais para avaliação longitudinal de pacientes, facilitando o diagnóstico de progressão do glaucoma.

III - SUMÁRIO

Serão recrutados pacientes do ambulatório de oftalmologia da Unicamp com glaucoma, hipertensos oculares e normais para um exame oftalmológico completo. Os resultados dos exames formarão um banco de dados que será utilizado para formação de um sistema computacional híbrido.

IV - COMENTÁRIOS DOS RELATORES

O projeto é relevante já que pretende diagnosticar o glaucoma precocemente, tratá-lo e assim, evitar maiores danos à visão. O Termo de Consentimento Livre e Esclarecido explica completa e pormenorizadamente, a natureza da pesquisa, seus objetivos, métodos, benefícios, potenciais riscos e incômodos eventuais, além de garantir a liberdade do sujeito em participar ou não da pesquisa e assegurar o sigilo dos sujeitos envolvidos.

V - PARECER DO CEP

O Comitê de Ética em Pesquisa da Faculdade de Ciências Médicas da UNICAMP, após acatar os pareceres dos membros-relatores previamente designados para o presente caso e atendendo todos os dispositivos das Resoluções 196/96 e complementares, bem como ter aprovado o Termo do Consentimento Livre e Esclarecido, assim como todos os anexos incluídos na Pesquisa, resolve aprovar sem restrições o Protocolo de Pesquisa supracitado.

O conteúdo e as conclusões aqui apresentados são de responsabilidade exclusiva do CEP/FCM/UNICAMP e não representam a opinião da Universidade Estadual de Campinas nem a comprometem.

VI - INFORMAÇÕES COMPLEMENTARES

O sujeito da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (Res. CNS 196/96 – Item IV.1.f) e deve receber uma cópia do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado (Item IV.2.d).

Pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado e descontinuar o estudo somente após análise das razões da descontinuidade pelo CEP que o aprovou (Res. CNS Item III.1.z), exceto quando perceber risco ou dano não previsto ao sujeito participante ou quando constatar a superioridade do regime oferecido a um dos grupos de pesquisa (Item V.3.).

O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo (Res. CNS Item V.4.). É papel do pesquisador assegurar medidas imediatas adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP e à Agência Nacional de Vigilância Sanitária – ANVISA – junto com seu posicionamento.

Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas. Em caso de projeto do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma junto com o parecer aprovatório do CEP, para serem juntadas ao protocolo inicial (Res. 251/97, Item III.2.e)

Relatórios parciais e final devem ser apresentados ao CEP, de acordo com os prazos estabelecidos na Resolução CNS-MS 196/96.

VII - DATA DA REUNIÃO

Homologado na VIII Reunião Ordinária do CEP/FCM, em 23 de agosto de 2005.


Prof. Dra. Carmen Sílvia Bertuzzo
PRESIDENTE DO COMITÊ DE ÉTICA EM PESQUISA
FCM / UNICAMP

7.2. ANEXO 2. Tabela de conversão da acuidade visual medida em escala de Snellen para escala de LogMAR (134).

Tabela 1. Notações mais usadas para representar acuidade visual

logMAR	Ângulo (minuto de arco)	Decimal	Imperial	Métrico	Frequência espacial (C/°)
1,3	20,0	0,05	20/400	6/120	600
1,2	15,8	0,06	20/317	6/95	475
1,1	12,6	0,08	20/252	6/76	378
1,0	10,0	0,10	20/200	6/60	300
0,9	7,9	0,13	20/159	6/48	238
0,8	6,3	0,16	20/126	6/38	189
0,7	5,0	0,20	20/100	6/30	150
0,6	4,0	0,25	20/80	6/24	119
0,5	3,2	0,32	20/63	6/19	95
0,4	2,5	0,40	20/50	6/15	75
0,3	2,0	0,50	20/40	6/12	60
0,2	1,6	0,63	20/32	6/10	48
0,1	1,3	0,79	20/25	6/8	38
0	1,0	1,00	20/20	6/6	30
-0,1	0,8	1,26	20/16	6/5	24
-0,2	0,6	1,58	20/13	6/4	19
-0,3	0,5	2,00	20/10	6/3	15

APÊNDICES

8.1. APÊNDICE 1. Quadro de dados coletados

Quadro 1. Dados coletados: indivíduos do grupo normal.

Nome	Olho	idade	sexo	raça	AV logMAR	refra ção (EE)	PIO	n med	MD	PSD	G H																	
											T	AT	T	S	N	I	1	2	3	4	5	6	7	8	9	10	11	12
APS	1	55	2	3	0	1,5	21	0	-5,91	5,22	3	110	71	142	83	144	129	117	67	65	117	177	140	68	56	88	150	146
AAP	2	51	1	1	0	3,25	12	0	-1,58	2,02	1	89	61	105	80	110	118	89	68	82	96	94	139	64	48	71	105	92
APS	1	78	2	1	0,2	3,75	17	0	-2,38	1,42	1	84	48	110	76	100	100	93	53	83	106	126	69	43	42	59	95	134
AFS	1	47	1	1	0	-0,5	16	0	-2,95	2,84	2	77	48	109	66	86	105	83	61	55	60	106	91	50	33	61	115	107
ASM	1	45	2	2	0	-0,75	14	0	-0,56	1,66	3	112	71	133	78	165	135	106	63	64	140	203	153	71	59	84	111	153
AG	1	76	1	1	0	2,25	16	0	0,22	1,5	1	86	64	103	81	95	73	93	82	68	66	83	137	74	47	71	128	109
AI	1	73	1	1	0	plano	13	0	-2,68	1,73	1	84	54	106	66	110	99	81	55	64	99	138	92	54	46	61	91	128
AMS	1		2	2	0	plano	12	0	-0,79	1,38	1	77	56	87	62	104	71	85	54	46	84	111	117	59	47	61	93	98
ASA	1	48	1	1	0,05	plano	16	0	-3,26	3,88	1	93	59	118	76	120	100	83	71	75	96	120	144	62	47	69	141	113
AEOF	1	54	2	1	0	plano	10	0	-2,97	2,58	2	80	53	82	74	111	68	68	74	81	102	128	103	46	45	68	117	61
APR	1	52	2	1	0	2,75	16	0	-0,6	1,7	1	88	64	107	69	112	98	84	57	66	79	114	142	66	49	78	129	93
AC	2	55	1	1	0	plano	16	0	-2,5	2,1	2	91	54	120	70	120	125	105	51	55	91	120	148	65	38	58	117	117
CC	1	57	2	1	0	1,25	13	0	-0,95	2	3	110	55	139	1	143	158	128	88	88	122	154	152	58	39	69	112	147
CRS	1	52	2	1	0	-0,75	12	0	-2,04	2,21	1	94	67	11	82	115	95	88	72	86	94	133	119	57	54	89	140	98
CSM	1	49	1	1	0	plano	12	0	-1,08	1,19	1	95	81	110	70	119	101	84	58	68	109	127	121	81	0	62	88	141
CSS	2	64	2	1	0	4,25	11	0	-4,12	1,37	4	94	60	127	66	121	115	76	61	62	102	161	101	56	51	74	123	142
CM	1	59	2	2	0	plano	10	0	-3,05	2,05	2	98	77	132	59	126	124	73	44	58	87	126	164	87	59	86	137	135
DAS	1	49	2	2	0,2	plano	20	0	-3,85	1,84	1	89	55	116	71	113	86	102	58	54	77	136	127	46	45	74	154	109
DABB	1	48	2	1	0	1,5	14	0	-1,95	1,98	1	102	66	134	69	138	108	78	73	55	120	141	153	74	50	73	131	164
DRS	1	54	1	1	0,05	plano	18	0	-2,11	1,55	1	102	69	125	73	141	100	76	61	83	131	172	122	49	62	95	154	121
DAS	1	45	1	2	0,15	0,75	19	0	-1,84	1,67	1	70	39	93	49	100	75	58	41	48	87	122	91	38	48	32	88	116
ERC	1	52	1	2	0,2	-2,75	18	0	-2,93	1,29	1	93	65	126	82	100	110	102	62	81	80	102	119	63	55	79	125	144
EMRB	1	54	2	1	0,15	4,25	16	0	-0,07	1,46	1	98	61	131	79	123	117	98	66	74	100	131	138	60	46	77	139	136
EES	2	51	2	1	0,2	1	12	0	0,02	1,05	1	91	67	110	60	126	102	86	52	41	92	130	155	69	58	74	134	95
GJQS	2	64	2	1	0	5	16	0	-3,5	1,47	1	93	59	121	75	116	118	71	50	56	104	134	109	64	56	106	113	131
GJS	1	59	1	1	0	3,25	16	0	0,45	1,31	1	95	60	112	65	142	87	90	54	52	114	163	149	60	48	72	130	121
HMR	1	69	2	1	0,05	5,5	11	0	-2,91	2,09	1	80	62	101	56	103	81	61	47	60	81	99	127	62	50	74	127	94
HCC	2	63	1	1	0,05	plano	15	0	-1,83	1,68	3	82	55	102	60	112	90	85	51	45	93	130	113	57	44	65	93	124
IC	1	65	2	1	0	3	20	0	-2,6	2,08	1	90	53	117	72	117	102	109	54	53	86	132	132	51	43	64	104	145
IPS	1	54	2	1	0	1,5	18	0	-0,31	1,37	1	102	86	120	68	132	82	81	62	61	108	158	129	79	79	101	152	127
JBSC	2	65	1	1	0	plano	16	0	1,89	1,12	1	92	50	123	74	119	112	84	62	71	89	152	115	45	41	66	119	146
JJV	1	82	1	1	0,5	plano	14	0	-2,73	1,88	1	78	50	105	68	89	100	92	52	60	78	95	95	51	43	58	80	135
JRDF	1	65	1	2	0,05	plano	14	0	-2,45	2,63	1	99	60	118	76	143	106	97	64	66	117	171	141	54	44	82	142	107

