

Marcelo Dias

**Geração de Populações Artificiais para
Treinamento de Classificadores de Apoio ao
Diagnóstico de Glaucoma**

São Paulo
2009

Marcelo Dias

Geração de Populações Artificiais para Treinamento de Classificadores de Apoio ao Diagnóstico de Glaucoma

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do Título de Doutor em Engenharia Elétrica.

Área de concentração:
Sistemas Digitais

Orientador:
Prof. Dr. Edson Satoshi Gomi

São Paulo
2009

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com anuência de seu orientador.

São Paulo, de junho de 2009.

Assinatura do autor

Assinatura do orientador

Ficha Catalográfica

Dias, Marcelo

Geração de Populações Artificiais para Treinamento de Classificadores de Apoio ao Diagnóstico de Glaucoma. / M.Dias. – ed.rev. – São Paulo, 2009. 109 p.

Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1. Inteligência artificial 2. Aprendizado computacional 3. Métodos estatísticos para aprendizagem 4. Glaucoma 5. Diagnóstico por computador I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais. II. t.

Ao meu amado filho Sacha (*in memoriam*),
cuja alegria de perceber um lindo dia de verão em toda a sua plenitude lhe foi furtada
pelo glaucoma.

Agradecimentos

Agradecimento especial ao Prof. Dr. Edson Satoshi Gomi, orientador, que me acolheu na Escola Politécnica e, com muita paciência, amizade e efetivo comprometimento, possibilitou a realização deste estudo.

Ao Prof. Dr. Vital Paulino Costa, pela receptividade e apoio no projeto e pelas verdadeiras aulas de glaucoma para engenheiros.

Ao Prof. Dr. Edison Spina e Prof. Dr. Marco Túlio de Andrade, pelos ensinamentos dentro e fora das salas de aula, e pela amizade.

Aos meus colegas do laboratório KNOMA, especialmente ao Fabrício Barth, companheiro de jornada.

Aos meus pais Osmar Francisco Dias e Ariene Pregun Dias, pela dádiva da vida e por serem meus primeiros e eternos professores.

Aos meus filhos Sacha Dias e João Pedro Estevo Dias, pela torcida silenciosa e por sempre tolerarem minha ausência.

À minha esposa Alice Estevo Dias, pelo amor, apoio e incentivo constantes.

Agradecimento especial à Fundação de Amparo à Pesquisa do Estado de São Paulo -
FAPESP - processo número 2007/51281-9.

Resumo

Glaucoma é uma neuropatia óptica cuja progressão gera comprometimento no campo visual e cegueira. Devido aos danos irreversíveis, a identificação precoce e correta é vital para o controle da progressão da doença. Para diagnóstico de glaucoma, oftalmologistas analisam dados de campo visual e da anatomia ocular obtidos através de testes. Para reduzir a quantidade de resultados falso negativos e falso positivos, diversas técnicas tem sido desenvolvidas para incrementar a sensibilidade e especificidade dos testes diagnósticos de glaucoma. Para aplicações médicas, existem estudos que indicam como uma abordagem promissora o uso de classificadores de aprendizagem de máquina baseados em diferentes conceitos como Árvores de Decisão, Redes Neurais Artificiais e abordagem Bayesiana. Apesar da disponibilidade de sofisticados algoritmos para desenvolvimento de classificadores, o sucesso do treinamento destes classificadores é altamente dependente de dados de qualidade. Dados de qualidade significam que os exemplos fornecidos para treinamento devem representar a maior quantidade possível de situações encontradas no mundo real. Estes requisitos são normalmente atendidos se dados de uma grande quantidade de pacientes estiver disponível. Entretanto, diversos fatores como o perfil da população avaliada, a duração das tarefas de coleta de dados, disponibilidade de equipamentos e de profissionais de saúde, além do comprometimento dos pacientes com o programa de pesquisa, restringem o tamanho do conjunto de dados. Uma possível estratégia para resolver o problema da escassez de dados para a tarefa de treinamento dos classificadores é o emprego de dados artificiais que representam populações reais. Estes dados artificiais serão adequados ao treinamento dos classificadores se possuírem características estatísticas de populações reais. O uso de populações artificiais possibilitará a criação de conjuntos de dados com número adequado de pacientes, sem gastar anos coletando dados. Neste trabalho é apresentado um gerador de dados artificiais denominado GLOR, baseado em um método de Monte Carlo, que é adequado ao treinamento de classificadores para o diagnóstico de glaucoma. A população gerada é caracterizada por dados funcionais e estruturais fornecidos pelos instrumentos de perimetria computadorizada padrão ou *Standard Automated Perimetry* (SAP) e tomografia de coerência óptica de alta definição ou *High Definition Optical Coherence Tomography* (HD-OCT). Os resultados experimentais, obtidos após o treinamento de uma Rede Neural Artificial empregando população gerada pelo GLOR contendo 4500 indivíduos normais e 500 glaucomatosos e avaliação com dados de uma população real formada por 44 indivíduos normais e 26 glaucomatosos, foram: acurácia total de 87,1%, sensibilidade de 80,8%, especificidade de 90,9% e área sob curva ROC de 0,941. Tais resultados indicam que o GLOR pode ser empregado satisfatoriamente no desenvolvimento de novos métodos que possam elevar a sensibilidade e especificidade no diagnóstico de glaucoma.

Palavras-chaves: inteligência artificial, classificadores de aprendizagem de máquina, método de Monte Carlo, diagnóstico de glaucoma, simulação computacional.

Abstract

Glaucoma is an optical neuropathy, whose progression results in visual field impairments and blindness. Due to its irreversible damages, early and correct identification is very important to control glaucoma's progression. For glaucoma diagnosis, ophthalmologists analyze patient's visual field and eyes structural data obtained by using eyes' test equipments. In order to reduce the quantity of false-negative and false-positive results, several new techniques have been developed to increment the sensitivity and specificity of glaucoma diagnostic tests. A promising approach is the use of machine learning classifiers. Classifiers based on different concepts like Decision Trees, Artificial Neural Networks, and Bayesian approach, have been developed for medical applications. Despite the availability of sophisticated algorithms for classifiers development, successful training of classifiers is highly dependent of good training data. Good data means that examples provided for classifier training should represent the many different situations found in real world. These requirements are usually accomplished if data from a large number of patients is available. However, several factors like profile of evaluated population, duration of the data acquisition activities, existence of healthcare professionals, and equipment availability, and people's commitment to the research program, restrict the size of patient's dataset. A possible approach to overcome the lack of patient's data to perform the classifier's training task is to use artificial data that represent a real population. This artificial data would be suitable for classifiers training if it has similar statistical properties of a real population. The use of artificial population will enable the creation of datasets with required number of patients, and without spending years measuring patients. It will also be possible to simulate scenarios and strategies before a long term research program starts. In this work is presented an artificial data generator named GLOR, based on a Monte Carlo method, and suitable for the training of classifiers for glaucoma diagnosis. The generated population is characterized by eyes'functional and structural data provided by Standard Automated Perimetry (SAP) and High Definition Optical Coherence Tomography (HD-OCT) instruments. The experimental results, obtained after an Artificial Neural Network training employing a population generated by GLOR comprising of 4500 normal and 500 glaucomatous individuals and evaluated by using real population data from 44 normal and 26 glaucomatous subjects, were: 87.1% for overall accuracy, 80.8% for sensitivity, 90.9% for specificity and 0.941 for the area under ROC curve. These results show that GLOR can be used as a promising approach to accelerate the development of new methods to increment sensitivity and specificity of glaucoma diagnosis.

Keywords: artificial intelligence, machine learning classifiers, Monte Carlo method, glaucoma diagnosis, computer simulation.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

1	Introdução	1
1.1	Objetivos da Pesquisa	9
1.2	Metodologia da Pesquisa	10
1.3	Estrutura do trabalho	10
2	Diagnóstico de Glaucoma	11
2.1	Outros Métodos Diagnósticos	19
2.1.1	Métodos Funcionais	19
2.1.2	Métodos Estruturais	21
2.2	Interpretação de testes diagnósticos	25
2.2.1	Sensibilidade	26
2.2.2	Especificidade	26
2.2.3	Curvas ROC (<i>Receiver Operating Characteristic</i>)	27
3	Classificadores	31
3.1	Categorias de classificadores	33
3.1.1	Árvores de decisão	34
3.1.2	Naive-Bayes	38

3.1.3	Redes neurais artificiais de múltiplas camadas	41
3.1.4	<i>Bagging</i>	44
3.2	Classificadores híbridos	45
3.3	Avaliando a robustez de classificadores	46
3.4	Classificadores para Glaucoma	49
4	O Gerador GLOR	51
4.1	Geração dos Atributos Idade, PIO e Espessuras Setoriais do OCT	53
4.2	Geração dos Demais Parâmetros de OCT	61
4.3	Bancos de Dados Normativos dos Instrumentos OCT e SAP	64
4.3.1	Reconstrução do Banco de Dados Normativo do Cirrus HD-OCT	64
4.3.2	Obtenção do Banco de Dados Normativo do HFA-II	67
4.4	A Geração de Dados de SAP a partir de Dados de OCT	72
4.5	A Saída Gráfica do GLOR	75
5	Avaliação dos Resultados	79
5.1	Avaliação Qualitativa do GLOR	79
5.2	Avaliação Quantitativa do GLOR	80
5.3	Avaliação dos Classificadores com Dados Artificiais	83
5.4	Avaliação dos Classificadores com Dados Reais	88
6	Considerações Finais	95
6.1	Contribuições deste Trabalho	98
6.2	Trabalhos Futuros	98
	Referências	100
	Apêndice A – Projeto FAPESP: Critérios de Inclusão e Exclusão	107
A.1	Grupo 1 - Indivíduos Normais	107

A.1.1	Critérios de Inclusão	107
A.1.2	Critérios de Exclusão	107
A.2	Grupo 2 - Indivíduos Glaucomatosos	108
A.2.1	Critérios de Inclusão	108
A.2.2	Critérios de Exclusão	108
A.3	Grupo 3 - Hipertensos Oculares	109

Lista de Figuras

1.1	Anatomia do olho humano. Adaptado de (DUANE'S, 2002)	2
1.2	Drenagem do humor aquoso (MERCK & CO. INC., 2004)	3
1.3	Glaucoma primário de ângulo aberto (A) e fechado (B). Adaptado de (DUANE'S, 2002)	3
1.4	Variação da pressão intra-ocular ao longo do dia. Adaptado de (AZUARA-BLANCO; COSTA; WILSON, 2002).	4
1.5	Imagem em três dimensões do fundo de olho obtida através do instrumento HRT II destacando-se o disco óptico ao centro, a área de escavação (CUP), a rima neural (RIM) e a camada de fibras nervosas (RNFL) (HEIDELBERG-ENGINEERING, 2009).	6
2.1	Resultado de perimetria computadorizada indicando defeitos no exame de campo visual (AZUARA-BLANCO; COSTA; WILSON, 2002).	15
2.2	Segmentos da retina que são avaliados pelo OCT (ZEISS, 2009).	16
2.3	Comparação entre as estratégias de varredura do Cirrus HD-OCT à esquerda e Stratus OCT à direita (ZEISS, 2009).	17
2.4	Resultado de tomografia de coerência óptica de alta definição (Cirrus HD-OCT) de um indivíduo normal voluntário.	18
2.5	Exemplos de resultado de SWAP (A) e SAP (B) para um mesmo paciente (ZEISS, 2009).	20
2.6	Exemplo de resultado de FDT (ZEISS, 2009).	21
2.7	Princípio de funcionamento da polarimetria de varredura a laser (ZEISS, 2009).	22
2.8	Imagens de reflectância (esquerda) e retardo (direita) obtidas através de GDx-VCC (ZEISS, 2009).	22

2.9	Exemplo de imagem em 3 dimensões de indivíduo com glaucoma avançado obtida através do instrumento HRT II (HEIDELBERG-ENGINEERING, 2009).	23
2.10	Exemplo de imagem de um olho glaucomatoso com indicações de segmentos de acordo com o índice de Classificação de Regressão de Moorfields (HEIDELBERG-ENGINEERING, 2009).	24
2.11	Limiar de decisão em um teste diagnóstico (TAPE, 2006)	27
2.12	Curva ROC para teste diagnóstico de glaucoma de acordo com valor da medida da pressão intra-ocular.	30
3.1	O problema de classificação de glaucoma.	31
3.2	Exemplo de árvore de decisão para representação do conceito JogarTênis. Adaptado de (MITCHELL, 1997).	35
3.3	Grafo estrutural de um <i>perceptron</i> de múltiplas camadas com duas camadas ocultas.	42
3.4	Grafo estrutural de um <i>perceptron</i> de múltiplas camadas com duas camadas ocultas e retro propagação de erro.	42
3.5	Grafo estrutural de uma rede neural artificial de três camadas para representar a decisão JogarTênis.	44
3.6	Conjunto de dados de treinamento do conceito JogarTênis no formato ARFF usado no ambiente de aprendizagem WEKA.	47
3.7	Saída em formato texto do experimento realizado com o classificador ID3 no WEKA.	48
3.8	Saída em formato texto do experimento realizado com o classificador Naive-Bayes no WEKA.	48
3.9	Saída em formato texto do experimento realizado com o classificador <i>Multilayer Perceptrons</i> no WEKA.	49
4.1	Processo de geração de amostras populacionais artificiais pelo GLOR.	52
4.2	Figura de um quadrado contendo uma circunferência inscrita de raio r utilizada na aproximação do valor de π através da aplicação do método Monte Carlo.	54
4.3	Exemplo de $f(x)$ implementada através da função <i>DIST.NORM</i> .	56

4.4	Exemplo de $F(x)$ implementada através da função <i>DIST.NORM.</i>	57
4.5	Exemplo de $F^{-1}(y)$ implementada através da função <i>INV.NORM.</i>	59
4.6	Espessuras médias por setor, quadrantes e total obtidas através da saída gráfica do instrumento Cirrus HD-OCT.	62
4.7	Processo de obtenção das médias por setor, quadrantes e total a partir dos dados discretos que compõem o gráfico de espessuras da camada de fibras nervosas.	63
4.8	Resumo das 4 etapas envolvidas na geração artificial inversa dos dados de Cirrus HD-OCT.	64
4.9	Gráfico normativo das espessuras da camada de fibras nervosas (<i>RNFL Thickness Normative Data</i>) de um indivíduo voluntário obtido com o instrumento Cirrus HD-OCT.	65
4.10	Método para identificação dos limiares de espessura por intervalo de confiança.	66
4.11	Valores normativos (μm) por intervalo de confiança e faixa etária para 27 pontos obtidos através da digitalização das saídas gráficas do Cirrus HD-OCT.	68
4.12	Exemplo de saída gráfica do exame de campo visual produzida pelo equipamento Humphrey HFA-II.	69
4.13	Mapa de campo visual obtido através do programa C30-2 <i>full threshold</i>	71
4.14	Mapeamento entre pontos do campo visual (A) e setores da cabeça do nervo óptico (B). Adaptado de (HARWERTH; WHEAT, 2008)	73
4.15	Exemplo de saída gráfica do GLOR para uma determinada instância obtida artificialmente.	76
4.16	Saída gráfica do GLOR destacando os parâmetros epidemiológicos e clínicos (em verde), estruturais (em vermelho) e funcionais (em azul) gerados pelo GLOR.	78
5.1	Distribuição de frequências de ocorrência para valores de pressão intraocular gerados pelo GLOR de acordo com características de uma população normal (n=4500).	82

5.2	Distribuição de frequências de ocorrência para valores de pressão intra-ocular gerados pelo GLOR de acordo com características de uma população com glaucoma (n=500).	82
-----	---	----

Lista de Tabelas

2.1	Modelo de tabela 2x2 para interpretação de testes diagnósticos	25
2.2	Exemplo de tabela 2x2 para interpretação de testes diagnósticos	26
2.3	Exemplo de tabela com resultados da classificação de glaucoma de acordo com a medida da pressão intra-ocular.	28
2.4	Tabela de classificação de glaucoma de acordo com limiar 21 para valor da medida da pressão intra-ocular.	28
2.5	Tabela de classificação de glaucoma de acordo com limiar 27 para valor da medida da pressão intra-ocular.	29
2.6	Tabela de classificação de glaucoma de acordo com limiar 21 para valor da medida da pressão intra-ocular.	29
2.7	Tabela de sensibilidades e especificidades de acordo com variação do limiar para valor da medida da pressão intra-ocular.	29
2.8	Tabela base para construção da Curva ROC	30
3.1	Exemplos de treinamento para o conceito JogarTênis. Adaptado de (MITCHELL, 1997)	36
4.1	Lista de atributos e correspondentes médias e desvios padrões em população normal e com glaucoma.	60
5.1	Médias e desvios padrões por atributos gerados artificialmente e respectivos erros percentuais em relação aos valores esperados definidos a priori.	81
5.2	Lista de atributos numéricos gerados pelo GLOR e utilizados no treinamento e teste dos classificadores, agrupados por categoria.	83
5.3	Resultados dos testes dos classificadores com dados artificiais organizados por classificador e <i>dataset</i>	86
5.4	Resultados dos testes dos classificadores com dados artificiais de acordo com o critério de acurácia total.	87

5.5	Resultados dos testes dos classificadores com dados artificiais de acordo com o critério de área sob a curva ROC.	88
5.6	Resultados dos testes dos classificadores com dados reais por classificador e <i>dataset</i>	89
5.7	Resultados dos testes dos classificadores com dados reais de acordo com o critério de acurácia total.	90
5.8	Resultados dos testes dos classificadores com dados reais de acordo com o critério de área sob a curva ROC.	91
5.9	Resultados dos testes dos classificadores com dados reais de acordo com o critério de sensibilidade.	92
5.10	Resultados dos testes dos classificadores com dados reais de acordo com o critério de especificidade.	93
5.11	Resultados dos testes dos classificadores com dados reais segmentados por <i>dataset</i>	94

Lista de Abreviaturas

ARFF Attribute-Relation File Format

AROC Área sob curva ROC

Bagged-CTREE Bagged Classification Tree

CSV Comma Separated Values

CTREE Classification Tree

dB Decibels

FDT Frequency Doubling Perimetry

GAM Generalized Additive Model

GDx-VCC Scanning Laser Polarimetry with Variable Corneal Compensation

GHT Glaucoma Hemifield Test

GLM Generalized Linear Model

GLOR Glaucoma Oracle

HD-OCT High Definition Optical Coherence Tomography

HRT Heidelberg Retina Tomograph

KBANN Knowledge-Based Artificial Neural Network

LDA Linear Discriminant Analysis

MC Monte Carlo

MD Mean Deviation

MGG Mixture of Generalized Gaussian

MLP Multilayer Perceptrons

MoG Mixture of Gaussian

OCT Optical Coherence Tomography

PIO Pressão Intra-ocular

PSD Pattern Standard Deviation

RNFL Retinal Nerve Fiber Layer

ROC Receiver Operating Characteristic

RPART Recursive Partitioning and Regression Tree

SAP Standard Automated Perimetry

SD-OCT Spectral Domain Optical Coherence Tomography

SVM Support Vector Machines

SWAP Short Wavelength Perimetry

TD-OCT Time Domain Optical Coherence Tomography

1 Introdução

O glaucoma é uma doença crônica, assintomática, de lenta progressão e multifatorial. Caracteriza-se por perda de fibras nervosas, danos ao nervo óptico e defeitos do campo visual (WEINREB; KHAW, 2004), e, se não for corretamente diagnosticada e tratada, pode levar à cegueira.

O glaucoma é comum após os 35 anos de idade, mas, ocasionalmente, ocorre em crianças. A Organização Mundial de Saúde estima que esta doença seja a segunda maior causa de cegueira na população mundial, e a primeira causa de cegueira irreversível (KINGMAN, 2004). A prevalência do glaucoma varia conforme a localização geográfica, a etnia e condições sócio-econômicas da população estudada (BANK, 1993). Sua prevalência varia entre 1,1% e 2,1% em caucasianos (KLEIN, 1992; DIELEMANS, 1994), sendo 3 a 4 vezes maior em negros (TIELSH, 1991; MASON, 1989).

Segundo os resultados do The Eye Diseases Prevalence Research Group (LESKE, 1994), uma análise sobre glaucoma nos Estados Unidos da América (EUA), a prevalência dessa doença foi estimada em 1,86% na população acima de 40 anos, o que equivale a 1,57 milhão de brancos e 398.000 negros afetados. Ainda segundo esse estudo, considerando-se a taxa atual de envelhecimento da população daquele país, o número de portadores de glaucoma crescerá em 50%, acometendo 3,36 milhões de indivíduos em 2020.

Na figura 1.1 pode-se observar a representação das câmaras anterior e posterior do olho humano. Estas câmaras são preenchidas por um líquido de baixa densidade denominado humor aquoso. Este líquido é produzido na câmara posterior do olho, flui através da pupila para a câmara anterior, e em seguida é drenado pelos canais de saída para fora do olho. Na figura 1.2 pode-se observar uma ilustração deste fluxo. Quando este fluxo é dificultado por algum problema como uma obstrução, há uma conseqüente elevação da pressão interna que o mesmo exerce sobre as estruturas oculares. Esta pressão (ou tensão) é denominada pressão intra-ocular (PIO). A pressão intra-ocular elevada é o fator de risco mais importante para o desenvolvimento do glaucoma (WEINREB; KHAW, 2004; HEIJL et al., 2002) e o único passível de tratamento (WEINREB; KHAW, 2004; HEIJL et al.,

2002; KASS et al., 2002). Uma das principais estruturas que podem ser lesionadas por este aumento de pressão é o nervo óptico, gerando danos irreversíveis à visão dos pacientes, uma vez que as células nervosas não sofrem regeneração.

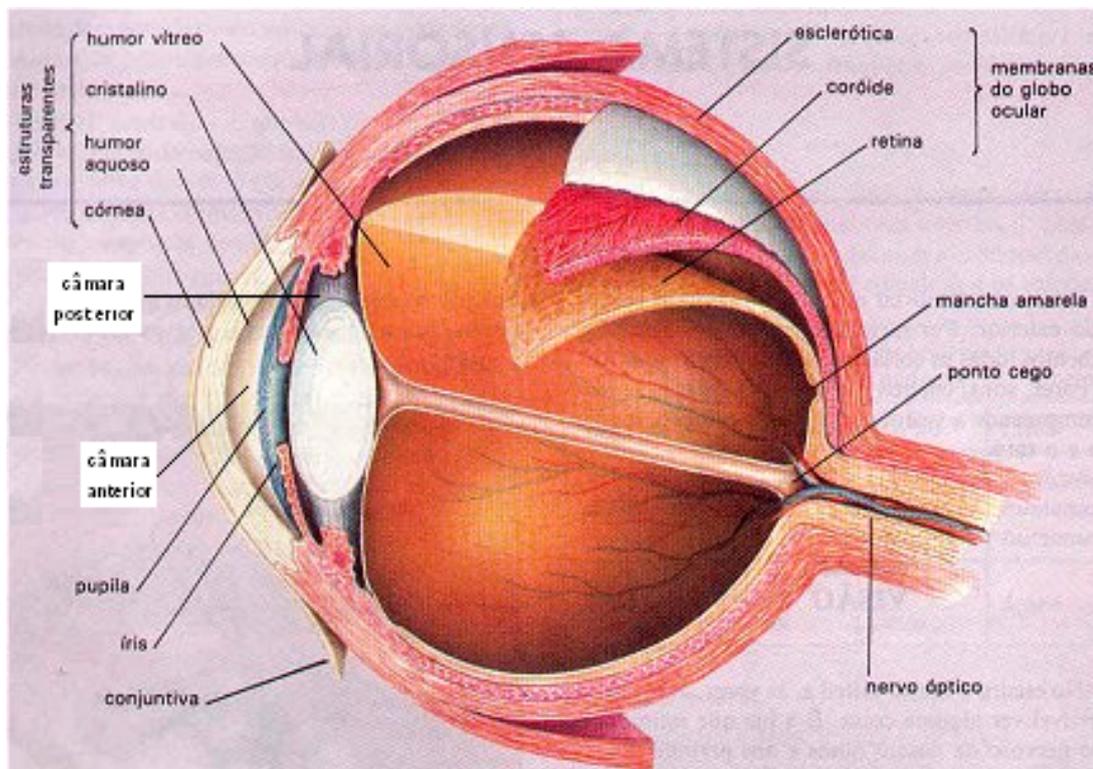


Figura 1.1: Anatomia do olho humano. Adaptado de (DUANE'S, 2002)

Dependendo do tipo de anomalia apresentada pelos canais de saída, o glaucoma recebe uma classificação correspondente: quando estão abertos, a doença é denominada glaucoma primário de ângulo aberto (GPAA), caso estejam obstruídos pela íris, tem-se o glaucoma primário de ângulo fechado (GPAF). A comparação entre as duas formas de glaucoma primário pode ser observada anatomicamente através da figura 1.3, onde o ângulo que a câmara anterior forma com a íris pode se apresentar aberto (A) ou fechado (B).

No GPAA, o líquido drena muito lentamente da câmara anterior. A pressão aumenta gradualmente, quase sempre em ambos os olhos, causando lesão do nervo óptico e uma perda da visão lenta e progressiva. Esta perda inicia-se na periferia do campo visual e, caso não controlada, pode comprometer inteiramente a visão do paciente, causando cegueira.

Um oftalmologista pode medir a pressão na câmara anterior através de um procedimento simples e indolor denominado tonometria. Dentre as diversas técnicas de medição de pressão, a mais utilizada é a tonometria Goldman. Neste exame, após a aplicação de um colírio anestésico, um pequeno instrumento é aproximado da córnea do paciente de

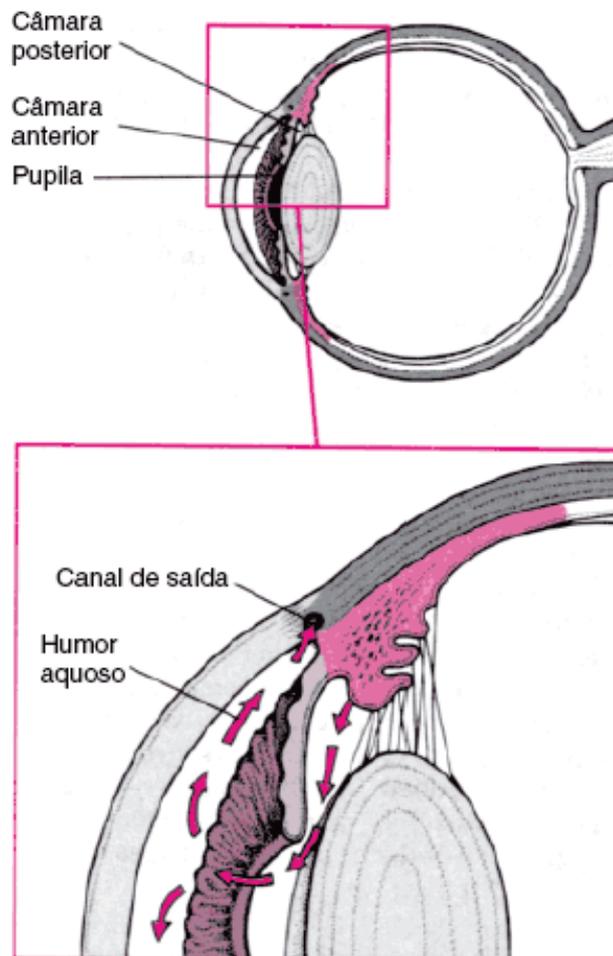


Figura 1.2: Drenagem do humor aquoso (MERCK & CO. INC., 2004)



Figura 1.3: Glaucoma primário de ângulo aberto (A) e fechado (B). Adaptado de (DUANE'S, 2002)

tal forma que a resistência apresentada pelo olho é medida, obtendo-se então a PIO.

Diversos estudos (AZUARA-BLANCO; COSTA; WILSON, 2002) apontam que a pressão intra-ocular em pacientes normais é geralmente definida entre 10 e 21 mmHg, com média de 16 mmHg e desvio padrão de 2.5. Geralmente, as medidas superiores a 22 milímetros indicam um aumento de pressão. Dado que os valores da PIO variam ao longo do dia (variação circadiana), normalmente diversas medidas em horários distintos devem ser tomadas e uma curva tensional então deve ser desenhada para identificar-se efetivamente a alteração dos níveis da PIO. A figura 1.4 apresenta um exemplo de curvas tensionais ao longo de um dia. A linha grafada em azul representa uma curva dentro dos limites considerados normais, enquanto que a vermelha demonstra um gráfico com níveis de variação de PIO típicos de glaucoma.

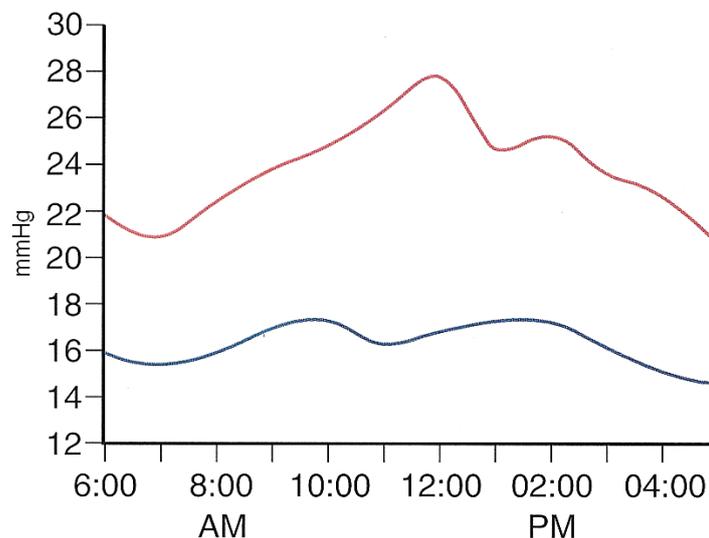


Figura 1.4: Variação da pressão intra-ocular ao longo do dia. Adaptado de (AZUARA-BLANCO; COSTA; WILSON, 2002).

No entanto, existem indivíduos que apesar de apresentarem PIO estatisticamente elevada não desenvolvem glaucoma, além de indivíduos glaucomatosos que possuem níveis de PIO pertencentes à faixa de normalidade (AZUARA-BLANCO; COSTA; WILSON, 2002).

Para diagnóstico do glaucoma atualmente existem diversos testes, cuja eficácia é avaliada em termos de dois indicadores: sensibilidade e especificidade. Sensibilidade é a proporção de indivíduos com a doença que foram corretamente identificados pelo teste, enquanto que especificidade é a fração de indivíduos normais, ou seja, sem a doença, que foram detectados como normais. Elevar a sensibilidade de um teste diagnóstico é uma forma de minimizar sua taxa de falsos negativos, ou seja, a quantidade percentual de pacientes com a doença que foram erroneamente classificados como normais. Aumentar

a especificidade significa reduzir a taxa de falsos positivos, ou melhor, a fração de pacientes que, apesar de normais, foram incorretamente classificados como portadores da doença pelo teste. Um resultado falso positivo para glaucoma apresenta impactos sócio-econômicos negativos, uma vez que implica no início de tratamento desnecessariamente, incorrendo em custos relacionados ao uso de medicação e de visitas ao serviço médico, além dos eventuais riscos e desconfortos ao paciente que qualquer tratamento pode causar. Já o custo humano de um resultado falso negativo em glaucoma é ainda maior. Um paciente glaucomatoso, quando erroneamente classificado como normal, pode ter sua visão irreversivelmente comprometida pela falta de intervenção médica, podendo até mesmo evoluir para cegueira caso o erro diagnóstico não seja corrigido em futuro próximo.

Desta forma, diversos testes funcionais e estruturais têm sido desenvolvidos e aprimorados ao longo dos últimos anos, sempre com o objetivo de aumentar a sensibilidade e especificidade do diagnóstico de glaucoma. Testes funcionais procuram identificar o comprometimento da visão dos pacientes (função) enquanto que os estruturais avaliam características anatômicas do olho (estruturas oculares) em busca de anomalias.