JAS	2	46	1	1	0	-1,75	14	0	-0,66	1,61	1	101	71	141	69	125	121	95	53	58	96	115	165	73	56	84	150	153
JBMO	1	57	1	3	0,05	1,25	12	0	-1,19	1,82	1	89	73	104	53	128	98	64	46	48	105	133	147	73	62	83	137	76
JMR	1	62	1	1	0,05	plano	18	0	1,49	1,35	1	110	69	142	3	126	101	139	78	92	101	121	154	88	50	68	153	172
JPS	2	61	1	1	0	0,25	16	0	-2	1,45	1	90	77	98	62	124	121	84	66	80	162	132	78	51	58	77	93	80
JFC	1	48	2	1	0	plano	15	0	-0,46	1,13	1	85	71	102	66	99	86	86	56	55	75	86	137	74	47	93	133	86
MS	1	49	2	1	0	0,5	15	0	0,86	1,06	1	103	73	125	86	127	108	110	74	73	99	138	144	77	56	85	137	130
MAF	2	56	2	2	0,1	1	14	0	-0,22	1,68	1	98	54	122	84	131	130	105	71	75	101	173	120	58	42	62	98	136
MFSA	1	47	2	1	0	plano	12	0	-1,98	1,79	1	98	68	123	67	133	114	82	55	63	108	134	157	83	47	75	112	144
MIMP	1	66	2	1	0	plano	13	0	0,05	1,54	1	89	67	106	74	111	77	80	58	83	85	120	127	72	48	80	132	109
MMG	2	66	2	3	0	1,5	16	0	-0,09	1,63	1	89	84	104	68	99	86	76	64	64	66	109	121	78	59	114	150	76
OCS	1	50	1	1	0	3,75	10	0	0,65	1,2	1	109	64	138	80	154	117	99	73	70	139	168	154	71	51	70	137	158
RS	1	47	1	1	0	-2,75	13	0	0,37	1,2	1	102	66	141	77	125	117	104	58	68	83	136	155	73	44	81	165	140
SS	1	67	1	1	0	plano	16	0	-0,63	1,89	1	107	69	143	79	137	113	93	64	79	111	143	156	77	46	83	155	162
TFM	1	49	2	1	0	0,75	12	0	-3,29	2,49	1	107	61	129	92	145	89	128	78	68	119	153	164	60	46	76	143	154
WA	2	55	1	1	0	plano	19	0	-1,19	3,67	3	95	63	111	76	131	123	101	58	69	126	148	120	60	62	67	97	114
Média		57,02			0,042	1,53	14,8	0	-1,50	1,85		93,5	,1	1	7	121	104	90,7	,2	66,1	99,6	133,3	1	,6	,9	75	3	123,3
Desvio												9,			10,			10						12	10	13,		
Padrão		9,22			0,091	2,15	2,8	0	1,6	0,8		9,9	7	21,5	7	17,1	18,5	16,9	,2	12,5	20,4	25,2	23,7	,1	,7	9	22,1	26,1

Nome: letras iniciais dos nomes dos pacientes, olho: 1=direito 2= esquerdo, idade: em anos, sexo: 1= masculino 2=feminino, raça: 1=branco 2=pardo 3=negro 4=amarelo, AV logMAR: acuidade visual em escala logMAR, EE: equivalente esférico, PIO: pressão intraocular, n med: número de medicações hipotensoras, MD=*mean deviation*, PSD: *pattern standard deviation*, GHT: *glaucoma hemifield test* 1=dentro dos limites de normalidade 2=fora dos limites de normalidade, 3=borderline, AT: *average thickness* do OCT, T: quadrante temporal do OCT, S: quadrante superior do OCT, N: quadrante nasal do OCT, I: quadrante inferior do OCT, 1 a 12: divisão da espessura do OCT nas 12 horas de relógio.

Quadro 2. Dados coletados: indivíduos do grupo glaucoma.