Um exemplo de teste diagnóstico amplamente utilizado para avaliação funcional de glaucoma é a perimetria computadorizada ou exame de campo visual computadorizado. Este exame consiste em apresentar ao paciente estímulos luminosos dispostos em uma semi-esfera e solicitar-lhe que, com a visão fixa em um ponto central, acione um pequeno dispositivo a cada vez que conseguir enxergar um destes sinais. Um software consolida o resultado de todos os pontos ativados e as respectivas identificações pelo paciente para então compor o mapa de campo visual.

Ao avaliar o resultado de uma perimetria computadorizada o oftalmologista pode identificar o grau de comprometimento visual do paciente através da localização das áreas afetadas e a intensidade em decibels (dB) da perda apresentada em relação a um campo visual de um indivíduo normal na mesma faixa etária. A perimetria computadorizada é classificada como teste funcional *gold standard* para diagnóstico de danos glaucomatosos.

Além da perimetria computadorizada padrão existem outros métodos mais recentes empregados para avaliação funcional, destacando-se os baseados na perimetria de frequência duplicada (PACZKA et al., 2001; FEUER et al., 2002; DELGADO et al., 2002; LANDERS; GOLDBERG; GRAHAM, 2003; MEDEIROS; SAMPLE; WEINREB, 2004) e na perimetria azul-amarelo (FEUER et al., 2002; PACZKA et al., 2001).

Já os testes estruturais são responsáveis pelo mapeamento das estruturas anatômicas oculares em busca de anomalias que possam ser indicadores da presença de glaucoma. Por

exemplo, um dos fatores já citados como prevalente entre os indivíduos glaucomatosos é a pressão intra-ocular elevada. Esta pressão acima dos limites normais pode ocasionar a morte de células ganglionares e a consequente perda de espessura na camada de fibras nervosas. Com isso há o aumento da área e volume da região denominada escavação do disco óptico. Disco óptico é a estrutura anatômica através da qual o nervo óptico penetra no olho, sendo também chamado de papila ou cabeça do nervo óptico. As elevações das relações numéricas entre áreas e volumes da escavação em relação ao disco óptico são importantes para o diagnóstico de glaucoma. Na figura 1.5 é possível observar-se estas estruturas.

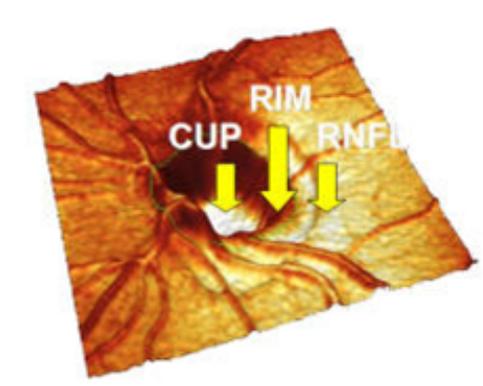


Figura 1.5: Imagem em três dimensões do fundo de olho obtida através do instrumento HRT II destacando-se o disco óptico ao centro, a área de escavação (CUP), a rima neural (RIM) e a camada de fibras nervosas (RNFL) (HEIDELBERG-ENGINEERING, 2009).

O estudo da camada de fibras nervosas pode propiciar o diagnóstico antecipado da doença, como mostrado pelo estudo de Sommer et al (SOMMER et al., 1991), que revelou que 60% dos pacientes que evoluíram para glaucoma tinham defeitos na camada de fibras nervosas seis anos antes do aparecimento de defeito de campo visual. Entretanto, a avaliação do nervo óptico e da camada de fibras nervosas, seja por observação direta ou por técnicas fotográficas, é subjetiva e requer examinador experiente (FEUER et al., 2002; HEIJL et al., 2003).

Para avaliação anatômica, destacam-se os métodos baseados na polarimetria de varredura a laser para análise da camada de fibras nervosas (DUARTE et al., 1999; PRATA et al., 2001; MOURA; MELLO; PRATA, 2003; TANITO et al., 2004), na oftalmoscopia confocal com varredura a laser para estudo do disco óptico (BARBOZA; JR, 1998; ZANGWILL et al., 2001) e na tomografia de coerência óptica (AYDIN et al., 2003) para ambas as estruturas.

No entanto, na literatura há estudos que sugerem que nenhum dos métodos atuais, quando empregados isoladamente, apresentam sensibilidade e especificidade que justifiquem o seu uso como ferramenta de rastreamento para glaucoma inicial a moderado

(SANCHEZ-GALEANA et al., 2001). Em contra-partida, outros estudos (BOWD et al., 2000; LAUANDE-PIMENTEL et al., 2001; GREANEY et al., 2002) relatam que a associação de métodos aumenta a capacidade diagnóstica de glaucoma. Pesquisa recente (MEDEIROS, 2004) revela que a combinação de resultados da aplicação dos métodos para avaliação estrutural pode fornecer informação adicional ao oftalmologista para diagnóstico do glaucoma, sugerindo que a integração dos dados clínicos, obtidos através de vários testes, ao conhecimento e experiência do médico pode melhorar o diagnóstico e até mesmo antecipá-lo, ou seja, identificar a presença do glaucoma antes que surjam danos irreversíveis à visão do paciente.

Porém, transformar conjuntos de dados em informação e descobrir suas inter relações para produzir conhecimento útil e aplicá-lo à solução de problemas em um determinado domínio são tarefas típicas da computação. Uma área da computação normalmente empregada na resolução de problemas no domínio médico é a inteligência artificial, sendo que uma sub-área de especial relevância para questões de diagnóstico médico é a aprendizagem de máquina. 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 (MITCHELL, 1999). Diversos estudos abordam o uso de algoritmos classificadores baseados em técnicas de aprendizagem para identificação de doenças.

Ao conduzir estudo sobre a aplicação de técnicas de aprendizagem de máquina no diagnóstico médico, Kononenko (KONONENKO, 2001) aponta como tendência a utilização cada vez maior pelos médicos destas técnicas como ferramentas para melhorar o desempenho da sua atividade. Estas técnicas constituem fonte adicional de informação para elevar a acurácia do diagnóstico, dado que o uso desta informação e o julgamento final continuarão sob responsabilidade do profissional. O mesmo estudo identificou como principais categorias de algoritmos de aprendizagem empregados ao longo dos anos nas tarefas de identificação da presença ou não de indícios de determinada enfermidade em dados de exames de um paciente são classificadores Bayesianos, árvores de decisão e redes neurais artificiais.

Do ponto de vista funcional, classificadores são algoritmos que recebem como entrada um vetor de características (como os valores extraídos de exames de um paciente visando ao diagnóstico de glaucoma), e apresentam como resultado a indicação de uma classe (categoria) à qual aquele vetor pode ser associado (por exemplo, glaucoma ou normal). Geralmente estes algoritmos se utilizam de uma forma de aprendizagem supervisionada, onde o programa aprende através de exemplos de treinamento positivos e negativos, com-

postos por valores para atributos selecionados que lhe são fornecidos juntamente com sua correspondente classificação. Um exemplo positivo para o treinamento de um classificador para glaucoma é um conjunto de valores de atributos como idade, PIO, volume da escavação, volume do disco óptico e espessura média da camada de fibras nervosas de um paciente com a doença, enquanto que um exemplo negativo seriam valores para os mesmos parâmetros, porém referentes a um paciente normal (sem a doença). O treinamento é repetido diversas vezes com o fornecimento de vários exemplos juntamente com a classificação positivo ou negativo realizada por um médico, até que o conceito (identificação da doença através do exame em questão) possa ter sido devidamente aprendido pelo sistema. Passa-se então a utilizar o algoritmo com novos dados, porém não previamente classificados por um especialista humano, para que o sistema então faça sua classificação automaticamente.

Um exemplo de categorias de algoritmos amplamente empregados em tarefas de classificação são as redes neurais artificiais. *Multilayer Perceptrons* (MLP) são um tipo de rede neural alimentada para frente, composta por um ou mais nós computacionais (neurônios) na camada de entrada, um ou mais neurônios na camada de saída, e uma ou mais camadas ocultas intermediárias de nós computacionais.

Neste tipo de rede o sinal de entrada é representado por atributos selecionados para a tarefa de classificação (como os parâmetros citados como exemplo para aprendizagem de um classificador para glaucoma). Este sinal, assim como em uma célula nervosa humana, propaga-se para a frente, camada a camada, a partir da entrada até a saída, representada pelas classes que se deseja obter (no exemplo, glaucoma e normal). Uma vez que uma saída apresentada pela rede não seja correta frente ao exemplo de treinamento fornecido como sinal de entrada, um erro de classificação é identificado. Através de um algoritmo de retro propagação destes erros (*error backpropagation*), a rede MLP pode ser eficientemente treinada para melhorar sua acurácia na tarefa de classificação.

A literatura apresenta resultados de estudos da aplicação de classificadores no diagnóstico de glaucoma ao longo dos últimos anos. Goldbaum (GOLDBAUM; AL, 2002) avaliou a eficácia de determinados classificadores na tarefa de identificação precoce de perda visual baseando-se em dados obtidos através do exame de perimetria computadorizada padrão. Hothorn e Lausen (HOTHORN; LAUSEN, 2003), Huang et al (HUANG; CHEN; HUANG, 2007) e Burgansky-Eliash et al (BURGANSKY-ELIASH, 2005) testaram classificadores sobre dados de espessura da camada de fibras nervosas. Bowd et al (BOWD et al., 2008) recentemente utilizaram classificadores em tarefas de diagnóstico de glaucoma baseado na integração de dados anatômicos e funcionais.

Todavia, para que algoritmos classificadores possam ser eficazes, a etapa de treinamento desempenha um papel de vital importância. Para que o treinamento possa ser realizado adequadamente, conjuntos de dados de qualidade devem ser empregados como exemplos. Dados de qualidade significam dados representativos de populações reais, e disponíveis em quantidade suficiente para que possam expressar a maior variedade de casos possível.

No entanto, uma questão relevante para a condução de estudos de métodos diagnósticos de glaucoma são as dificuldades para coleta de dados, tais como: o recrutamento e seleção de sujeitos de acordo com critérios de inclusão e exclusão do estudo, a disponibilidade dos equipamentos tecnológicos, materiais e profissionais de saúde para condução dos testes nos indivíduos ao longo do tempo e, principalmente, a manutenção da aderência dos pacientes ao programa, dificultada pela duração do estudo e pelo perfil sócio-econômico da população avaliada.

Diante destes problemas, normalmente os estudos são conduzidos com poucos indivíduos, o que pode constituir sensível restrição à eficácia de classificadores baseados em aprendizagem, visto que necessitam de conjuntos de dados com muitos elementos para que consigam desempenhar a tarefa de classificação com maiores possibilidades de sucesso.

Contudo, métodos estatísticos podem ser empregados para gerar amostras populacionais artificiais de forma a permitir a simulação computacional de estudos em glaucoma, e empregá-las no treinamento e testes de classificadores. Como benefícios desta abordagem pode-se citar a economia de tempo, dado que a escala cronológica artificial pode ser bem inferior à real, de recursos financeiros, visto que o uso de equipamentos e consumo de materiais pode ser reduzido, e sociais, uma vez que é possível conduzir experimentos sem a necessidade de coleta de dados em seres humanos em larga escala.

1.1 Objetivos da Pesquisa

Diante do exposto, o objetivo deste trabalho de pesquisa é criar um gerador artificial de dados epidemiológicos, clínicos, anatômicos e funcionais de pacientes oftalmológicos de acordo com as respectivas distribuições normais definidas a priori para cada parâmetro selecionado. Este gerador é capaz de gerar dados em quantidade e qualidade suficientes para possibilitar o treinamento de classificadores de aprendizagem de máquina, visando assim contribuir para o desenvolvimento de novos métodos que possam elevar a sensibilidade e especificidade no diagnóstico de glaucoma.

1.2 Metodologia da Pesquisa

A metodologia proposta para solução dos problemas identificados por este trabalho partiu do desenvolvimento de um método de Monte Carlo para produzir artificialmente as populações amostrais, visando o treinamento de classificadores de glaucoma. Monte Carlo tem sido amplamente utilizado em pesquisas científicas para simulação e geração de dados a partir de distribuições de probabilidade. O método desenvolvido utiliza distribuições de probabilidade e valores normais de dados epidemiológicos e clínicos, além de parâmetros obtidos pelos testes diagnósticos funcionais e estruturais, e assim gera artificialmente dados representativos de populações reais.

Na etapa seguinte, treinou-se e testou-se classificadores baseados em Naive-Bayes, Árvores de Decisão e Redes Neurais Artificiais com base em dados artificiais e reais. A avaliação da eficácia do gerador foi realizada através do desempenho dos classificadores no diagnóstico de glaucoma treinados com os dados artificiais. Tal avaliação foi conduzida através do cálculo da acurácia total, sensibilidade, especificidade e áreas sob curvas ROC (*Receiver Operating Characteristic*) dos classificadores.

Médicos oftalmologistas especialistas em glaucoma foram os avaliadores qualitativos da capacidade do gerador em produzir artificialmente dados que simulem indivíduos reais. Foram igualmente os responsáveis pela classificação quanto à presença de glaucoma na população de indivíduos reais utilizados para validação dos algoritmos desenvolvidos neste trabalho.

1.3 Estrutura do trabalho

No capítulo 2 é apresentado o glaucoma e seus métodos diagnósticos atuais. Os classificadores baseados em algoritmos de aprendizagem de máquina utilizados neste trabalho são descritos no capítulo 3.

Os métodos desenvolvidos visando sua resolução - a geração de dados artificiais para uso nas tarefas de classificação e a construção dos classificadores - são descritos no capítulo 4, enquanto que a avaliação dos resultados obtidos através da sua aplicação são apresentados no capítulo 5.

As considerações finais encerram o texto no capítulo 6.

2 Diagnóstico de Glaucoma

Apesar dos indivíduos com glaucoma geralmente apresentarem níveis estatisticamente elevados de pressão intra-ocular (MEDEIROS; WEINREB, 2002), a medida da pressão (ou tonometria) não pode ser empregada como método diagnóstico isolado, uma vez que nem todos os indivíduos com hipertensão ocular evoluem para glaucoma (KASS et al., 2002), enquanto vários indivíduos com PIO estatisticamente normal são portadores dessa doença (ANDERSON; DRANCE; SCHULZER, 2001). Além disso, técnicas de tonometria são passíveis de artefatos induzidos pela espessura da córnea, doenças oculares e pelo tipo de tonômetro utilizado (BRANDT et al., 2001; WHITACRE; STEIN, 1993). Outro fator importante é que a tonometria isolada representa a medida de um momento e não considera a variação circadiana da PIO, ou seja, as alterações fisiológicas que a pressão sofre ao longo de um dia (TAVARES et al., 2003).

O glaucoma é diagnosticado pelo oftalmologista especialista na doença, que se baseia na associação de dados como o histórico clínico do paciente e a medida da pressão intra-ocular, aos resultados de exames funcionais, que avaliam o campo visual do paciente para procurar identificar indícios de redução da capacidade visual, e aos exames estruturais, que buscam por alterações anatômicas nas estruturas oculares, como a perda acentuada da camada de fibras nervosas, que é um dos primeiros indícios de danos típicos do glaucoma. Estes resultados são comparados estatisticamente aos dados de pacientes normais, ou seja, sem glaucoma, na mesma faixa etária. Tal relação com bancos de dados normativos por faixa etária é importante para auxiliar o diagnóstico, uma vez que não se pode comparar dados obtidos para um indivíduo de 40 anos aos mesmos dados de um indivíduo com 50 anos, por exemplo. Isto se deve ao fato de que à medida que a idade avança ocorre a morte natural de células ganglionares, neurônios cujos axônios formam os feixes que constituem as fibras nervosas. Como consequência da falência destas células ao longo dos anos, ocorre a redução de espessura da camada de fibras nervosas. Como estrutura e função são relacionados, a redução da quantidade de células ganglionares e da camada de fibras nervosas prejudicam a transmissão das informações dos fotorreceptores ao cérebro,

ocasionando a redução gradativa de sensibilidade no campo visual dos indivíduos ao longo dos anos.