Nome	Idade	Sexo	Raça	AV logMAR	refração			n med	MD	PSD	G H																	
					(EE)	PIO					T	AT	T	S	N	I	1	2	3	4	5	6	7	8	9	10	11	12
APBS	1	64	2	1	0,2	3,25	20	2	-1,68	2,58	1	94	67	112	80	118	75	88	68	85	86	125	143	63	51	87	157	102
ASS	1	48	2	1	0	plano	12	4	-1,52	2,28	1	83	58	102	84	87	66	84	90	79	101	96	65	38	42	94	108	133
AMS	2	68	2	2	0,3	1,75	13	1	-3,38	6,85	3	86	63	110	73	98	108	91	66	63	73	94	129	67	50	73	107	115
ATOS	1	69	2	1	0,3	4	18	1	-2,78	2,81	2	74	55	84	51	105	73	63	47	44	79	134	103	50	49	66	91	88
ALSS	1	59	2	1	0,15	plano	14	1	-4,77	4,47	3	96	83	108	77	116	100	92	70	70	92	119	137	93	70	85	97	128
AANO	2	48	2	2	0,05	1,25	13	3	-4,33	1,81	4	63	42	77	53	82	76	62	50	48	72	89	85	46	37	42	76	77
ACC	1	60	1	1	0	0,25	14	2	-3,17	2,97	2	68	53	99	51	70	87	57	45	51	67	75	68	51	48	59	96	114
AV	2	57	1	1	0,1	plano	13	2	-12	11,04	3	75	48	116	59	75	135	75	53	49	108	69	48	46	41	57	72	139
ASF	1	62	2	1	0,3	-3,25	12	4	-3,8	2,27	2	78	51	95	73	95	75	83	65	70	79	116	90	49	47	58	85	125
AG	2	78	1	1	0,3	plano	11	4	-6,84	5,81	3	76	74	59	75	94	73	79	80	65	71	110	102	69	65	90	60	45
AC	2	73	1	1	0,15	3	12	4	-12	12,58	3	72	50	95	70	72	98	85	72	54	74	63	80	44	44	64	99	89
BIBR	1	71	2	1	0,3	1,25	16	2	-1,92	3,5	3	80	67	78	78	97	78	66	74	93	88	104	98	53	62	84	106	51
ESS	1	69	2	1	0	1	11	2	0,15	4,41	1	89	63	110	68	117	93	93	53	58	101	125	124	78	47	64	103	134
ECS	1	60	1	2	0	plano	14	1	-0,69	2,67	2	71	40	82	70	91	75	79	66	64	117	89	67	35	39	47	99	72
ES	1	60	1	2	0,2	-1,5	14	2	-3,28	2,27	3	71	37	88	65	96	74	67	55	72	86	120	81	35	28	49	92	96
EM	1	48	1	1	0	plano	21	2	-5,68	7,81	3	86	71	111	62	102	110	77	61	48	74	93	138	80	61	72	92	129
ESF	1	67	2	1	0,3	-0,5	12	3	-1,45	3,09	2	71	47	83	59	94	80	62	63	52	81	114	86	51	38	53	87	83
FCS	1	60	1	3	0,2	plano	12	4	-6,17	6,37	3	81	55	91	69	107	74	73	61	75	97	120	103	54	47	64	106	93
GB	2	52	1	1	0	0,5	16	0	-2,07	2,87	3	104	85	138	60	135	153	67	55	58	87	144	174	107	64	83	112	148
GCS	2	70	1	2	0,15	1	15	3	-3,24	3,63	3	64	48	91	53	65	79	71	39	50	79	76	41	43	42	58	100	95
GM	1	69	1	1	0,2	1	14	2	-5,89	3,99	2	92	65	111	72	122	101	75	64	76	114	148	103	61	54	79	109	124
ISA	2	70	1	2	0,05	1,5	10	2	-1,94	3,15	3	72	42	83	69	96	63	3	50	53	82	102	104	39	37	49	76	110
JMT	1	64	2	1	0,2	plano	14	1	-9,29	8,47	3	100	82	130	70	118	108	83	54	73	99	97	157	94	58	95	154	129
JBL	1	57	1	2	0	plano	13	3	-2,62	3,8	3	74	40	85	71	101	97	97	58	57	82	131	90	45	34	42	60	96
JCPS	2	58	1	2	0	0,75	12	2	-1,33	1,9	3	76	38	107	59	101	100	68	51	57	98	122	82	43	31	40	81	140
JRDF	2	66	1	3	0,05	plano	14	1	-5,75	3,89	3	87	53	121	66	108	112	86	55	58	102	120	101	62	45	51	144	107
JTO	1	53	1	1	0	1	13	4	-4,39	4,44	3	71	54	84	57	88	71	68	49	53	71	100	95	47	50	67	105	78
JASF	2	63	1	1	0,1	0,75	14	3	-2,67	3,49	3	82	52	94	65	117	93	93	50	52	105	130	115	50	49	58	113	76
JAR	2	51	1	3	0,15	-4,75	10	3	-3,42	2,76	2	78	46	103	69	95	91	78	67	61	67	128	90	37	40	62	126	91
JSN	1	67	1	1	0	2,25	13	4	-2,57	2,43	3	67	43	77	56	90	74	69	45	55	82	114	74	49	44	35	77	81
JJA	2	67	1	1	0	1,5	12	4	-6,43	4,95	3	74	49	104	64	79	96	59	64	69	68	54	115	68	30	48	90	126
JRF	1	57	1	1	0,3	-2,5	13	3	-2,8	1,78	3	69	48	89	63	77	82	66	49	73	100	89	41	45	44	54	90	94
JGS	1	56	2	1	0,2	0,75	16	2	-1,94	4,75	3	73	36	103	69	84	84	83	55	68	109	125	18	45	2	62	113	113
JDS	1	53	2	1	0	1,5	16	1	-3,68	6,16	3	95	62	128	76	115	118	92	63	73	107	109	130	59	50	77	117	148
LP	2	54	2	1	0,2	plano	14	1	-3,62	3,69	2	92	63	112	81	113	93	99	64	79	93	117	131	64	49	75	123	118
LES	1	45	2	1	0	-1,25	18	2	-3,8	3,1	3	83	89	96	66	82	94	87	54	57	71	78	97	91	112	65	94	100
ML	1	64	1	1	0,15	2	14	3	-6,31	5,66	3	73	38	90	66	98	67	66	66	68	70	128	97	33	30	50	108	95
MSC	1	78	2	1	0,3	4,25	14	3	-7,07	5,79	3	77	46	108	81	75	100	3	65	76	96	80	48	45	38	54	100	123

MNOS	1	52	2	1	0,05	2,75	18	1	-2,81	2,36	2	79	50	92	65	108	81	86	55	55	92	141	92	49	35	65	123	72
MLSC	1	46	2	3	0	plano	13	4	-3,62	1,97	3	63	54	69	48	80	68	52	48	45	75	72	92	61	42	60	65	73
MJOR	1	67	2	1	0	4	12	1	-2,73	2,33	2	78	69	94	65	84	70	76	58	60	81	71	101	79	58	70	106	105
MJRT	2	59	2	1	0,05	1,25	10	1	-4,38	4,39	3	85	52	122	62	104	105	72	53	61	97	122	92	53	38	65	151	108
MNSR	2	52	2	1	0	2,25	11	3	-2,22	2,77	1	75	41	115	53	92	106	67	48	43	80	114	83	34	38	52	104	134
MF	2	43	2	1	0,3	plano	12	1	-4,59	4,13	3	90	62	114	63	123	106	82	50	58	104	126	138	79	44	62	100	137
MPS	1	50	2	1	0	2	16	2	-4,45	7,27	3	75	62	85	71	82	75	84	69	60	61	77	107	58	51	76	112	67
NLM	1	78	2	1	0,15	plano	12	2	-5,87	5,46	3	86	75	105	58	104	77	73	56	46	84	120	108	77	63	84	122	117
NCM	1	50	2	1	0	0,75	18	3	-2,71	1,96	3	86	56	104	69	114	91	89	62	56	91	128	125	66	44	58	87	133
NSS	2	58	2	1	0,05	plano	14	4	-6,08	7,39	3	86	56	100	62	127	78	68	59	57	104	140	138	58	47	63	95	127
NTSS	2	50	2	1	0,05	1,25	12	1	-1,16	1,47	1	89	68	121	52	114	116	56	59	40	78	121	145	78	51	75	122	125
ON	1	73	1	1	0	3,5	18	1	-7,4	11,75	3	71	50	87	55	93	91	68	40	56	69	81	130	51	33	66	76	95
OB	1	55	1	1	0,05	plano	16	1	-4,13	3,47	2	86	53	116	70	104	122	88	48	76	116	127	68	49	46	65	108	118
PF	2	71	1	1	0,1	3,5	12	0	-3,19	3,97	3	74	51	90	59	96	85	72	45	59	70	98	121	59	37	58	99	84
PDL	1	53	1	3	0	plano	16	3	-2,16	3,13	3	53	43	74	34	60	85	46	45	11	35	74	71	43	35	51	54	84
RFG	1	46	2	1	0	-0,25	16	2	-4,4	3,46	3	79	63	86	72	94	69	79	65	73	101	110	73	62	54	74	83	105
RAF	1	58	1	1	0,05	1,5	13	3	-3,43	3,73	3	75	58	87	63	92	75	69	57	62	85	91	100	70	48	56	92	95
SCL	1	61	1	1	0	4	13	1	-2,39	2,18	3	86	53	107	71	114	90	85	65	62	100	128	114	61	42	56	106	126
SO	1	75	1	4	0,1	0,25	10	2	-2,66	2,07	2	105	76	128	68	149	96	78	61	66	126	180	143	77	68	83	144	143
SAS	1	54	1	2	0,05	0,5	12	1	-5,53	5,96	3	99	79	128	71	118	94	67	68	79	84	112	160	81	61	95	186	104
SABT	1	50	2	1	0,2	2	19	0	-4,11	6,85	3	97	71	139	72	104	116	93	63	60	59	115	138	63	53	96	143	159
TMO	2	59	2	3	0,1	2,75	10	3	-4,77	7,07	3	95	43	126	94	115	144	1	90	82	113	139	92	35	38	56	104	130
VDG	1	66	1	1	0,1	-0,75	12	1	-6,11	2,88	3	78	57	97	62	94	90	72	61	51	88	115	79	62	42	67	82	120
ZPF	2	54	2	1	0,05	1,5	14	2	-3,35	2,82	2	92	63	115	70	121	110	90	54	65	98	136	129	58	49	81	131	104

Média	59,9				0,1	1,2	13,8	2,2	-4,1	4,3		80,6	,6	100,9	,8	99,4	91,4	,6	,6	61,1	87,4	109,4	101,4	57,9	46,5	,2	103,5	107,6
Desvio												12	9,					13	10							14		
Padrão	9,0				0,1	1,9	2,5	1,2	2,4	2,4		10,8	,9	17,4	8	17,5	19,1	,3	,4	13,0	16,8	24,4	31,5	16,5	14,0	,7	24,9	24,8

Nome: letras iniciais dos nomes dos pacientes, olho: 1=direito 2=esquerdo, idade: em anos, sexo: 1=masculino 2=feminino, raça: 1=branco 2=pardo 3=negro 4=amarelo, AV logMAR: acuidade visual em escala logMAR, EE: equivalente esférico, PIO: pressão intraocular, n med: número de medicações hipotensoras, MD=*mean deviation*, PSD: *pattern standard deviation*, GHT: *glaucoma hemifield test* 1=dentro dos limites de normalidade 2=fora dos limites de normalidade, 3=borderline, AT: *average thickness* do OCT, T: quadrante temporal do OCT, S: quadrante superior do OCT, N: quadrante nasal do OCT, I: quadrante inferior do OCT, 1 a 12: divisão da espessura do OCT nas 12 horas de relógio.

8.2. APÊNDICE 2. Termo de Consentimento Livre e Esclarecido

FORMULÁRIO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título do Projeto: Diagnóstico Precoce e de Progressão do Glaucoma Baseados em Sistema de Aprendizagem de Classificadores Híbridos

Investigador principal: Prof. Dr. Vital Paulino Costa

OBJETIVO DA PESQUISA

Eu entendo que fui convidado(a) a participar de um projeto de pesquisa envolvendo indivíduos normais, pacientes com glaucoma e hipertensos oculares. O objetivo geral do estudo é coletar dados provenientes de exames para gerar um sistema de computador que auxiliará o diagnóstico de glaucoma e sua progressão. Esse estudo poderá levantar dados que possibilitarão diagnosticar glaucoma mais cedo e, assim, tratá-lo antes que cause problemas para a visão. O sigilo será mantido em todo o estudo através da utilização de um código para a identificação dos indivíduos participantes.

PROCEDIMENTO

Entendo que se eu concordar em participar desse estudo, os pesquisadores participantes farão perguntas a respeito dos meus antecedentes médicos e familiares. Eu serei submetido a um exame oftalmológico para confirmar meu estado clínico. O exame oftalmológico tem o objetivo de observar se há alterações no olho que lembrem aquelas que vemos no glaucoma. Além disso, minha pressão ocular e espessura da córnea serão medidas, o que pode causar um leve ardor ocular ou embaçamento transitório da visão, exame de campo visual e fotografias do fundo do olho também serão realizados. As fotografias serão realizadas em dois tipos diferentes de aparelhos fotográficos: retinógrafo e OCT. Estes procedimentos são normalmente realizados em qualquer exame oftalmológico, são seguros e não apresentam riscos à minha vista. Os procedimentos mencionados acima serão realizados a cada seis meses ao longo de quatro anos.

RISCO E DESCONFORTO

Os riscos associados aos procedimentos a serem realizados são mínimos, podendo ocorrer leve ardor ao pingar os colírios anestésico (para medida da pressão do olho) e midriático (para examinar e fotografar o fundo do olho). Esse último deixa a visão embaçada, especialmente para perto, e incômodo com a luz por, aproximadamente, seis horas. O desconforto será mínimo, pois se tratam de exames rotineiros realizados frequentemente e por profissional treinado para realizar todos esses procedimentos.

VANTAGENS

Entendo que não obterei nenhuma vantagem direta com a participação neste estudo e que o diagnóstico e o tratamento provavelmente não serão modificados com base neste estudo. Os resultados dos testes realizados estarão disponíveis através do acompanhamento no ambulatório de origem. No entanto, é possível que este estudo possibilite diagnosticar glaucoma mais cedo em muitos pacientes, e assim, tratá-lo antes que cause problemas para a visão.

SIGILO

Entendo que toda a informação médica, assim como os resultados dos testes decorrentes desse projeto de pesquisa, farão parte do meu prontuário médico e serão matidos em sigilo. Se os resultados ou informações fornecidas forem utilizados para fins de publicação científica, nenhum nome será utilizado.

FORNECIMENTO DE INFORMAÇÃO ADICIONAL

Entendo que posso requisitar informações adicionais relativas ao estudo a qualquer momento. Os Drs Vital (19-35217818) e Álvaro (19-35217769) estarão disponíveis para responder minhas questões e preocupações. Em caso de recurso, dúvidas ou reclamações, contactar a secretaria da comissão de ética da Faculdade de Ciências Médicas – UNICAMP, tel 19-35217232.

RECUSA OU DESCONTINUAÇÃO DA PARTICIPAÇÃO

Entendo que a participação é voluntária e que posso me recusar a participar ou retirar meu consentimento e interromper a minha participação no estudo a qualquer momento sem comprometer os cuidados médicos que recebo atualmente ou receberei no futuro nas instituições participantes. Eu reconheço também que o Dr. Vital Paulino Costa pode interromper a minha participação nesse estudo a qualquer momento que julgar apropriado.

FORMULÁRIO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título do Projeto: Diagnóstico Precoce e de Progressão do Glaucoma Baseados em Sistema de Aprendizagem de Classificadores Híbridos

Investigador principal: Prof. Dr. Vital Paulino Costa

Eu confirmo que o(a) Dr.(a) _____ me explicou o objetivo do estudo, os procedimentos que serão realizados, os riscos, desconfortos e possíveis vantagens advindas desse projeto de pesquisa. Eu li e/ou me foi explicado, assim como compreendi esse formulário de consentimento e estou de pleno acordo em participar desse estudo.

Nome do participante ou responsável, idade, RG e Registro Hospitalar

Assinatura do participante ou responsável

data

Endereço e telefone

RESPONSABILIDADE DO PESQUISADOR

Eu expliquei a _____ o objetivo do estudo, os procedimentos requeridos e os possíveis riscos e vantagens que poderão advir do estudo, usando o melhor do meu conhecimento. Eu me comprometo a fornecer uma cópia desse formulário de consentimento ao participante ou responsável.

Nome do pesquisador ou associado

Assinatura do participante ou responsável

data

Telefones para contato: UNICAMP (19) 3521-8360/(19) 3521-8936

8.3. APÊNDICE 3. Artigo publicado

ORIGINAL ARTICLE

Sensitivity and specificity of machine learning classifiers and spectral domain OCT for the diagnosis of glaucoma

Vanessa G. Vidotti¹, Vital P. Costa¹, Fabrício R. Silva¹, Graziela M. Resende¹, Fernanda Cremasco¹, Marcelo Dias², Edson S. Gomi²

¹Glaucoma Service, Department of Ophthalmology, University of Campinas, Campinas - Brazil

²Department of Computer Engineering School of Engineering, University of São Paulo, São Paulo - Brazil

PURPOSE. To investigate the sensitivity and specificity of machine learning classifiers (MLC) and spectral domain optical coherence tomography (SD-OCT) for the diagnosis of glaucoma.

METHODS. Sixty-two patients with early to moderate glaucomatous visual field damage and 48 healthy individuals were included. All subjects underwent a complete ophthalmologic examination, achromatic standard automated perimetry, and RNFL imaging with SD-OCT (Cirrus HD-OCT; Carl Zeiss Meditec, Inc., Dublin, California, USA). Receiver operating characteristic (ROC) curves were obtained for all SD-OCT parameters. Subsequently, the following MLCs were tested: Classification Tree (CTREE), Random Forest (RAN), Bagging (BAG), AdaBoost M1 (ADA), Ensemble Selection (ENS), Multilayer Perceptron (MLP), Radial Basis Function (RBF), Naive-Bayes (NB), and Support Vector Machine (SVM). Areas under the ROC curves (aROCs) obtained for each parameter and each MLC were compared.

RESULTS. The mean age was 57.0±9.2 years for healthy individuals and 59.9±9.0 years for glaucoma patients ($p=0.103$). Mean deviation values were $-4.1±2.4$ dB for glaucoma patients and $-1.5±1.6$ dB for healthy individuals ($p<0.001$). The SD-OCT parameters with the greater aROCs were inferior quadrant (0.813), average thickness (0.807), 7 o'clock position (0.765), and 6 o'clock position (0.754). The aROCs from classifiers varied from 0.785 (ADA) to 0.818 (BAG). The aROC obtained with BAG was not significantly different from the aROC obtained with the best single SD-OCT parameter ($p=0.93$).

CONCLUSIONS. The SD-OCT showed good diagnostic accuracy in a group of patients with early glaucoma. In this series, MLCs did not improve the sensitivity and specificity of SD-OCT for the diagnosis of glaucoma.

KEY WORDS. Machine learning classifiers, Optical coherence tomography, Sensitivity, Specificity, Spectral domain

Accepted: June 4, 2012

INTRODUCTION

Glaucoma is a progressive optic neuropathy characterized by structural changes to the optic nerve head (ONH) and retinal nerve fiber layer (RNFL), with corresponding functional changes, particularly visual field (VF) loss. Histological studies have demonstrated that 30%-50% of the

retinal ganglion cells may be lost before VF defects appear on standard automated perimetry (SAP) (1). Reports have also shown that subtle structural ONH and peripapillary RNFL defects can precede the development of detectable SAP VF loss in early glaucomatous optic neuropathy (1, 2). Therefore, several techniques have been introduced over the past years aiming at detecting morphologic and func-

tional glaucomatous abnormalities earlier than conventional methods.