Entre os testes para avaliação de campo visual existentes, a perimetria computadorizada padrão ou *Standard Automated Perimetry* (SAP) é considerado pelos oftalmologistas o padrão-ouro (*gold standard*) para realizar o diagnóstico funcional e avaliar a progressão do glaucoma (LIU et al., 2003). A evolução deste exame, graças à análise computadorizada dos dados, permitiu uma maior reprodutibilidade dos resultados e comparações estatísticas com um extenso banco de dados normativo embutido no software do equipamento (DIAS; LIMA, 1998). Devido a isso, a perimetria computadorizada padrão é amplamente utilizada nos protocolos clínicos de glaucoma.

O exame de campo visual padrão consiste basicamente em apresentar ao paciente estímulos luminosos e solicitar ao mesmo que, com a visão fixa em um ponto central do equipamento, acione um botão a cada vez que conseguir perceber um destes sinais. Como estes estímulos são apresentados na forma de uma luz branca, a perimetria computadorizada padrão é também chamada de perimetria acromática. O estímulo apresentado possui intensidade variável dentro de uma faixa determinada para cada ponto, iniciando-se em um valor base pré-determinado para o ponto a ser avaliado. Em seguida, esta intensidade de luz é gradualmente elevada até que haja a ativação do botão pelo paciente, indicando o momento que a luz foi identificada (e portanto o limiar de intensidade luminosa percebida), ou então a intensidade atinja o valor máximo da faixa, evidenciando que o paciente não percebeu o estímulo apresentado. Os valores das intensidades dos estímulos percebidos em cada ponto são denominados sensibilidade. Estes valores de sensibilidade podem ser obtidos em números absolutos de intensidade luminosa por área, expressas em apostlib ($1 \text{ asb} = 1/\pi \text{ candela}/m^2$), ou em valores relativos aos limiares de teste (valores iniciais das faixas de intensidade de apresentação por ponto), expressos neste caso em decibels (dB). Como a variabilidade de estímulos é grande, variando de menos de 1 asb até 10000 asb, normalmente as sensibilidades são expressas em valores relativos (AZUARA-BLANCO; COSTA; WILSON, 2002). Valores próximos a zero indicam áreas defeituosas do campo visual, ou seja, áreas de baixa sensibilidade, enquanto que valores significativamente maiores do que zero indicam áreas normais.

A quantidade total de medidas de sensibilidade relativa obtidos em um exame de campo visual depende do número total de pontos avaliados, que são determinados pela estratégia de teste empregada. Na estratégia denominada Central 30-2 (C30-2), os estímulos são apresentados ao longo de 30 graus em todas as direções ao redor do ponto de fixação e alinhados a cada 6 graus dos meridianos vertical e horizontal. Nesta estratégia são testados

76 pontos, sendo que 2 deles correspondem à área da retina desprovida de fotorreceptores denominada ponto cego (*blind spot*), região na qual os estímulos luminosos não são percebidos por incidirem diretamente sobre o disco óptico. Outra estratégia normalmente empregada nos exames de campo visual padrão é a Central 24-2 (C24-2) onde os estímulos são apresentados ao longo de 24 graus na região temporal e 30 graus na região nasal. Nesta estratégia são testados 54 pontos (sendo 2 correspondentes ao ponto cego). Com esta abordagem, reduz-se o tempo gasto no teste sem redução significativa da capacidade de identificação de danos no campo visual. Esta característica é importante pois quanto maior o tempo empregado em um teste, maior a probabilidade do resultado ser comprometido pela fadiga do paciente.

Cada estratégia de teste é associada a um algoritmo para determinação dos limiares de sensibilidade. Os mais empregados atualmente são o *Full Threshold* e o *SITA Standard*. O algoritmo *Full Threshold* apresenta boa acurácia quando comparado aos demais algoritmos, pois promove a estratégia de teste e reteste de pontos, todavia consome mais tempo para ser executado. No entanto, o algoritmo *SITA Standard*, desenvolvido recentemente, apresenta a mesma acurácia do *Full Threshold* mas com tempo de teste menor, sendo portanto cada vez mais empregado.

Após a coleta dos dados das sensibilidades relativas, um software estatístico denominado STATPAC realiza comparações entre as sensibilidades relativas ponto a ponto e os respectivos valores normais presentes em um banco de dados normativo relacionado à idade. Estes valores normais, embutidos no software, são representados por médias e desvios padrões de sensibilidade em cada ponto avaliado. Tais valores normativos são empregados pelo instrumento para construir os gráficos *total deviation* e *pattern deviation*, tanto em valores relativos, expressos em dB, quanto em faixas de probabilidades de normalidade.

O gráfico *total deviation* apresenta os valores em dB das diferenças ponto a ponto entre o valor de sensibilidade medido e o esperado de acordo com o banco de dados normativo, enquanto que os valores do gráfico *pattern deviation* são obtidos através de um processo de normalização aplicado sobre os valores do gráfico *total deviation*.

Os índices globais *Mean Deviation* (MD) e *Pattern Standard Deviation* (PSD) são obtidos através de cálculos sobre os valores em decibels dos gráficos *total deviation* e *pattern deviation*, respectivamente. O MD representa o desvio médio de sensibilidade em relação à população normal, enquanto o PSD representa o desvio padrão apresentado pelos dados normalizados. A principal função do MD é indicar eventual redução geral de sensibili-

dade. Já a função do índice PSD é oferecer ao especialista a medida de variabilidade da sensibilidade, indicando a possível existência de defeitos localizados no campo visual.

Outro índice que pode ser obtido é o GHT ou *Glaucoma Hemifield Test*. Este índice é calculado através de comparações de sensibilidade entre 5 áreas pré-determinadas do hemisfério superior do campo visual e as respectivas áreas simétricas no hemisfério inferior. O objetivo é identificar alterações significativas entre os valores nestas regiões, uma vez que os danos glaucomatosos são geralmente assimétricos, com um hemisfério sendo mais comprometido que o outro. O GHT pode apresentar 5 indicações: '*within normal limits*', '*borderline*', '*outside normal limits*', '*general reduction of sensitivity*' ou ainda '*abnormally high sensitivity*', indicando respectivamente resultado normal, limítrofe, fora dos valores normais (positivo para glaucoma), redução geral de sensibilidade e sensibilidade elevada de modo anormal.

Através do resultado de um exame de campo visual (figura 2.1), pode-se observar na região central superior o gráfico de sensibilidades relativas em dB (primeiro gráfico de cima para baixo na figura). Ao lado direito deste gráfico de sensibilidades relativas é exibido um gráfico em tons de cinza, o qual apresenta em tons mais escuros as áreas com maior redução de sensibilidade. Abaixo do gráfico de sensibilidades relativas, ao lado esquerdo, encontram-se os dois gráficos *total deviation* (um com valores em decibels e outro em faixas de probabilidades de normalidade). Ao lado direito dos gráficos *total deviation* estão os gráficos de *pattern deviation*. Nesta figura apresentada como exemplo, pode-se identificar redução significativa de sensibilidade em pontos do quadrante superior direito (com até 23 dB de redução frente a um indivíduo normal pelo gráfico *total deviation*), redução geral de sensibilidade indicada por MD de -3,79 dB além do PSD de 4,35 dB, sugerindo a presença de defeitos localizados. Além disso, o resultado '*outside normal limits*' apresentado pelo GHT indica maior probabilidade destes danos serem típicos de um olho glaucomatoso.

Entre os instrumentos de perimetria computadorizada padrão, o modelo Humphrey HFA-II (Carl Zeiss Meditec, Inc.) tem sido um dos mais utilizados.

Entretanto, evidências crescentes sugerem que a definição de dano glaucomatoso precoce através de alterações da perimetria acromática está longe do ideal. Estudos histológicos demonstraram que, para haver repercussão no campo visual, é necessário dano a aproximadamente 30-50% do número de fibras nervosas do nervo óptico (BRENTON; PHELPS, 1986). Como consequência deste achado, houve grande interesse em determinar os sinais precoces de glaucoma no nervo óptico, uma vez que este pode mostrar alterações

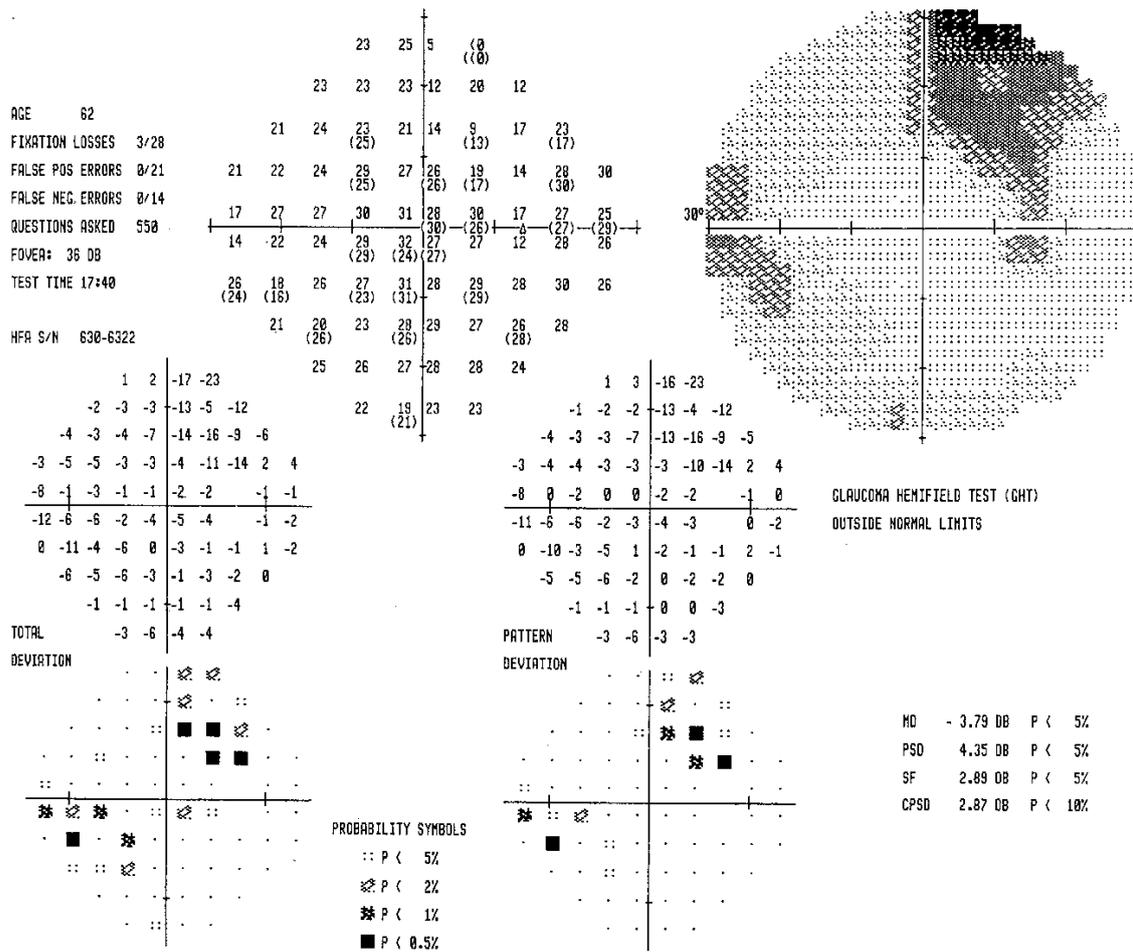


Figura 2.1: Resultado de perimetria computadorizada indicando defeitos no exame de campo visual (AZUARA-BLANCO; COSTA; WILSON, 2002).

antes da detecção do defeito no campo visual (QUIGLEY et al., 1983).

A tomografia de coerência óptica (OCT) é um método de imagem anatômico baseado na interferometria luminosa de baixa coerência. A interferometria se baseia na interferência de ondas, construtiva quando estas estão em fase, destrutiva quando estão fora de fase (180°) e todo o espectro intermediário de construção ou destruição parcial. Utilizando-se de um interferômetro de Michelson (JAFPE; CAPRIOLI, 2004) e uma fonte luminosa de diodo super luminescente próxima ao espectro infravermelho (810 nm), um feixe luminoso é dividido em dois, sendo um - o de referência - desviado para um espelho de referência, situado a uma distância variável e conhecida, enquanto o outro é direcionado ao tecido a ser examinado.

As estruturas intra-oculares são transparentes a este comprimento de onda, fazendo com que o feixe luminoso proveniente da fonte atravesse os tecidos. A cada interface óptica, ou seja, meios ópticos de diferentes índices refracionais, parte do feixe luminoso é

refletido, sendo analisado pelo fotodetector do aparelho em relação ao feixe de referência. As imagens são então reconstruídas de acordo com a proporção de interferência, e coloridas de acordo com o grau de reflectividade (vermelho a branco para estruturas bastante refletivas e azul a preto para as pouco refletivas).

Com esta tecnologia, a resolução da imagem das estruturas observadas alcança o nível de 7 a 8 μm no OCT. O protocolo utilizado para uso da OCT na análise anatômica do paciente glaucomatoso se divide em 2 etapas: a primeira consiste em secções transversais do disco óptico e da região peripapilar (ao redor do disco) em diferentes direções (0° , 30° , 60° , 90° , 120° e 150°), passando pelo centro do primeiro gerando-se seis imagens de 100 pixels na horizontal por 250 pixels na vertical, correspondendo a uma secção radial de 6,75 mm do disco óptico. A segunda consiste numa varredura circular ao redor do disco óptico na região peripapilar em duas linhas diferentes, ambas iniciando num ponto imaginário na região mais temporal do disco óptico. A primeira situa-se a 1,5 diâmetro de disco, ou seja, a meio diâmetro de disco da margem mais externa da rima neural (a borda do disco óptico) e a segunda a 2 diâmetros de disco, ou a 1 diâmetro de disco da margem mais externa da rima neural. As duas imagens de 100 pixels na horizontal por 250 pixels na vertical são apresentadas ao examinador como uma retificação do escaneamento no sentido anti-horário, ou seja, seguindo-se a imagem gerada da esquerda para a direita temos a secção transversal da retina peripapilar das regiões temporal, inferior, nasal, superior e temporal novamente. Na figura 2.2 podem ser identificadas as 9 regiões medidas pelo OCT em torno do centro do disco óptico.

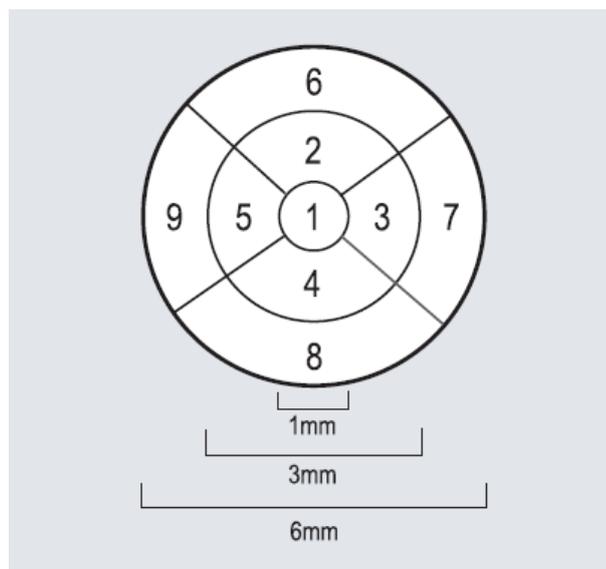


Figura 2.2: Segmentos da retina que são avaliados pelo OCT (ZEISS, 2009).

Existem atualmente dois tipos de tomógrafos: os de domínio temporal ou *Time Do-*

main OCT (TD-OCT), pioneiros e portanto bem mais disseminados, e os de domínio espectral ou *Spectral Domain OCT* (SD-OCT), que empregam o estado da arte desta tecnologia. Comercialmente, um dos mais utilizados modelos de TD-OCT é o Stratus OCT (Carl Zeiss Meditec, Inc.) enquanto que o modelo Cirrus HD-OCT (*High Definition OCT*) é a versão de SD-OCT da mesma empresa. Na tomografia de coerência óptica de domínio temporal ao todo são realizadas 6 varreduras transversais e 256 varreduras axiais, enquanto que na tomografia no domínio espectral são realizadas 200 varreduras lineares verticais e 200 horizontais, distribuídas em um quadrado de 6 mm de lado com centro no disco óptico. A figura 2.3 apresenta a comparação entre as duas estratégias de varredura. Esta grande diferença entre a quantidade de medições proporcionadas pelo uso das tecnologias temporal e espectral é responsável pela alta definição das imagens produzidas pelo instrumento Cirrus HD-OCT. Além disso, o Cirrus HD-OCT é 50 vezes mais veloz ao realizar seu processo de varredura do que o Stratus OCT.