Optical coherence tomography (OCT), first described in 1991 by Huang et al (3), is a noncontact, high-resolution technique using a scanning interferometer to produce cross-sectional images of the retina and peripapillary RNFL in vivo (3, 4). Previous reports have shown that time-domain (TD) OCT (Stratus, Carl Zeiss Meditec, Inc., Dublin, California, USA) has high sensitivity and specificity for diagnosing glaucoma, and has good correlation with VF findings detected with SAP (5-8).

Recently, several companies have developed newer versions of OCT employing spectral domain (SD) technology. The SD-OCTs have higher axial resolution and scan speed than conventional TD-OCTs. The Cirrus HD-OCT (Carl Zeiss Meditec, Inc.) has an axial resolution of 5 μm and a scan speed of 27,000 A-scans per second, whereas the Stratus OCT has an axial resolution of 8-10 μm and a scan speed of 400 A-scans per second. The higher sampling rates of the newer OCTs allow more data to be collected and shorter scan times. Studies comparing these 2 technologies have demonstrated that the sensitivity and specificity of various RNFL parameters using the Cirrus OCT are excellent and are equivalent to Stratus OCT (9-13).

Since 1990 (Goldbaum MH, et al. IOVS 1990;31:ARVO Abstract 503), machine learning classifier (MLC) techniques have been applied to optical imaging and visual function measurements to improve glaucoma detection, with results suggesting that these techniques are as good as (14, 15, 21, 25) or better than (16-20, 22-24) currently available methods at classifying eyes as glaucomatous or healthy. Classifiers usually employ a form of supervised learning, where the program learns from positive and negative training examples, representing cases where, for example, there are signs of glaucoma on data obtained by examination of the visual field (positive examples) or not (negative examples). The training is repeated several times with the provision of various training data, with the positive or negative classification previously performed by an ophthalmologist, until the concept (identification of signs of glaucoma in one test) can be properly learned by the system.

To our knowledge, there is no study in the literature evaluating the use of machine learning classifiers with SD-OCT for the diagnosis of glaucoma. The purpose of this study was to investigate the sensitivity and specificity of MLCs and SD-OCT for the diagnosis of glaucoma.

MATERIALS AND METHODS

Subjects

This prospective observational cross-sectional study included 110 eyes of 110 study participants (62 patients with glaucoma and 48 healthy control subjects) older than 40 years and examined at the Glaucoma Service of the University of Campinas, Brazil, between August 2008 and November 2010. Each study participant underwent a comprehensive ophthalmic evaluation, including review of medical history, best-corrected visual acuity, slit-lamp biomicroscopy, intraocular pressure (IOP) measurement with Goldmann applanation tonometry, gonioscopy, dilated slit-lamp fundus examination with a 78-D lens, SAP using the 24-2 SITA Standard (Humphrey Field Analyzer II, Carl Zeiss Meditec, Inc.), and RNFL imaging with Cirrus HD-OCT (Cirrus HD-OCT; Carl Zeiss Meditec, Inc.). To be included in the study, participants had to have a best-corrected visual acuity better than or equal to 20/40, spherical refraction within ± 5.0 D, cylinder correction within ± 3.0 D, open angles on gonioscopy, reliable SAP (false-positive errors <20%, false-negative errors <20%, and fixation losses <20%), absence of coexisting retinal disease, uveitis, nonglaucomatous optic neuropathy, pseudophakia, or aphakia, significant cataract according to the criteria of Lens Opacification Classification System III (defined as the maximum nuclear opacity [NC3, NO3], cortical [C3], and subcapsular [P3]), and family history of glaucoma for healthy subjects. Patients not meeting all criteria were not included. One eye was randomly selected if both eyes were found to be eligible.

The inclusion criteria for healthy individuals were IOP ≤ 21 mmHg, no history of elevated IOP, and 2 consecutive and reliable normal visual fields. Glaucomatous eyes were defined as those with 2 or more IOP measurements > 21 mmHg, glaucomatous changes in the optic nerve head and/or the RNFL, and 2 reliable VF examinations showing glaucomatous VF defects, with the most recent examination within 3 months of the enrollment date. Eyes with glaucomatous VF defects were defined as those that met 2 of the following criteria: 1) cluster of 3 points with a probability of $< 5\%$ on a pattern deviation map in a single hemifield including ≥ 1 point with a probability of $< 1\%$; 2) glaucoma hemifield test outside 99% of the age-specific normal limits; and 3) pattern standard deviation outside 95% of the normal limit. The severity of glaucomatous VF damage

was determined according to the following criteria: 1) early damage: mean deviation (MD) ≥ -6 dB; 2) moderate damage: MD between -6 dB and -15 dB; 3) advanced damage: MD ≤ -15 dB. Only patients with early or moderate damage were included in the study.

Informed consent was obtained from all participants before enrollment. All procedures conformed to the Declaration of Helsinki and the study was approved by the University of Campinas Medical Institutional Review Board.

Optical coherence tomography

Participants underwent ocular imaging with the commercially available Cirrus HD-OCT (software version 3.0, Carl Zeiss Meditec, Inc.), which uses spectral domain technology. The optic disc cube mode consists of 200 A-scans that are derived from 200 B-scans and covers a 6-mm² area centered on the optic disc. After creating a RNFL thickness map from the 3-dimensional cube dataset, the software automatically determines the center of the disc and extracts a circumpapillary circle (1.73-mm radius) for RNFL thickness measurement. All images were acquired with undilated pupils by a single, well-trained ophthalmologist (V.G.V.), who was masked to the diagnosis. The OCT technology provides RNFL thickness maps with 17 parameters: average thickness, 4 quadrants (superior, inferior, nasal, and temporal), and 12 clock hour measurements. All OCT data were aligned according to the orientation of the right eye. Hence, clock hour 9 of the circumpapillary scan represented the temporal side of the optic disc for both eyes. Only well-centered scans with a signal strength ≥ 7 were included.

Machine learning classifiers

Based on patient data obtained from the Cirrus OCT, machine learning classifiers were developed using the following algorithms: Classification Tree (CTREE), Random Forest (RAN), Bagging (BAG), AdaBoost M1 (ADA), Ensemble Selection (ENS), Multilayer Perceptron (MLP), Radial Basis Function (RBF), Naive-Bayes (NB), Linear Support Vector Machine (SVMl), and Gaussian Support Vector Machine (SVMg).

The CTREE is a classifier based on a disjunction (logical or operation) of conjunctions (logical and operation) of constraints on the attribute values of instances (26-28). This logical expression is usually represented as a tree struc-

ture, where each path from the root to a leaf represents a conjunction of attribute values tests. The tree may have many paths, and the tree itself represents the disjunction of all its paths. Another way to see a decision tree is to read it as a set of if-then rules (26). The classic algorithm for constructing decision trees is the ID3 (28). The ID3 is a recursive algorithm that creates the tree from the root to the leaves. It uses information entropy as a heuristic rule to choose the best attribute at each training step. Classification trees are robust to errors or missing attribute values in the training data.

The RAN is a classifier that works based on voting from a set of decision trees. Each tree is constructed using a random selection of features or input variables. After a large number of trees is generated, a decision is obtained from each tree, and the final decision corresponds to the most voted decision. Random Forests are useful in problems like medical diagnosis and document analysis, which have many input variables (29).

The BAG is a method that uses voting as decision criteria to predict the class of a given object. For example, if the classifier is a decision tree, a set of decision trees is generated using different random samples from a dataset. These random samples, taken with replacement from the training dataset, are named bootstrap samples. In the next step, each decision tree predicts the object's class and the final result is the most voted class. This procedure is called "bootstrap aggregating," which is the origin of the acronym bagging (30).

While bagging is based on a voting process to classify a given object by using classifiers trained from bootstrap samples, boosting considers a weighted combination of results from several classifiers. The ADA (Adaptive Boosting) algorithm is an interactive training method that uses a weighted sample set and assigns higher weights for misclassified examples in the next iteration (Freund Y, Schapire RE. Experiments with a new boosting algorithm. In: Proceedings of the 13th International Conference on Machine Learning 1996; 148-56).

The ENS is a classifier constructed from a weighted average or voting from the results of a set of classifiers. In each step of basic set construction procedure, the classifier that maximizes the ensemble's performance with respect to an error metric is added to the set. Possible performance metrics are accuracy, root mean square error, and receiver operating characteristic (ROC) area. The process is repeated for a fixed number of iterations or until all possible

classifiers have been used (Caruana R, Niculescu-Mizil A, Crew G, Ksikes A. Ensemble selection from libraries of models. In: Proceedings of the 21st International Conference on Machine Learning 2004; 137-44).

In MLP, a perceptron is a classifier that takes a vector of attribute's values of an object, calculates a weighted linear combination of these values, and outputs 1 if the result is greater than some threshold and -1 otherwise. Although a single perceptron can represent many boolean functions, MLP are used to express more complex functions. An MLP has multiple layers of interconnected single units, forming a feedforward multilayer network of perceptrons. The internal layers are called hidden layers. In order to be able to represent highly nonlinear functions, the simple threshold criteria are changed to the sigmoid function. The back-propagation algorithm can be used to learn the connection weights for a given fixed multilayer network. It uses gradient descent to minimize the squared error between the network output and the target values for the corresponding inputs, as given in the training set (26, 31).

The RBF is a special type of feed-forward multilayer network with 3 layers. The hidden layer uses as threshold criteria a bell-shaped Gaussian function instead of sigmoid. The RBF represents a hypersphere or hyperellipsoid in the classification space, while the MLP represents a hyperplane (31).

In the Bayesian approach, a given object instance is assigned to the most probable class, given the attribute values that describe the instance. The challenge in this approach is to estimate the conditional probabilities. The NB classifier uses a simplifying assumption that the attribute values are conditionally independent given the target value. The simplified formula is just the product of the probabilities for the individual attributes. An important difference between the NB method and other classifiers is that there is no explicit search through the space of possible models for the classifier (26) (John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the 11th Conf. Uncertainty in Artificial Intelligence 1995; 338-45).

The SVMs are classifiers that use as models linear functions or hyperplanes in a high-dimensional feature space. A SVM is built from a set of vectors whose inner-product with hyperplanes that separates the positive from the negative instances is constant (32, 33).

The SVMg is a classifier that uses a Gaussian radial basis function instead of linear functions as classification sepa-

ration surface for the positive and negative instances (32, 33).

Initially, MLC training sessions were supervised with all 17 parameters of the SD-OCT. Subsequently, a backward feature selection was used to find the minimal number of parameters that resulted in the highest area under the ROC curves (aROC) for each classifier. The analysis started with the evaluation of the classifier's performance over the full-dimensional feature set containing the 17 SD-OCT attributes. Sequentially, the attribute that presented the lowest aROC, computed over the SD-OCT data alone, was removed and the classifier's accuracy was computed. This process of dimension reduction, based on areas under ROC curves of the SD-OCT parameters sorted in descending order, was performed 16 times. It started with the exclusion of the parameter with the lowest aROC and stopped only when the parameter with the best aROC was used. The criterion for evaluating the algorithms involved the analysis of ROC curves generated from the results of several classification tasks for each classifier.

The classifiers were developed using a machine learning environment software called Weka (34) version 3.7.0 (Waikato Environment for Knowledge Analysis, The University of Waikato, New Zealand). The examination data and images were collected at the Department of Ophthalmology, University of Campinas, and temporarily stored in a local workstation. The data were transferred to the server located in the Laboratory of Knowledge Engineering (KNOMA), in the Department of Computer Engineering of School of Engineering of University of São Paulo.

To maximize the use of our collected data, the 10-fold cross-validation resampling method was employed. Accordingly, all data (110 subjects enrolled in the study) were randomly divided into 10 subsets, each containing approximately the same proportion of healthy and glaucomatous OCT thickness measurements. Nine subsets were used as training data, while the remaining subset was used as testing data. This process was performed 10 times, until each of the 10 folders had been used as a test set. The test results were averaged to estimate the classifier's accuracy.

Statistical analysis

All analyses were performed using MedCalc software version 11.0.1 (MedCalc Software, Mariakerke, Belgium). Continuous variables were compared using the Student *t* test, whereas categorical variables were analyzed using

the chi-square or the Fisher exact test.

The ROC curves were obtained for all SD-OCT parameters: average, 4 quadrants (superior, inferior, nasal, temporal), and 12 clock-hours RNFL thickness measurements. The aROCs obtained for each SD-OCT parameter and each machine learning classifier, before and after optimization, were compared using the z test.

Since multiple comparisons were made, $p < 0.01$ was considered to be statistically significant.

RESULTS

A total of 110 eyes of 110 individuals were included in this study. Among the 110 individuals, 62 patients had glaucoma and 48 were healthy control subjects.

Table I demonstrates the clinical characteristics of the study population. The mean age was 57.0 ± 9.2 years for healthy individuals and 59.9 ± 9.0 years for glaucoma patients ($p = 0.103$). There was no significant difference between the control and glaucoma groups regarding IOP (14.8 ± 2.8 mmHg and 13.8 ± 2.5 mmHg, respectively) ($p = 0.062$), although glaucoma patients were using a mean number of 2.2 ± 1.2 medications to lower IOP. Mean MD values were -4.1 ± 2.4 dB for glaucoma patients and -1.5 ± 1.6 dB for healthy individuals ($p < 0.001$). Among the glaucoma patients, 82.3% ($n = 51$) were classified as having early VF damage and 17.7% ($n = 11$) as having moderate VF damage.

Table II compares the mean RNFL thickness values in both groups. Mean RNFL thickness measurements were signifi-

cantly lower in the glaucoma group, except for the 3, 4, 8, and 9 o'clock positions.

The SD-OCT parameters with the greatest aROCs were as follows: inferior quadrant (0.813; 95% CI 0.727-0.881), average thickness (0.807; 95% CI 0.721-0.876), 7 o'clock position (0.765; 95% CI 0.674-0.840), and 6 o'clock position (0.754; 95% CI 0.663-0.831) (Tab. III). For a fixed specificity of 80%, the best sensitivities were observed with 7 o'clock position (64.5%), average thickness (62.1%), inferior quadrant (61.3%), and superior quadrant (56.4%) (Tab. III).

The aROCs with machine learning classifiers varied from 0.569 (CTREE; 95% CI 0.462-0.676) to 0.795 (SVMg; 95% CI 0.713-0.877) (Tab. IV) when trained with all parameters. All classifiers performed best when trained with a smaller number of parameters. When fewer parameters were used, aROCs varied from 0.785 (ADA; 95% CI 0.701-0.869) to 0.818 (BAG; 95% CI 0.740-0.896) (Tab. V). The best aROC obtained with BAG trained with 2 parameters (0.818) was not significantly different from the aROC obtained with the best single SD-OCT parameter (0.813) ($p = 0.93$) (Fig. 1).

DISCUSSION

Many studies have been published assessing the sensitivity and specificity of TD-OCT, and some compared the diagnostic ability of TD-OCT and SD-OCT, demonstrating that they show similar and good performances (9-13). The MLC techniques have been employed with various technologies designed to perform structural and functional

TABLE I - DEMOGRAPHIC CHARACTERISTICS OF THE STUDY POPULATION

	Healthy (n=48)	Glaucoma (n=62)	p
Age, y, mean \pm SD	57.0 \pm 9.2	59.9 \pm 9.0	0.103
Age, y, range	45-82	43-78	—
Male: female, n (%)	23 (47.9): 25 (52.1)	31 (50): 31 (50)	0.830
Caucasian: African American, n (%)	37 (77.1): 11 (22.9)	46 (74.2): 16 (25.8)	0.461
VA logMAR, mean \pm SD	0.04 \pm 0.09	0.1 \pm 0.1	0.003
SE, D, mean \pm SD	1.53 \pm 2.15	1.2 \pm 1.9	0.467
IOP, mmHg, mean \pm SD	14.8 \pm 2.8	13.8 \pm 2.5	0.062
Medications, mean \pm SD	0	2.2 \pm 1.2	<0.001
MD, dB, mean \pm SD	-1.5 \pm 1.6	-4.1 \pm 2.4	<0.001
PSD, dB, mean \pm SD	1.8 \pm 0.8	4.3 \pm 2.4	<0.001

IOP = intraocular pressure; MD = mean deviation; PSD = pattern standard deviation; SE = spherical equivalent; VA = visual acuity.