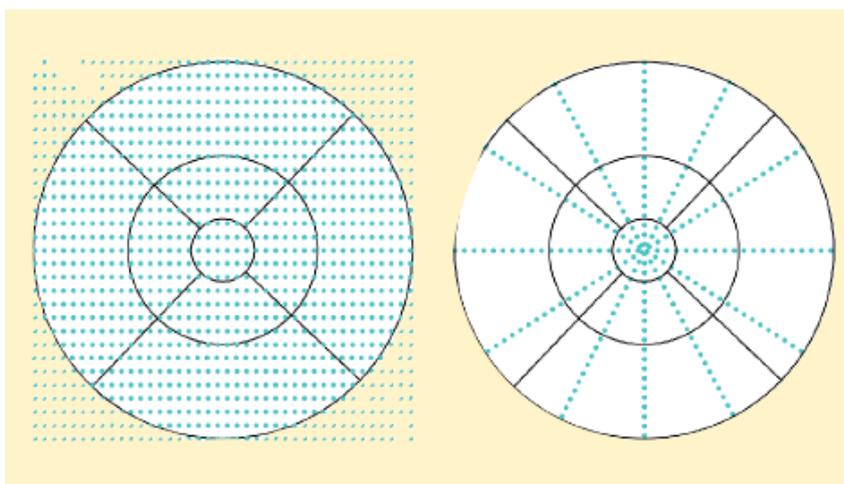


Figura 2.3: Comparação entre as estratégias de varredura do Cirrus HD-OCT à esquerda e Stratus OCT à direita (ZEISS, 2009).

Na figura 2.4 é apresentado um resultado de tomografia de coerência óptica de alta definição (Cirrus HD-OCT) destacando de cima para baixo o gráfico de espessura média em microns da camada de fibras nervosas total (*Average Thickness*), as espessuras médias nos setores temporal, superior, nasal e inferior (*Quadrants*), espessuras médias a cada hora de relógio (*Clock Hours*) e o gráfico da espessura da camada de fibras nervosas (*RNFL TSNIT Normative Data*). Este último gráfico, obtido através dos valores de espessura em 256 pontos ao longo de uma circunferência com centro no disco óptico, apresenta as três curvas normativas das probabilidades de normalidade (preenchidas em vermelho, amarelo e verde para indicar faixas de valores encontrados em menos de 1%, 5% e 95% da população normal, respectivamente), as quais dependem da idade do paciente e são construídas de acordo com o banco de dados normativo do software embutido no instrumento.

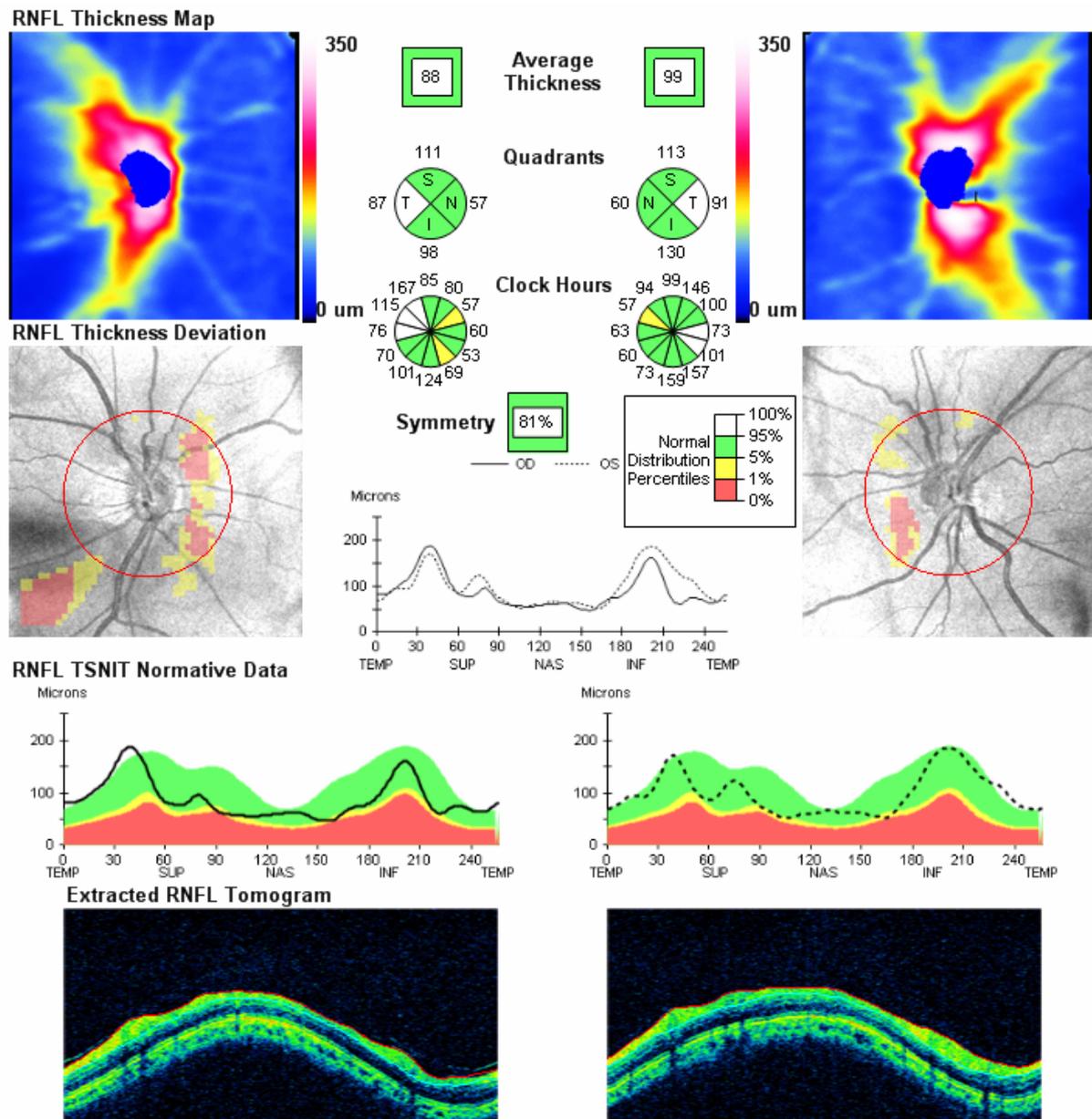


Figura 2.4: Resultado de tomografia de coerência óptica de alta definição (Cirrus HD-OCT) de um indivíduo normal voluntário.

O OCT permite obter medidas objetivas e reprodutíveis do disco óptico e da camada de fibras nervosas que não são influenciadas por erro refracional, diâmetro axial e cataratas nucleares iniciais (SCHUMAN et al., 1995). Em estudos iniciais, mostrou diferença significativa quando utilizado para comparar pacientes glaucomatosos e indivíduos normais e hipertensos oculares (BOWD et al., 2000).

2.1 Outros Métodos Diagnósticos

No presente trabalho utilizou-se dados de tonometria, perimetria automatizada padrão e tomografia de coerência óptica. No entanto existem atualmente outros métodos diagnósticos como a perimetria azul-amarelo ou *Short Wavelength Perimetry* (SWAP) e a perimetria de frequência duplicada ou *Frequency Doubling Perimetry* (FDT), para avaliação de campo visual, e a polarimetria de varredura a laser (GDx-VCC) e a oftalmoscopia confocal com varredura a laser (HRT II) para avaliação anatômica.

2.1.1 Métodos Funcionais

A perimetria azul-amarelo (SWAP) é um exame de campo visual que, ao contrário da perimetria padrão (SAP), que utiliza estímulos acromáticos, emprega cores na avaliação do paciente com glaucoma (AYDIN et al., 2003). A perimetria azul-amarelo avalia uma subpopulação de células ganglionares denominadas células parvocelulares (células P). Sensíveis a cores, as células P são o tipo mais abundante de células ganglionares. Estudos mostraram que os pacientes com glaucoma apresentam uma maior incidência de defeitos de cores no espectro azul-amarelo (PACHECO-CUTILLAS; SAHRAIE; EDGAR, 1999). A perimetria azul-amarelo é realizada com estímulo azul-violeta em fundo amarelo.

Estudos demonstraram que os déficits funcionais mostrados no SWAP em pacientes com hipertensão ocular e glaucoma são detectados anos antes da perimetria acromática convencional (POLO; LARROSA; PABLO, 1999). Johnson et al (JOHNSON; ADAMS; LEWIS, 1993) seguiram longitudinalmente uma população de 38 hipertensos oculares durante 5 anos. No início do estudo, todos os pacientes apresentaram perimetria acromática normal e 9 mostraram defeito no SWAP. Destes 9 pacientes, 5 vieram a desenvolver, ao longo do seguimento, defeitos de campo visual na perimetria acromática em locais correspondentes aos defeitos no SWAP. Nenhum dos pacientes com SWAP normal no início do estudo veio a desenvolver defeito na perimetria acromática. Este estudo confirmou a capacidade do SWAP de detectar defeitos de campo visual 3 a 4 anos antes da perimetria acromática. Na

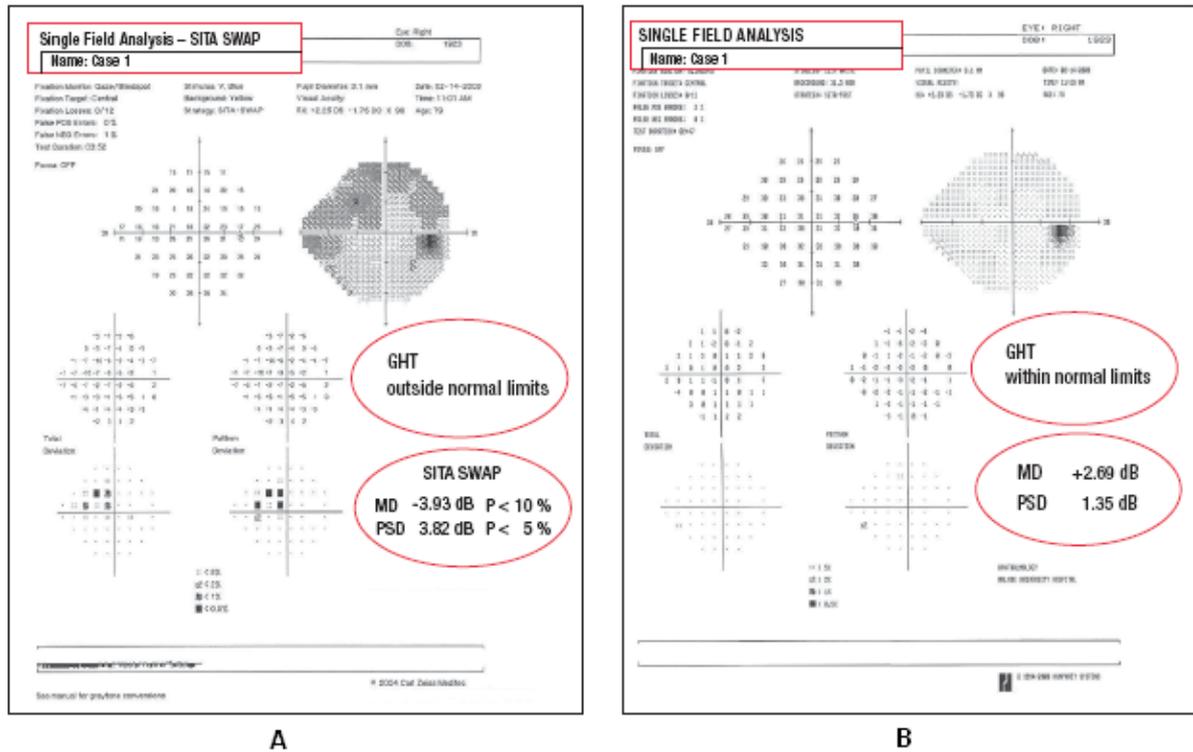


Figura 2.5: Exemplos de resultado de SWAP (A) e SAP (B) para um mesmo paciente (ZEISS, 2009).

figura 2.5 pode-se observar resultados de SWAP (A) e SAP (B) para um mesmo paciente lado a lado. Na imagem A tem-se o resultado do SWAP indicando que o paciente está fora da faixa de normalidade ($GHT = \textit{outside normal limits}$) para a idade, enquanto que na imagem B pode-se notar o resultado da perimetria acromática indicando que os valores estão dentro da faixa de normalidade ($GHT = \textit{'within normal limits'}$). A diferença entre os resultados também pode ser notada nos respectivos valores de MD e PSD.

Já a perimetria de frequência duplicada (FDT) avalia as células koniocelulares, que são outro tipo de células ganglionares. Como as células koniocelulares são mais sensíveis a movimentos, o instrumento apresenta estímulos formados por faixas pretas e brancas alternando-se num padrão senoidal, em frequência espacial de 0,25 Hz e temporal de 25 Hz. Nestas condições, o paciente percebe o dobro do número de faixas apresentadas, fenômeno conhecido como ilusão de duplicação. O contraste entre as faixas aumenta e diminui para determinar o menor contraste (limiar) que permita a percepção da duplicação (MOK; LEE, 2000).

Na figura 2.6 tem-se um exemplo de resultado de FDT indicando os limiares (*threshold*) de sensibilidade (dB) em cada ponto do campo visual avaliado, bem como os gráficos de desvio total (*total deviation*) e desvio do padrão (*pattern deviation*), ambos

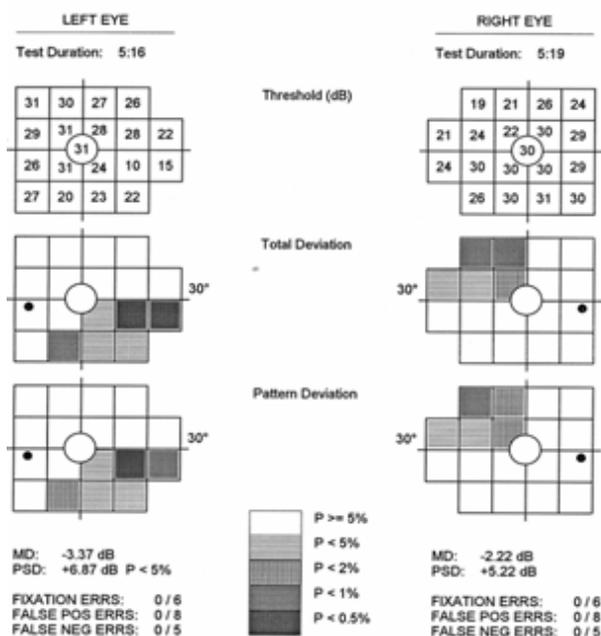


Figura 2.6: Exemplo de resultado de FDT (ZEISS, 2009).

expressos graficamente em probabilidades de normalidade, além do MD (*Mean Deviation*) e PSD (*Pattern Standard Deviation*).

Medeiros et al (MEDEIROS; SAMPLE; WEINREB, 2004), analisando 105 suspeitos de glaucoma por 41 meses, revelaram que em 59% daqueles que desenvolveram glaucoma, as alterações campimétricas do FDT precederam as da SAP em 4 anos. Constataram também que o desenvolvimento de perda do campo visual medida pela SAP ocorreu nas regiões previamente alteradas ao FDT. Assim, esses autores confirmaram que anormalidades funcionais detectadas pelo FDT podem ser preditivas da instalação e localização de defeitos campimétricos na SAP em suspeitos de glaucoma.

2.1.2 Métodos Estruturais

Já entre os métodos para avaliação estrutural, o laser confocal polarizado, cuja última versão é comercialmente conhecida como GDx-VCC, é um oftalmoscópio confocal associado a um modulador de polarização e uma unidade detectora de polarização, que obtém e grava imagens em alta resolução do fundo do olho. Utiliza como fonte luminosa um laser polarizado de diodo, com baixa potência e de espectro próximo ao infravermelho (comprimento de onda = 780 nm) (WEINREB et al., 1998; GREENFIELD et al., 2002).

O instrumento emite um feixe de laser que passa por um polarizador interno de forma a obter luz polarizada, que é defletida por um complexo sistema de cristais na unidade de varredura nas direções horizontal e vertical (eixos x e y). O uso da polarimetria baseia-se

na propriedade birrefringente da camada de fibras nervosas, que tem suas fibras dispostas em um arranjo espacial organizado, caracterizado por microtúbulos que se dispõem paralelamente (figura 2.7). Isto permite que parte da luz polarizada sofra um atraso, ou seja, uma alteração de velocidade, à medida que cruza a camada de fibras nervosas. Esta alteração no estado da polarização da luz é proporcional ao número de microtúbulos atravessados pelos raios luminosos e pode ser quantificada. Desta forma, o atraso sofrido pela luz é diretamente proporcional à espessura da camada de fibras nervosas (WEINREB et al., 1998). Antes de deixar o aparelho, o raio passa por um modulador de polarização que visa compensar a birrefringência exercida pelo segmento anterior do olho (córnea e cristalino).

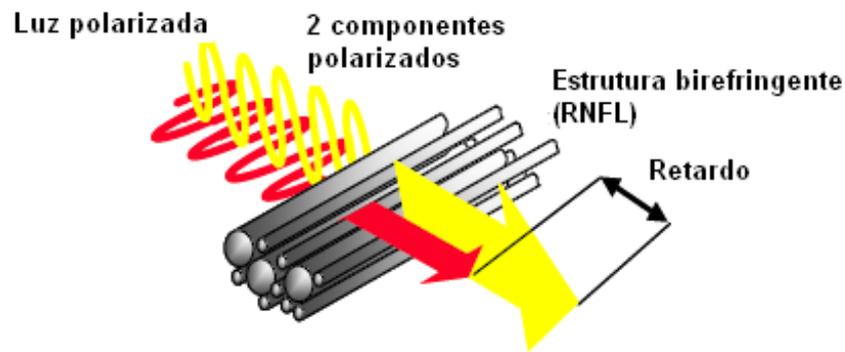


Figura 2.7: Princípio de funcionamento da polarimetria de varredura a laser (ZEISS, 2009).

Para cada medida, o GDx-VCC gera duas imagens: uma de reflectância e outra de retardo (figura 2.8). A imagem de reflectância registra a reflexão da luz diretamente sobre a superfície da retina. A imagem de retardo é um mapa de valores de retardo convertidos para espessura da camada de fibras nervosas através de um fator de conversão de $0,67 \text{ nm}/\mu\text{m}$. Na imagem de retardo, áreas em azul e verde indicam regiões de menor espessura, enquanto áreas em amarelo e vermelho indicam maiores espessuras.

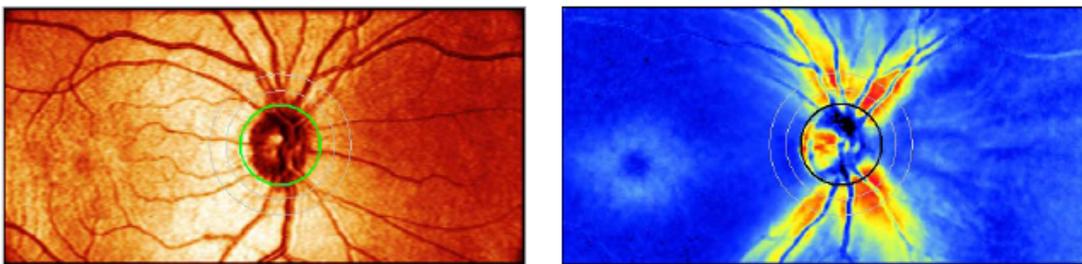


Figura 2.8: Imagens de reflectância (esquerda) e retardo (direita) obtidas através de GDx-VCC (ZEISS, 2009).

A oftalmoscopia confocal com varredura a laser, realizada pelo HRT II, utiliza um

laser de diodo de 670 nm de comprimento de onda, projetado para a retina, por um sistema confocal. Esse feixe move-se pela retina num padrão de varredura e uma pequena abertura é colocada à frente do fotodetector para eliminar a difusão de luz. A abertura é conjugada com um foco de laser, o que permite que a imagem obtida seja confocal (SANCHEZ-GALEANA et al., 2001), (GREANEY et al., 2002).

A profundidade de foco do HRT II varia de 0,5 a 4,0 mm. Com essa variação de profundidade, o HRT II faz até 64 varreduras consecutivas em três dimensões (x, y e z). O primeiro plano de imagem refere-se aos vasos retinianos e o último à porção mais profunda da área de escavação.

Desde o início de seu emprego na avaliação do glaucoma, o HRT tem demonstrado boa acurácia e reprodutibilidade, apresentando excelente correlação tanto intra quanto inter-observadores (SUNG; BHAN; VERNON, 2002). O HRT apresenta o índice de Classificação de Regressão de Moorfields, que utiliza áreas globais e setoriais do anel neuro-retiniano corrigidas para as áreas globais e setoriais da cabeça do nervo óptico para produzir três classificações: pertencente aos limites normais (*normal*), limítrofe (*borderline*) e fora dos limites normais (*outside normal limits*).

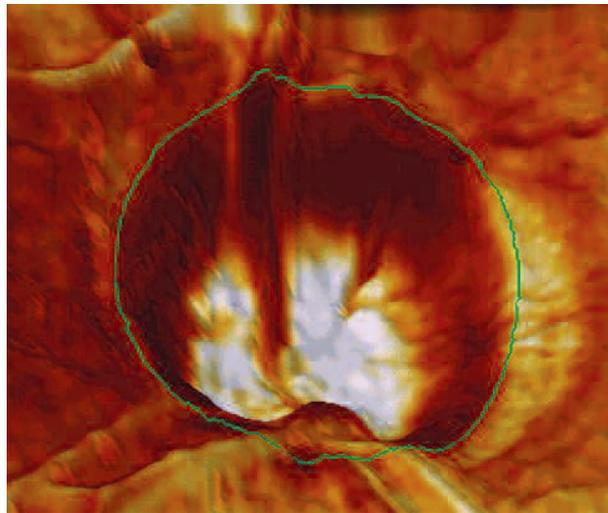


Figura 2.9: Exemplo de imagem em 3 dimensões de indivíduo com glaucoma avançado obtida através do instrumento HRT II (HEIDELBERG-ENGINEERING, 2009).

As figuras 2.9 e 2.10 apresentam exemplos de imagens obtidas através do HRT II. Na figura 2.9 tem-se uma imagem em 3 dimensões de indivíduo com glaucoma avançado. Nela pode-se observar a profundidade da escavação, inclinação abrupta das paredes e a redução de tecido na rima neural. Já na figura 2.10 pode-se observar um exemplo de imagem de um olho glaucomatoso no qual a área da rima neuro retinal em alguns segmentos do disco óptico são exibidos com indicações de 'pertencente aos limites normais'

ou '*within normal limits*' (marca verde), outros como 'limítrofes' ou '*borderline*' (ponto de exclamação amarelo), e um como 'fora dos limites normais' ou '*outside normal limits*' (X vermelho) Estas indicações são obtidas através do cálculo do índice de Classificação de Regressão de Moorfields para cada segmento.

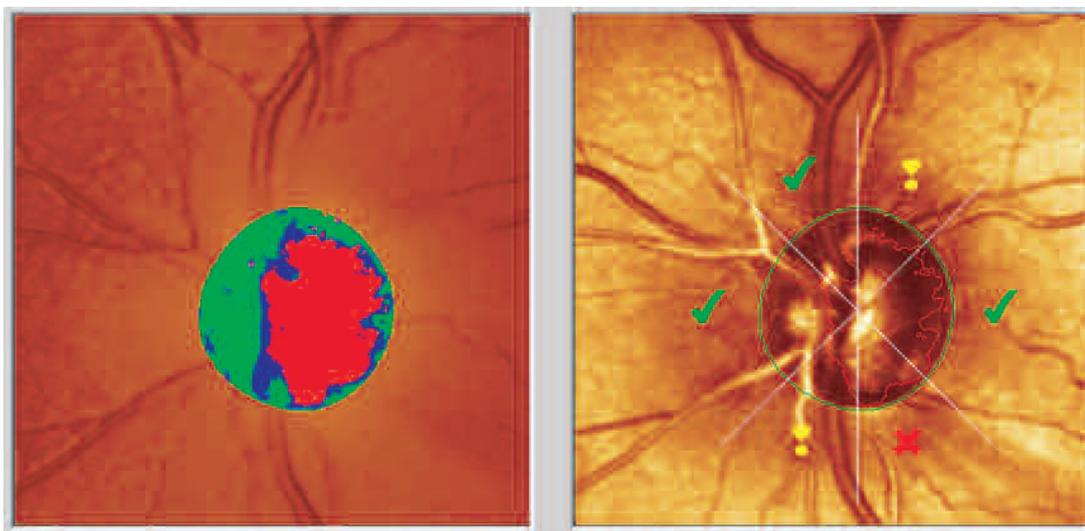


Figura 2.10: Exemplo de imagem de um olho glaucomatoso com indicações de segmentos de acordo com o índice de Classificação de Regressão de Moorfields (HEIDELBERG-ENGINEERING, 2009).

Chauhan et al (CHAUHAN et al., 2001), estudando 77 pacientes glaucomatosos por um período médio de 5,5 anos, reportaram que 40% tiveram diagnóstico de progressão da doença evidenciado somente pelo HRT, 4% somente pelo CV e 29% por ambas as tecnologias. Desses 29%, 45% tiveram progressão demonstrada primeiramente pelo HRT. Esses autores concluem que alterações da cabeça do nervo óptico são mais frequentes e mais precocemente determinadas pelo HRT do que pela perimetria acromática convencional.

2.2 Interpretação de testes diagnósticos

Duas questões importantes para a avaliação da eficácia de um teste para diagnóstico médico são a capacidade deste teste em identificar a presença da doença em um paciente que efetivamente a apresenta (indicado pelo índice de sensibilidade) e a capacidade de identificação de pacientes normais (avaliado pelo índice de especificidade). Para calcular-se estes índices utiliza-se uma tabela 2x2 (TAPE, 2006) como apresentado na tabela 2.1.

Tabela 2.1: Modelo de tabela 2x2 para interpretação de testes diagnósticos

Resultado	Presença da doença	Ausência da doença	Totais
Positivo	Verdadeiro positivo	Falso positivo	Total positivo
Negativo	Falso negativo	Verdadeiro negativo	Total negativo
Totais	Total de pacientes com a doença	Total de pacientes sem a doença	Total de pacientes testados

Caso um resultado de um exame diagnóstico seja positivo, indicando a provável existência da doença que está sendo investigada, e o paciente estiver com a moléstia presente em seu organismo, define-se que o teste produziu um resultado verdadeiro positivo (VP). Entretanto, se a doença não estiver efetivamente presente no paciente, o teste produziu um resultado falso positivo (FP). Analogamente, se o resultado do exame foi negativo, indicando a inexistência da doença em um paciente onde esta efetivamente está ausente, o teste gerou um resultado verdadeiro negativo (VN). Porém, caso a doença esteja presente, o teste produziu um resultado falso negativo (FN).

Ao aplicar-se o exame diagnóstico em diversos pacientes e submeter os resultados à avaliação de um médico especialista, que os classificará entre as categorias VP, FP, FN e VN, pode-se totalizar a quantidade de resultados em cada categoria e construir uma tabela. A título de exemplo, considere-se a realização de um determinado exame em um grupo de 20 pacientes, onde os resultados indicaram que 12 pacientes apresentavam diagnóstico positivo para a doença e os 8 restantes apresentavam resultados negativos. Após avaliação por um médico especialista na doença em questão, verificou-se que dos 12 resultados apontados como positivos pelo teste, apenas 10 pacientes efetivamente estavam doentes, enquanto que dos 8 resultados negativos somente 3 pacientes seguramente não encontravam-se enfermos. Assim, pode-se elaborar a seguinte tabela:

Ao avaliar os resultados da tabela 2.2, nota-se que o exame diagnóstico produziu 10 resultados verdadeiros positivos para 15 pacientes efetivamente doentes (índice de acerto de 2/3), ao mesmo tempo que identificou corretamente a inexistência da enfermidade em

Tabela 2.2: Exemplo de tabela 2x2 para interpretação de testes diagnósticos

Resultado	Presença da doença	Ausência da doença	Totais
Teste positivo	10	2	12
Teste negativo	5	3	8
Totais	15	5	20

3 pacientes de um total de 5 que não possuíam a doença (índice de acerto de 3/5). Estes números serão utilizados no cálculo de dois importantes fatores para medição da acurácia de um teste diagnóstico: a sensibilidade (índice de acerto para presença da doença) e a especificidade (índice de acerto para ausência da doença).

2.2.1 Sensibilidade

A sensibilidade de um exame diagnóstico mede a proporção de pacientes com uma determinada doença que conseguem ser identificados positivamente a partir do resultado de um teste. A medida da sensibilidade, expressa através de um número no intervalo entre 0 e 1, pode ser calculada dividindo-se a quantidade de resultados verdadeiros positivos pela quantidade total de pacientes com a doença, ou seja, o somatório de verdadeiros positivos e falsos negativos.

$$S(T) = \frac{VP}{VP + FN}$$

No exemplo apresentado na tabela 2.2, a sensibilidade do exame diagnóstico é de 0,66 (2/3).

2.2.2 Especificidade

A especificidade de um exame diagnóstico mede a proporção de pacientes normais (sem a doença) que conseguem ser identificados como tal através de um teste. Assim como a sensibilidade, a medida da especificidade é igualmente expressa através de um número no intervalo entre 0 e 1, sendo calculada através da razão entre a quantidade de resultados verdadeiros negativos e a quantidade total de pacientes normais, ou seja, o somatório de verdadeiros negativos e falsos positivos.

$$E(T) = \frac{VN}{VN + FP}$$

Retomando o exemplo da tabela 2.2, a especificidade do exame é de 0,60 (3/5).

2.2.3 Curvas ROC (*Receiver Operating Characteristic*)

Curvas ROC foram usadas pela primeira vez durante a Segunda Guerra Mundial para análise de sinais de radar antes de serem empregadas na teoria de detecção de sinais (GREEN; SWETS, 1966). Na medicina tem sido utilizadas extensivamente para avaliação de testes diagnósticos e efetividade de novas drogas (ZWEIG; CAMPBELL, 1993; PEPE, 2003) assim como para avaliação de técnicas em radiologia (OBUCHOWSKI, 2003). Fawcett constitui uma boa referência sobre utilização de curvas ROC em projetos de pesquisa (FAWCETT, 2004).

A sensibilidade e especificidade de um teste diagnóstico dependem não apenas da qualidade do exame em si, mas sim dos conceitos de normalidade e anormalidade para o mesmo.

No exemplo a seguir encontra-se a distribuição de quantidade de pacientes dado um valor de teste.