TABLE II - MEAN RETINAL NERVE FIBER LAYER THICKNESS MEASUREMENTS (\pm SD) OBTAINED BY SD-OCT IN BOTH POPULATIONS

OCT	Healthy (n=48)	Glaucoma (n=62)	p
Average (μm)	93.5 \pm 9.9	80.6 \pm 10.8	<0.001
Quadrant, μm			
Temporal	63.1 \pm 9.7	56.6 \pm 12.9	0.004
Superior	115.1 \pm 21.5	100.9 \pm 17.4	<0.001
Nasal	72.7 \pm 10.7	65.8 \pm 9.8	<0.001
Inferior	121.0 \pm 17.1	99.4 \pm 17.5	<0.001
Clock hour, μm			
1	104.0 \pm 18.5	91.4 \pm 19.1	<0.001
2	90.7 \pm 16.9	77.6 \pm 13.3	<0.001
3	61.2 \pm 10.2	58.6 \pm 10.4	0.188
4	66.1 \pm 12.5	61.1 \pm 13.0	0.047
5	99.6 \pm 20.4	87.4 \pm 16.8	<0.001
6	133.3 \pm 25.2	109.4 \pm 24.4	<0.001
7	130.1 \pm 23.7	101.4 \pm 31.5	<0.001
8	63.6 \pm 12.1	57.9 \pm 16.5	0.050
9	50.9 \pm 10.7	46.5 \pm 14.0	0.074
10	75.0 \pm 13.9	65.2 \pm 14.7	<0.001
11	124.3 \pm 22.1	103.5 \pm 24.9	<0.001
12	123.3 \pm 26.1	107.6 \pm 24.8	0.001

OCT = optical coherence tomography; SD-OCT = spectral domain optical coherence tomography.

TABLE III - AREAS UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE (AROC) FOR EACH SD-OCT PARAMETER AND SENSITIVITIES (%) WITH FIXED SPECIFICITIES OF 80% AND 90%

OCT	aROC (95% CI)	Specificity 80%	Specificity 90%
Average thickness	0.807 (0.721-0.876)	62.1	54.0
Quadrant			
Temporal	0.675 (0.579-0.761)	50.0	33.1
Superior	0.737 (0.645-0.816)	56.4	46.7
Nasal	0.685 (0.590-0.771)	49.2	26.6
Inferior	0.813 (0.727-0.881)	61.3	53.2
Clock hour			
1	0.703 (0.608-0.786)	45.2	29.8
2	0.723 (0.630-0.804)	51.6	34.7
3	0.574 (0.476-0.668)	29.8	22.6
4	0.605 (0.507-0.696)	29.0	11.3
5	0.671 (0.575-0.757)	43.5	27.4
6	0.754 (0.663-0.831)	47.6	33.9
7	0.765 (0.674-0.840)	64.5	38.7
8	0.631 (0.534-0.721)	42.7	27.4
9	0.625 (0.527-0.715)	44.3	36.3
10	0.699 (0.604-0.783)	50.8	42.0
11	0.740 (0.648-0.819)	40.3	33.9
12	0.672 (0.576-0.759)	38.7	21.0

CI = confidence interval; OCT = optical coherence tomography; SD-OCT = spectral domain optical coherence tomography.

TABLE IV - AREAS UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE (AROC) OBTAINED WITH MACHINE LEARNING CLASSIFIERS AND SENSITIVITIES (%) WITH FIXED SPECIFICITIES OF 80% AND 90%

MLC	aROC (95% CI)	Specificity 80%	Specificity 90%
BAG	0.752 (0.662-0.842)	64.5	40.3
NB	0.788 (0.704-0.872)	66.1	56.4
SVMl	0.788 (0.704-0.872)	62.9	56.4
SVMg	0.795 (0.713-0.877)	62.9	56.4
MLP	0.591 (0.485-0.697)	59.7	51.6
RBF	0.672 (0.572-0.772)	57.2	50.0
RAN	0.758 (0.669-0.847)	61.3	46.8
ENS	0.721 (0.627-0.815)	73.4	46.8
CTREE	0.569 (0.462-0.676)	59.7	52.4
ADA	0.744 (0.653-0.835)	62.1	50.0

ADA = Ada Boost M1; BAG = Bagging; CI = confidence interval; CTREE = Classification Trees; ENS = Ensemble Selection; MLC = machine learning classifier; MLP = Multilayer Perceptron; NB = Naive-Bayes; RAN = Random Forest; RBF = Radial Basis Function; SVMl = Linear Support Vector Machine; SVMg = Gaussian Support Vector Machine.

TABLE V - AREAS UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE (AROC) OBTAINED WITH MACHINE LEARNING CLASSIFIERS TRAINED WITH SOME SD-OCT PARAMETERS

MLC	aROC (no. parameters)	95% CI
BAG	0.818 (2)	0.740-0.896
NB	0.817 (2)	0.739-0.895
SVMl	0.815 (2)	0.737-0.893
SVMg	0.813 (2)	0.734-0.892
MLP	0.804 (1)	0.723-0.885
RBF	0.801 (1)	0.720-0.882
RAN	0.797 (7)	0.715-0.879
ENS	0.797 (1)	0.715-0.879
CTREE	0.789 (1)	0.706-0.872
ADA	0.785 (2)	0.701-0.869

ADA = Ada Boost M1; BAG = Bagging; CI = confidence interval; CTREE = Classification Trees; ENS = Ensemble Selection; MLC = machine learning classifier; MLP = Multilayer Perceptron; NB = Naive-Bayes; RAN = Random Forest; RBF = Radial Basis Function; SD-OCT = spectral domain optical coherence tomography; SVMl = Linear Support Vector Machine; SVMg = Gaussian Support Vector Machine.

evaluation of glaucoma, including TD-OCT (15, 17, 19, 23), HRT (18, 20, 22), GDx (16), and visual field (14, 19-22, 24, 25).

Although 3 groups (15, 17, 23) have employed TD-OCT in conjunction with MLCs, there is no study evaluating the use of MLCs along with SD-OCT data. Bizios et al (17) applied a nonlinear dimensionality reduction method based on manifold learning for the transformation of Stratus OCT A-scan measurements, and optimized 2 MLCs (MLP and SVM) for the classification of OCT test

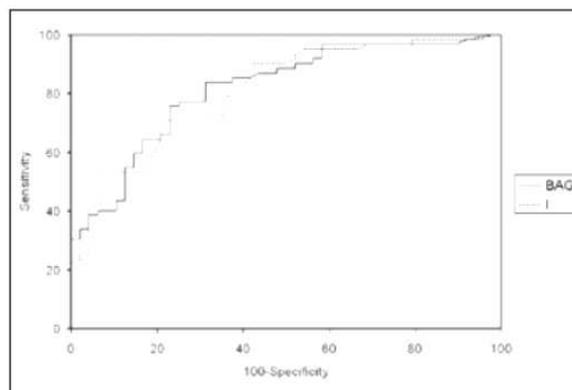


Fig. 1 - Areas under the receiver operating characteristic curve (aROC) of the best classifier trained with the number of spectral domain optical coherence tomography (SD-OCT) parameters which allowed the best performance (BAG = 0.818) and aROC of the best SD-OCT parameter (I = inferior quadrant = 0.813) ($p=0.93$).

data. The authors observed that, despite differences in the way MLP and SVM perform classification tasks, both methods were successful in recognizing pertinent patterns in the training data and applying this knowledge to classify new data, achieving similar performance. The aROCs were 0.982 for MLP (95% CI 0.966-0.999) and 0.989 for SVM (95% CI 0.979-1.0). In this study, patients with visual field damage (MD ≥ -12 dB) corresponding to glaucomatous changes in the ONH and/or the RNFL judged by examination of photographs were included.

Burgansky-Eliash et al (15) used the Stratus OCT and showed that the largest aROC among the RNFL parameters was obtained with average thickness (0.938) and that the best classifier was the SVM using 8 parameters (aROC=0.981). Similar to our findings, they demonstrated that the reduction in the number of OCT RNFL thickness parameters used by MLCs can yield improvements in classification performance. Furthermore, the accuracies and sensitivities obtained with the best individual parameter and with the classifiers were not significantly different ($p=0.05$), a finding that we also observed in our series.