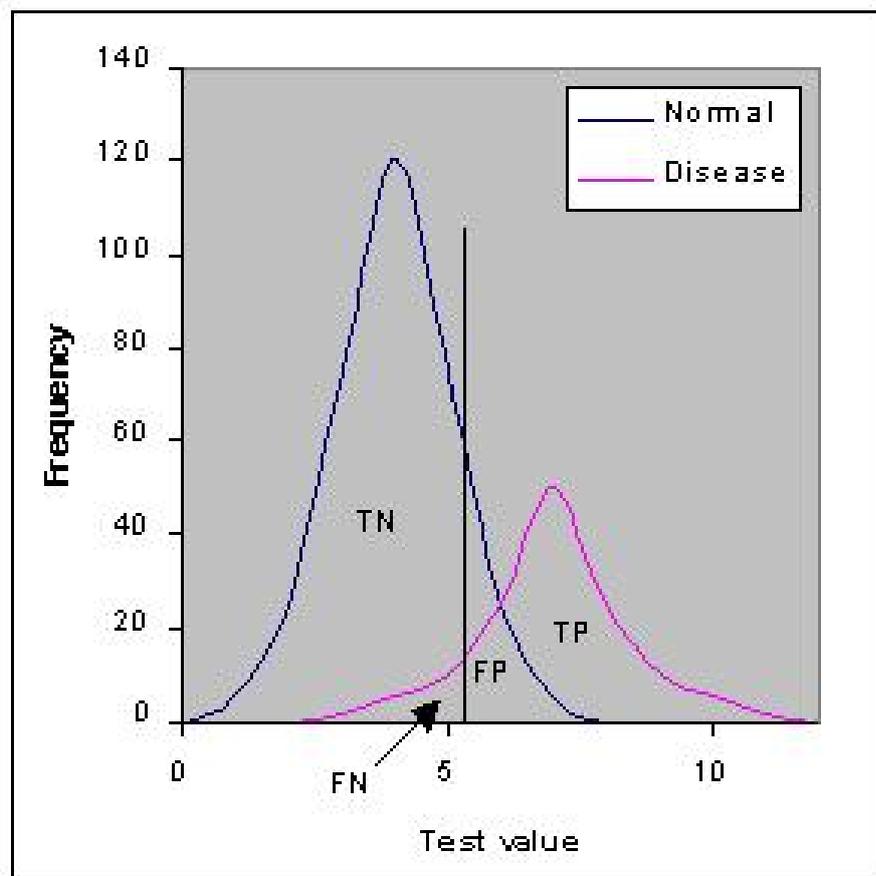


Figura 2.11: Limiar de decisão em um teste diagnóstico (TAPE, 2006)

Note-se que a área de sobreposição (*overlap*) das curvas indica a região onde o teste não consegue distinguir entre pacientes com e sem a doença com 100 % de acurácia. Na prática

é definido um limiar de corte acima ou abaixo do qual considera-se a doença presente. A partir da definição do limiar pode-se obter a quantidade de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos. Pode-se utilizar diferentes limiares para diferentes situações clínicas.

Ao variar os pontos de corte obtém-se novos valores para sensibilidade e especificidade do teste. Tome-se como exemplo fictício os resultados de um teste diagnóstico para identificação de pacientes com glaucoma através da medida da pressão intra-ocular (PIO) obtida pelo exame de tonometria (tabela 2.3). Nesta tabela encontram-se as faixas de valores encontrados para a medida da PIO, juntamente com as respectivas quantidades de indivíduos com glaucoma e normais.

Tabela 2.3: Exemplo de tabela com resultados da classificação de glaucoma de acordo com a medida da pressão intra-ocular.

Valor PIO	Glaucoma	Normal
Inferior a 21	3	39
21 - 26.9	4	36
27 - 28.9	7	17
Superior a 29	18	1
Totais	32	93

As quantidades de indivíduos com teste positivo (glaucoma) e negativo (normal) são alteradas de acordo com a escolha do limiar de PIO. À medida que se varia este limiar de decisão, as respectivas sensibilidades e especificidades do teste podem ser calculadas:

- **Limiar 21 ou superior** (tabela 2.4) - sensibilidade 0.91 (29/32) e especificidade 0.42 (39/93).
- **Limiar 27 ou superior** (tabela 2.5) - sensibilidade 0.78 (25/32) e especificidade 0.81 (75/93).
- **Limiar 29 ou superior** (tabela 2.6) - sensibilidade 0.56 (18/32) e especificidade 0.99 (92/93).

Tabela 2.4: Tabela de classificação de glaucoma de acordo com limiar 21 para valor da medida da pressão intra-ocular.

Valor PIO	Glaucoma	Normal
Inferior a 21	3	39
21 ou superior	29	54
Totais	32	93

Tabela 2.5: Tabela de classificação de glaucoma de acordo com limiar 27 para valor da medida da pressão intra-ocular.

Valor PIO	Glaucoma	Normal
Inferior a 27	7	75
27 ou superior	25	18
Totais	32	93

Tabela 2.6: Tabela de classificação de glaucoma de acordo com limiar 21 para valor da medida da pressão intra-ocular.

Valor PIO	Glaucoma	Normal
Inferior a 29	14	92
29 ou superior	18	1
Totais	32	93

A tabela 2.7 apresenta os resultados de especificidade e sensibilidade obtidos de acordo com a variação do limiar de PIO. Note-se que à medida que o limiar varia de 21 a 29 as sensibilidades diminuem ocorre o aumento da especificidade. Assim, a análise de um teste diagnóstico deve ponderar sensibilidade e especificidade, buscando maximizar ambos os índices em conjunto.

Tabela 2.7: Tabela de sensibilidades e especificidades de acordo com variação do limiar para valor da medida da pressão intra-ocular.

Limiar	Sensibilidade	Especificidade
21	0,91	0,42
27	0,78	0,81
29	0,56	0,99

2.2.3.1 Construindo a Curva ROC

Uma Curva ROC consiste em uma representação gráfica da taxa de verdadeiros positivos (sensibilidade) versus a taxa de falsos positivos (1-especificidade) para diferentes valores de limiar em um teste diagnóstico.

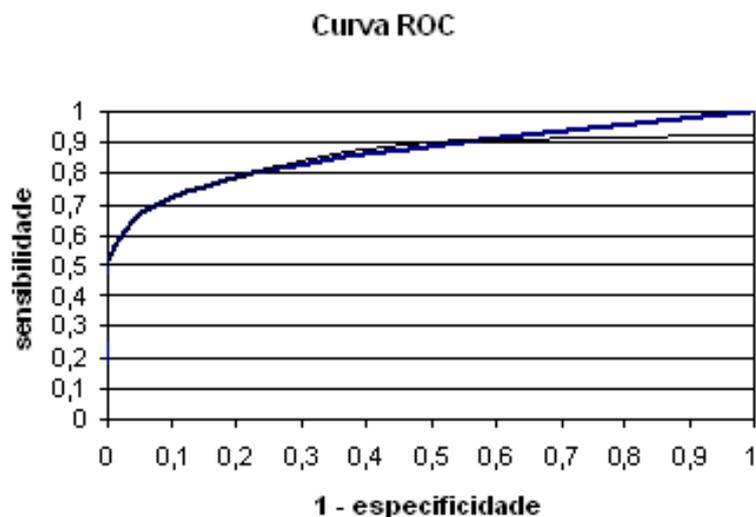
Para cada valor de limiar, além de se calcular a sensibilidade e especificidade, deve-se calcular então o valor de 1-especificidade (tabela 2.8).

A seguir constrói-se uma curva (figura 2.12) onde o eixo das abscissas é representado pelos valores de 1-especificidade, e o eixo das ordenadas pela sensibilidade dos testes para diferentes limiares.

Uma curva ROC demonstra a relação entre sensibilidade e especificidade, onde aumentos na sensibilidade são acompanhados de diminuições na especificidade.

Tabela 2.8: Tabela base para construção da Curva ROC

Limiar	Sensibilidade	1-Especificidade
21	0,56	0,01
27	0,78	0,19
29	0,91	0,58

**Figura 2.12:** Curva ROC para teste diagnóstico de glaucoma de acordo com valor da medida da pressão intra-ocular.

O teste apresentará maior acurácia quando a curva se aproximar da borda esquerda e superior do gráfico (maior área sob a curva) Quanto mais a curva se aproximar da reta bissetriz do espaço da ROC menor a acurácia do teste.

A área sob a curva é a medida da acurácia do teste (TAPE, 2006):

- 0,90-1,00 = excelente;
- 0,80-0,90 = bom ;
- 0,70-0,80 = regular;
- 0,60-0,70 = pobre;
- 0,50-0,60 = falhou.

3 Classificadores

O diagnóstico de glaucoma baseado na integração de dados epidemiológicos, clínicos, anatômicos e funcionais de pacientes pode ser claramente definido como um problema de classificação. De um lado dispõe-se dos dados obtidos através de anamneses e dos resultados de testes funcionais e estruturais, e de outro deseja-se como saída a correspondente classificação Glaucoma ou Normal (figura 3.1).



Figura 3.1: O problema de classificação de glaucoma.

Do ponto de vista funcional, classificadores são algoritmos que recebem como entrada um vetor de características (como os valores extraídos de exames de um paciente visando ao diagnóstico de glaucoma), e apresentam como resultado a indicação de uma classe (categoria) à qual aquele vetor pode ser associado (glaucoma ou normal).

Geralmente estes algoritmos se utilizam de uma forma de aprendizagem supervisionada, onde o programa aprende através de exemplos de treinamento positivos e negativos, compostos por valores para atributos selecionados que lhe são fornecidos juntamente com sua correspondente classificação. Um exemplo positivo para o treinamento de um classificador para glaucoma é um conjunto de valores de atributos como idade, PIO, volume da escavação, volume do disco óptico e espessura média da camada de fibras nervosas de um paciente com a doença, enquanto que um exemplo negativo seriam valores para os mesmos parâmetros, porém referentes a um paciente normal (sem a doença). O treinamento é repetido diversas vezes com o fornecimento de vários exemplos juntamente com a classificação positivo ou negativo realizada por um médico, até que o conceito (identificação da doença através do exame em questão) possa ter sido devidamente aprendido pelo sistema. Passa-se então a utilizar o algoritmo com novos dados, porém não previamente classificados por um especialista, para que o sistema então faça sua classificação

automaticamente.

O aprendizado supervisionado de um algoritmo classificador automático é baseado em um conjunto de exemplos de treinamento representativos da população, devidamente rotulados em relação à classe à qual pertencem. Estes dados serão empregados como experiência para o algoritmo. Segundo Mitchell (MITCHELL, 1997), diz-se que um programa é capaz de aprender com a experiência E em relação a uma classe de tarefas T se seu desempenho P ao realizar as tarefas em T melhora com a experiência E .

Por exemplo, um programa de computador que aprende a classificar dados obtidos através de testes diagnósticos de um paciente como sendo de um indivíduo com glaucoma pode ter seu desempenho medido através da sensibilidade e especificidade na tarefa, e pode ter esse desempenho incrementado através da experiência de treinamento com diversos registros clínicos de pacientes com glaucoma e normais classificados por médicos especialistas. Neste caso a definição deste problema de aprendizagem seria a seguinte:

- **Tarefa T:** classificar registros clínicos quanto à presença de indícios de glaucoma;
- **Medida de desempenho P:** cálculo da sensibilidade e especificidade na tarefa;
- **Experiência de treinamento E:** banco de dados contendo registros clínicos e respectiva classificação - glaucoma ou normal.

Uma vez definido o problema de aprendizagem, para se construir um sistema que aprenda com a experiência é fundamental selecionar-se o conjunto de dados que será utilizado no treinamento. Uma característica importante deste conjunto de treinamento é a sua capacidade de representar adequadamente a distribuição destes dados no domínio real do problema. Para que um programa consiga aprender a diferenciar indivíduos normais de glaucomatosos, os dados de treinamento devem ser estatisticamente representativos da população em geral, levando-se em consideração a prevalência da doença. Por exemplo, apresentar ao mecanismo de aprendizagem durante seu treinamento apenas dados de pacientes normais fará com que o sistema falhe ao tentar reconhecer futuramente dados de indivíduos com glaucoma.

Após a definição formal do problema e da seleção da experiência de treinamento faz-se necessário estabelecer:

1. O exato tipo de conhecimento a ser aprendido pelo sistema;
2. A forma de representação deste conhecimento;
3. Um mecanismo de aprendizagem.

No caso do diagnóstico de glaucoma, o tipo de conhecimento que se deseja que o sistema aprenda é a capacidade de identificar indícios de glaucoma e normais através de dados extraídos de exames diagnósticos de forma a produzir uma provável classificação sobre a doença com acurácia adequada. Este conhecimento será representado através de uma função-alvo que determine uma curva no espaço n -dimensional (onde n é o número total de atributos disponíveis, ou seja, o número de parâmetros obtidos dos exames diagnósticos) que separe dados de indivíduos glaucomatosos de normais. Na prática, a aplicação desta função sobre um determinado conjunto de atributos pertencentes a um indivíduo, produzirá como resultado dois possíveis valores: glaucoma ou normal. Como mecanismo de aprendizagem podem ser empregadas redes neurais artificiais treinadas com algoritmo de retro propagação de erros, que se adequam à tarefa por possuírem forte poder de aproximação de diversos tipos de funções.

A função-alvo ou conceito a ser aprendido pelo sistema sobre um conjunto de dados-exemplo ou instâncias X é denominado conceito-alvo c , que pode apresentar, no caso de funções booleanas por exemplo, os valores 0 ou 1. Escreve-se, portanto que $c : X \rightarrow \{0, 1\}$ para indicar que o conceito-alvo c é definido sobre os valores de instâncias X e pode apresentar valores $\{0, 1\}$ quando aplicado sobre cada uma das instâncias $x \in X$.

Em muitos problemas de aprendizagem, diversas funções h , onde H é o conjunto de possíveis hipóteses que podem ser elaboradas sobre X , podem ser definidas de forma que $h(x) = c(x)$. Por armazenar as possíveis hipóteses de representação de um conceito-alvo, H é também chamado de espaço de hipóteses. Assim, o processo de aprendizagem de conceitos busca identificar a melhor hipótese $h \in H$ onde $h : X \rightarrow \{0, 1\}$ para todo $x \in X$.

Quando a única informação disponível sobre o conceito c é seu valor para cada instância x , o aprendizado é chamado indutivo. Neste caso, os algoritmos de aprendizagem podem garantir no máximo que a h possui a capacidade de representação de c sobre os dados pertencentes ao conjunto de treinamento. Contudo, uma definição importante para aprendizagem indutiva é que, se h pode representar adequadamente c sobre as instâncias de treinamento X , então h pode representar igualmente c sobre novas instâncias não apresentadas ao algoritmo de aprendizagem.

3.1 Categorias de classificadores

Um problema de classificação consiste em identificar entre possíveis valores do conceito c , o valor mais provável para uma dada instância x apresentada, ou seja, obter-se $c(x)$. Nestes

problemas, o conjunto de valores para representar o conceito é previamente conhecido (como {Glaucoma, Normal} no caso do diagnóstico de glaucoma) e o que se espera do algoritmo é a previsão de um possível valor de classe dado um conjunto de valores de atributos a ele submetidos.

Entre diversas técnicas empregadas na construção de classificadores estão os algoritmos simbólicos de regras CN2 de Clark e Niblett (CLARK; NIBLETT, 1989), árvore de decisão como o C4.5 de Quinlan (QUINLAN, 1993), programação lógica indutiva (MUGGLETON, 1994), algoritmos Bayesianos (HECKERMAN; GEIGER; CHICKERING, 1995) e redes neurais artificiais (CHAUVIN; RUMELHART, 1995).

Diante do longo histórico de utilização em tarefas de diagnóstico médico (KONONENKO, 2001) e da capacidade de expressão de conceitos, as seguintes categorias de algoritmos foram selecionadas para aplicação neste projeto:

- Árvores de decisão;
- Naive-Bayes;
- Redes Neurais Artificiais de Múltiplas Camadas;
- Bagging.

A literatura apresenta diversas referências a essas classes de algoritmos e suas aplicações, sendo que apenas uma pequena parte delas são citadas neste trabalho diante do seu escopo. Contudo, cada um destes algoritmos será sucintamente apresentado a seguir.

3.1.1 Árvores de decisão

Árvores de decisão são métodos para aproximação de funções-alvo baseadas em valores discretos. A representação desta função é dada através de uma árvore de regras se-então onde cada nó representa uma regra e as folhas representam o conceito-alvo.