To our knowledge, this study was the first to use MLCs with data obtained from SD-OCT. We compared the performance of SD-OCT parameters and MLCs in discriminating between healthy and glaucomatous eyes. The best individual parameter was inferior quadrant (aROC=0.813), the best classifier trained with all OCT parameters was SVMg (aROC=0.795), and the best classifier trained with some SD-OCT parameters was BAG (aROC=0.818). Although the aROC was slightly higher for this classifier, the difference was not statistically significant when compared to inferior quadrant ($p=0.93$). Hence, there was no significant difference in the diagnostic ability of the MLCs and the individual parameters of the OCT. Complex machine classifiers that use numerous input parameters tend to perform better in larger databases (15). Our sample size was probably small to allow the use of all SD-OCT parameters.

Comparisons across studies are very difficult, due to differences in population demographics, in the definition and severity of glaucoma, and because very few studies used OCT parameters with MLCs, and none of them with SD-OCT. The diagnostic ability of the parameters tested in this study was lower when compared with some previously published studies, which can be attributed to several reasons. First, a large proportion (82.3%) of our patients had early glaucomatous VF damage. Second, differences in mean age between the groups may also influence the accuracy of MLCs. Although age was not included in the training of the MLCs used by Bizios et al (17) and Burgansky-Eliash et al (15), it is well known that a reduction in RNFL thickness occurs with age, favoring the diagnosis of glaucoma in older patients.

In fact, the study by Huang and Chen (23), where patients had early glaucomatous damage ($MD > -6$ dB), and where there was no significant difference in age between the

glaucoma and control groups ($p=0.07$), showed aROCs similar to ours. The inferior quadrant thickness measured with the Stratus OCT was also the best individual parameter for differentiating between normal and glaucomatous eyes (aROC=0.832). The aROCs were 0.824 for linear discriminant analysis (LDA), 0.849 for Mahalanobis distance, and 0.821 for MLP. Principal component analysis (PCA) was performed on 25 Stratus OCT parameters and the results were reprocessed by LDA, Mahalanobis distance, and MLP, generating an improvement in the performance for each classifier, increasing aROCs to 0.915, 0.991, and 0.874, respectively. Interestingly, we attempted to use PCA in our series and were unable to improve our results.

In conclusion, the current study demonstrated that SD-OCT showed good diagnostic accuracy in a group of patients with early to moderate glaucoma. The MLCs did not improve the sensitivity and specificity of SD-OCT for the diagnosis of glaucoma in this series. Other studies including functional parameters in the training of the classifiers should be performed in order to evaluate if this association leads to increased diagnostic accuracy.

ACKNOWLEDGEMENTS

The authors thank FAPESP (grant #07/51281-9: Vital P. Costa) for support.

The authors report no proprietary interest.

Address for correspondence:
Vanessa Gonçalves Vidotti, MD
Av. José Bonifácio, 1425, apto 42
Campinas, S.P. 13091-140
Brazil
vavidotti@yahoo.com.br

REFERENCES

1. Sommer A, Katz J, Quigley HA, et al. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol* 1991; 109: 77-83.
2. Johnson CA, Cioffi GA, Liebmann JR, Sample PA, Zangwill LM, Weinreb RN. The relationship between structural and functional alterations in glaucoma: a review. *Semin Ophthalmol* 2000; 15: 221-33.
3. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science* 1991; 254: 1178-81.
4. Schuman JS, Hee MR, Arya AV, et al. Optical coherence tomography: a new tool for glaucoma diagnosis. *Curr Opin Ophthalmol* 1995; 6: 89-95.
5. Schuman JS, Hee MR, Puliafito CA, et al. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. *Arch Ophthalmol* 1995; 113: 586-96.
6. Harwerth RS, Vilupuru AS, Rangaswamy NV, Smith EL III. The relationship between nerve fiber layer and perimetry measurements. *Invest Ophthalmol Vis Sci* 2007; 48: 763-73.
7. Zangwill LM, Bowd C, Berry CC, et al. Discriminating between normal and glaucomatous eyes using the Heidelberg retina tomograph, GDx nerve fiber analyzer, and optical coherence tomograph. *Arch Ophthalmol* 2001; 119: 985-93.
8. Guedes V, Schuman JS, Hertzmark E, et al. Optical coherence tomography measurement of macular and nerve fiber layer thickness in normal and glaucomatous human eyes. *Ophthalmology* 2003; 110: 177-89.
9. Chang RT, Knight OJ, Feuer WJ, Budenz DL. Sensitivity and specificity of time-domain versus spectral-domain optical coherence tomography in diagnosing early to moderate glaucoma. *Ophthalmology* 2009; 116: 2294-9.
10. Jeoung JW, Park KH. Comparison of Cirrus OCT and Stratus OCT on the ability to detect localized retinal nerve fiber layer defects in preperimetric glaucoma. *Invest Ophthalmol Vis Sci* 2010; 51: 938-45.
11. Park SB, Sung KR, Kang SY, Kim KR, Kook MS. Comparison of glaucoma diagnostic capabilities of Cirrus HD and Stratus optical coherence tomography. *Arch Ophthalmol* 2009; 127: 1603-9.
12. Moreno-Montañés J, Olmo N, Alvarez A, Garcia N, Zarranz-Ventura J. Cirrus high-definition optical coherence tomography compared with Stratus optical coherence tomography in glaucoma diagnosis. *Invest Ophthalmol Vis Sci* 2010; 51: 335-43.
13. Knight OJ, Chang RT, Feuer WJ, Budenz DL. Comparison of retinal nerve fiber layer measurements using time domain and spectral domain optical coherent tomography. *Ophthalmology* 2009; 116: 1271-7.
14. Bizios D, Heijl A, Bengtsson B. Trained artificial neural network for glaucoma diagnosis using visual field data: a comparison with conventional algorithms. *J Glaucoma* 2007; 16: 20-8.
15. Burgansky-Eliash Z, Wollstein G, Chu T, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Invest Ophthalmol Vis Sci* 2005; 46: 4147-52.
16. Bowd C, Medeiros FA, Zhang Z, et al. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci* 2005; 46: 1322-9.
17. Bizios D, Heijl A, Hougaard JL, Bengtsson B. Machine learning classifiers for glaucoma diagnosis based on classification of retinal nerve fibre layer thickness parameters measured by Stratus OCT. *Acta Ophthalmol* 2010; 88: 44-52.
18. Townsend KA, Wollstein G, Danks D, et al. Heidelberg Retina Tomograph 3 machine learning classifiers for glaucoma detection. *Br J Ophthalmol* 2008; 92: 814-8.
19. Bowd C, Hao J, Tavares IM, et al. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2008; 49: 945-53.
20. Horn FK, Lämmer R, Mardin CY, et al. Combined evaluation of frequency doubling technology perimetry and scanning laser ophthalmoscopy for glaucoma detection using automated classification. *J Glaucoma* 2012; 21: 27-34.
21. Boden C, Chan K, Sample PA, et al. Assessing visual field clustering schemes using machine learning classifiers in standard perimetry. *Invest Ophthalmol Vis Sci* 2007; 48: 5582-90.
22. Mardin CY, Peters A, Horn F, Jünemann AG, Lausen B. Improving glaucoma diagnosis by the combination of perimetry and HRT measurements. *J Glaucoma* 2006; 15: 299-305.
23. Huang ML, Chen HY. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci* 2005; 46: 4121-9.
24. Horn FK, Brenning A, Jünemann AG, Lausen B. Glaucoma detection with frequency doubling perimetry and short-wavelength perimetry. *J Glaucoma* 2007; 16: 363-71.
25. Tucker A, Vinciotti V, Liu X, Garway-Heath D. A spatio-temporal Bayesian network classifier for understanding visual field deterioration. *Artif Intell Med* 2005; 34: 163-77.
26. Mitchell TM. *Machine Learning*. Singapore: McGraw-Hill, 1997; 81-126; 154-99.
27. Quinlan JR. *Programs for Machine Learning*. San Mateo: Morgan Kaufmann, 1993; 17-25.
28. Quinlan JR. Induction of decision trees. *Mach Learn* 1986; 1: 81-106.
29. Breiman L. Random forests. *Mach Learn* 2001; 45: 5-32.
30. Breiman L. Bagging predictors. *Mach Learn* 1996; 24: 123-40.
31. Haykin S. *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall, Inc., 1999; 183-282.
32. Keerthi SS, Shevade SK, Bhattacharyya C, et al. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Comput* 2001; 13: 637-49.
33. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Mining Knowledge Discovery* 1998; 2: 121-67.
34. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco: Morgan Kaufmann, 2005; 365-8.