No exemplo da figura 3.2, o conceito JogarTênis é representado por uma árvore onde cada nó (retângulo na figura) 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 chamado de raiz enquanto que os últimos são denominados folhas. O processo de classificação inicia-se com a apresentação de uma instância de valores de atributos 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

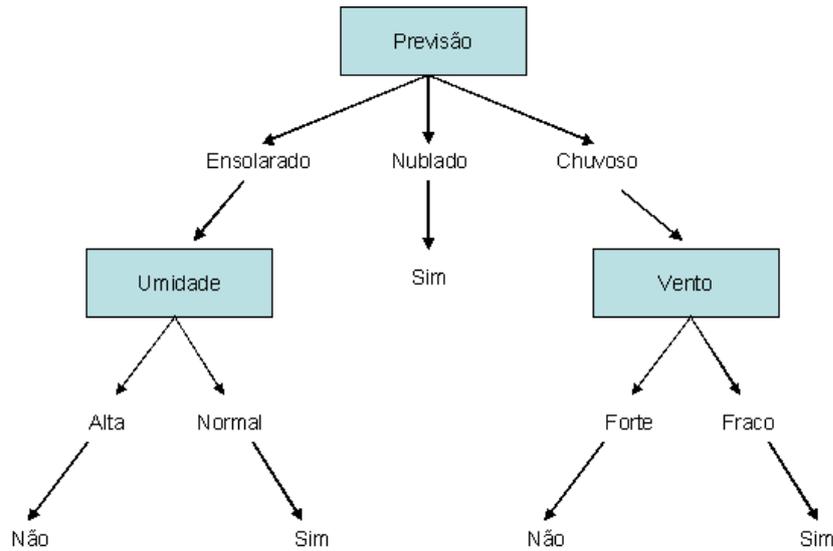


Figura 3.2: Exemplo de árvore de decisão para representação do conceito JogarTênis. Adaptado de (MITCHELL, 1997).

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.

Geralmente, árvores de decisão representam conceitos na forma de disjunções (cláusulas OU - \vee) de conjunções (cláusulas E - \wedge) de restrições sobre os valores dos atributos. Retomando o exemplo da figura 3.2, a árvore de decisão em questão pode ser representada pela seguinte expressão:

$$\begin{aligned}
 & (Previsão = Ensolarado \wedge Umidade = Alta) \\
 & \vee \\
 & (Previsão = Nublado) \\
 & \vee \\
 & (Previsão = Chuvoso \wedge Vento = Fraco)
 \end{aligned}$$

A maioria dos algoritmos para árvores de decisão, como os desenvolvidos por Quinlan, ID3 (QUINLAN, 1983) e seu sucessor C4.5 (QUINLAN, 1993), promovem uma busca pelo espaço de hipóteses, representado pelo conjunto de possíveis árvores, em busca da hipótese (árvore) que melhor represente o conceito-alvo. No ID3, por exemplo, a busca por esta hipótese inicia-se pela escolha do primeiro atributo a ser testado, ou seja, o atributo-

raiz da árvore. Para tanto, uma avaliação sobre o conjunto formado pelos exemplos de treinamento do algoritmo busca identificar o atributo que isoladamente possui maior poder de separação entre exemplos positivos e negativos. Tal avaliação é conduzida sobre uma propriedade estatística denominada ganho de informação (*information gain*).

Comumente aplicada na teoria da informação, a homogeneidade ou (im)pureza de um conjunto (S) de exemplos positivos e negativos pode ser medida através do cálculo da sua entropia de acordo com a fórmula:

$$Entropia(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

Neste cálculo de entropia, p_{\oplus} é a proporção de exemplos positivos em S e p_{\ominus} a de exemplos negativos. Se Entropia(s)=0 então todos os exemplos do conjunto pertencem à mesma classe. Entropia(S)=1 indica que a quantidade de exemplos positivos em S é igual à de exemplos negativos.

Tabela 3.1: Exemplos de treinamento para o conceito JogarTênis. Adaptado de (MITCHELL, 1997)

Ex.	Previsão	Temperatura	Umidade	Vento	JogarTênis
1	Ensolarado	Quente	Alta	Fraco	Não
2	Ensolarado	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuvoso	Temperado	Alta	Fraco	Sim
5	Chuvoso	Frio	Normal	Fraco	Sim
6	Chuvoso	Frio	Normal	Forte	Não
7	Nublado	Frio	Normal	Forte	Sim
8	Ensolarado	Temperado	Alta	Fraco	Não
9	Ensolarado	Frio	Normal	Fraco	Sim
10	Chuvoso	Temperado	Normal	Fraco	Sim
11	Ensolarado	Temperado	Normal	Forte	Sim
12	Nublado	Temperado	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuvoso	Temperado	Alta	Forte	Não

Dados os 14 exemplos de treinamento apresentados na tabela 3.1 é possível observar-se que são 9 positivos (para os quais o conceito JogarTênis=Sim) e 5 negativos (onde o conceito JogarTênis=Não). Assim, o cálculo da entropia deste conjunto S de exemplos pode ser calculada da seguinte forma:

$$\begin{aligned} Entropia(S) &= -(9/14) \log_2(9/14) - (5/14) \log_2(5/14) \\ &= 0,940 \end{aligned}$$

Um vez determinada a entropia do conjunto, é possível calcular o ganho de informação de um determinado atributo deste conjunto. O ganho de informação $Ganho(S,A)$ é a medida da capacidade de separação entre exemplos positivos e negativos que um atributo A apresenta em relação a S, sendo portanto assim definido:

$$Ganho(S, A) \equiv Entropia(S) - \sum_{v \in Valores(A)} \frac{[S_v]}{[S]} Entropia(S_v)$$

Onde $Valores(A)$ é o conjunto de possíveis valores para o atributo A e S_v é o subconjunto de S para os quais o atributo A possui valor v. A primeira parte da fórmula apresenta a entropia original de S, enquanto que a segunda representa a entropia após a segmentação de S de acordo com os valores do atributo A.

Retornando ao exemplo da tabela 3.1, o cálculo do ganho de informação para o atributo Vento pode-se obter da seguinte forma:

$$\begin{aligned} Valores(Vento) &= \{Fraco, Forte\} \\ S_{Fraco} &\leftarrow [6+, 2-] \\ S_{Forte} &\leftarrow [3+, 3-] \\ Ganho(S, Vento) &= Entropia(S) - \sum_{v \in \{Forte, Fraco\}} \frac{[S_v]}{[S]} Entropia(S_v) \\ &= Entropia(S) - \frac{8}{14} Entropia(S_{Fraco}) - \frac{6}{14} Entropia(S_{Forte}) \\ &= 0,940 - \frac{8}{14} 0,811 - \frac{6}{14} 1,000 \\ &= 0,048 \end{aligned}$$

Após se calcular o ganho de informação de cada atributo do conjunto, 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.

$$\begin{aligned} Ganho(S, Previsão) &= 0,246 \\ Ganho(S, Umidade) &= 0,151 \end{aligned}$$

$$\text{Ganho}(S, \text{Vento}) = 0,048$$

$$\text{Ganho}(S, \text{Temperatura}) = 0,029$$

Em seguida, o processo se repete para cada valor do atributo raiz, calculando-se o ganho de informação de cada um dos demais atributos em relação ao subconjunto de exemplos positivos e negativos que apresentam este valor de ramo. O processo de escolha de novos nós em um determinado ramo se encerrará quando todos os atributos já tiverem sido utilizados previamente ou quando todos os exemplos referentes a um determinado nó tenham o mesmo valor (entropia zero).

Apesar de se destinarem a problemas onde as instâncias dos conjuntos de dados são representadas por valores discretos, é possível estender o uso das árvores de decisão para dados contínuos. No exemplo citado, o atributo temperatura, que representa uma grandeza física expressa em valores reais, apresenta 3 possíveis valores: Frio, Temperado e Quente. Para tanto deve-se utilizar um mapeamento por faixas de valores, para a temperatura (T):

- $T \leq 10 \Rightarrow \text{Frio}$;
- $10 < T < 20 \Rightarrow \text{Temperado}$;
- $T \geq 20 \Rightarrow \text{Quente}$.

Contudo, a seleção de quais valores de limiar devem ser usados é o fator principal para se aplicar este mapeamento de valores contínuos em discretos. Autores como Fayyad e Irani (FAYYAD; IRANI, 1993), Utgoff e Brodley (UTGOFF P E ANS BRODLEY, 1991) e Murthy et al (MURTHY; KASIF; SALZBERG, 1994) apresentam abordagens para este questão.

3.1.2 Naive-Bayes

Em aprendizagem de máquina há o interesse na determinação da melhor hipótese de um espaço de hipóteses H dado um exemplo de treinamento D observado. Por melhor hipótese entenda-se a mais provável hipótese em H , dados D e as probabilidades a priori das diversas hipóteses em H . O teorema de Bayes apresenta-se como um método para calcular estas probabilidades.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Onde:

- $P(h)$ é a probabilidade inicial ou probabilidade a priori da hipótese h antes da observação do exemplo de treinamento D e reflete o conhecimento de fundo sobre h . Caso não se disponha deste conhecimento, deve-se considerar que todas as hipóteses são equiprováveis em H .
- $P(D)$ é a probabilidade inicial ou probabilidade a priori do exemplo de treinamento D , ou seja, a probabilidade do exemplo D ser observado sem que o conhecimento sobre qual hipótese h é verdadeira para D .
- $P(D|h)$ é a probabilidade de D condicional a h , ou seja, a probabilidade de D ocorrer sabendo-se que h é verdadeira.
- $P(h|D)$ é a probabilidade a posteriori de h dado D , ou seja, a probabilidade da hipótese h ser verdadeira face a ocorrência do exemplo de treinamento D .

A abordagem de um classificador Bayesiano para classificar uma nova instância é atribuir o valor objetivo mais provável, v_{MAP} , dados os valores de atributos (a_1, a_2, \dots, a_n) que descrevem a instância.

$$v_{MAP} = \arg \max_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n)$$

Utilizando-se o teorema de Bayes tem-se:

$$v_{MAP} = \arg \max_{v_j \in V} \frac{P(a_1, a_2, \dots, a_n | v_j) \cdot P(v_j)}{P(a_1, a_2, \dots, a_n)}$$

$$v_{MAP} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) \cdot P(v_j)$$

Um classificador Naive-Bayes (ou Bayes ingênuo) é aquele que considera que todos os valores de atributos são condicionalmente independentes. Assim, dado um valor objetivo para uma instância, a probabilidade de observar-se a conjunção a_1, a_2, \dots, a_n corresponde apenas ao produto dos atributos individuais: $P(a_1, a_2, \dots, a_n | v_j) = \prod_i P(a_i | v_j)$. Assim, substituindo este resultado na equação anterior tem-se a expressão para o classificador

Bayesiano ingênuo:

$$v_{NB} = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad (3.1)$$

Utilizando-se como exemplo novamente a tabela 3.1, deseja-se obter o valor mais provável do conceito JogarTênis (Sim ou Não) para a seguinte instância D de valores de atributos:

$$\{Previsão = Ensolarado, Temperatura = Frio, Umidade = Alta, Vento = Forte\}$$

Para tanto deve-se calcular o valor v_{NB} de acordo com a equação 3.1:

$$v_{NB} = \arg \max_{v_j \in \{Sim, Não\}} P(v_j) \cdot P(Previsão = Ensolarado | v_j) \cdot P(Temperatura = Frio | v_j) \cdot P(Umidade = Alta | v_j) \cdot P(Vento = Forte | v_j)$$

$$P(JogarTênis = Sim) = \frac{9}{14} = 0,64$$

$$P(JogarTênis = Não) = \frac{5}{14} = 0,36$$

$$P(Previsão = Ensolarado | JogarTênis = Sim) = \frac{2}{9} = 0,22$$

$$P(Previsão = Ensolarado | JogarTênis = Não) = \frac{3}{5} = 0,60$$

$$P(Temperatura = Frio | JogarTênis = Sim) = \frac{3}{9} = 0,33$$

$$P(Temperatura = Frio | JogarTênis = Não) = \frac{1}{5} = 0,20$$

$$P(Umidade = Alta | JogarTênis = Sim) = \frac{4}{9} = 0,44$$

$$P(Umidade = Alta | JogarTênis = Não) = \frac{3}{5} = 0,60$$

$$P(Vento = Forte | JogarTênis = Sim) = \frac{3}{9} = 0,33$$

$$P(Vento = Forte | JogarTênis = Não) = \frac{3}{5} = 0,60$$

Assim, para calcular $P(\text{JogarTênis}|D)$ para a instância apresentada tem-se:

$$\begin{aligned}
 P(\text{JogarTênis} = \text{Sim}|D) &= P(\text{JogarTênis} = \text{Sim}). \\
 &P(\text{Previsão} = \text{Ensolarado}|\text{JogarTênis} = \text{Sim}). \\
 &P(\text{Temperatura} = \text{Frio}|\text{JogarTênis} = \text{Sim}). \\
 &P(\text{Umidade} = \text{Alta}|\text{JogarTênis} = \text{Sim}). \\
 &P(\text{Vento} = \text{Forte}|\text{JogarTênis} = \text{Sim}) \\
 &= 0.007 \\
 P(\text{JogarTênis} = \text{Não}|D) &= P(\text{JogarTênis} = \text{Não}) \\
 &P(\text{Previsão} = \text{Ensolarado}|\text{JogarTênis} = \text{Não}). \\
 &P(\text{Temperatura} = \text{Frio}|\text{JogarTênis} = \text{Não}). \\
 &P(\text{Umidade} = \text{Alta}|\text{JogarTênis} = \text{Não}). \\
 &P(\text{Vento} = \text{Forte}|\text{JogarTênis} = \text{Não}) \\
 &= 0,015
 \end{aligned}$$

Portanto, como $P(\text{JogarTênis} = \text{Sim}|D) < P(\text{JogarTênis} = \text{Não}|D)$ então o classificador bayesiano ingênuo atribuiria o valor Não para a função $\text{JogarTênis}(D)$.

3.1.3 Redes neurais artificiais de múltiplas camadas

Multi-Layer Perceptrons (MLP) são um tipo de rede neural artificial alimentada para frente, composta por um ou mais nós computacionais (neurônios) na camada de entrada, um ou mais neurônios na camada de saída, e uma ou mais camadas ocultas intermediárias de nós computacionais (HAYKIN, 1999).

Neste tipo de rede o sinal de entrada propaga-se para a frente, camada a camada, a partir da entrada até a saída. MLPs têm sido aplicadas com sucesso à resolução de diversos problemas envolvendo reconhecimento de padrões. Um dos motivos deste sucesso reside no fato de que a rede MLP pode ser eficientemente treinada de forma supervisionada através de algoritmos de retro propagação de erro (*error backpropagation*).

A aprendizagem por retro propagação de erro (figura 3.4) consiste em dois passos através das diversas camadas da rede: na primeira passagem - para frente - o vetor de

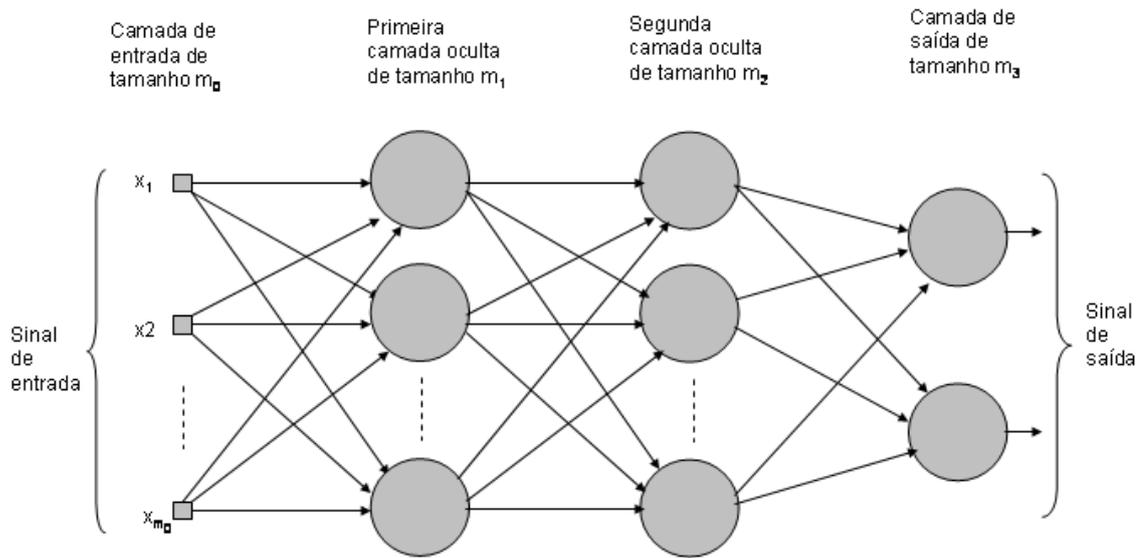


Figura 3.3: Grafo estrutural de um *perceptron* de múltiplas camadas com duas camadas ocultas.

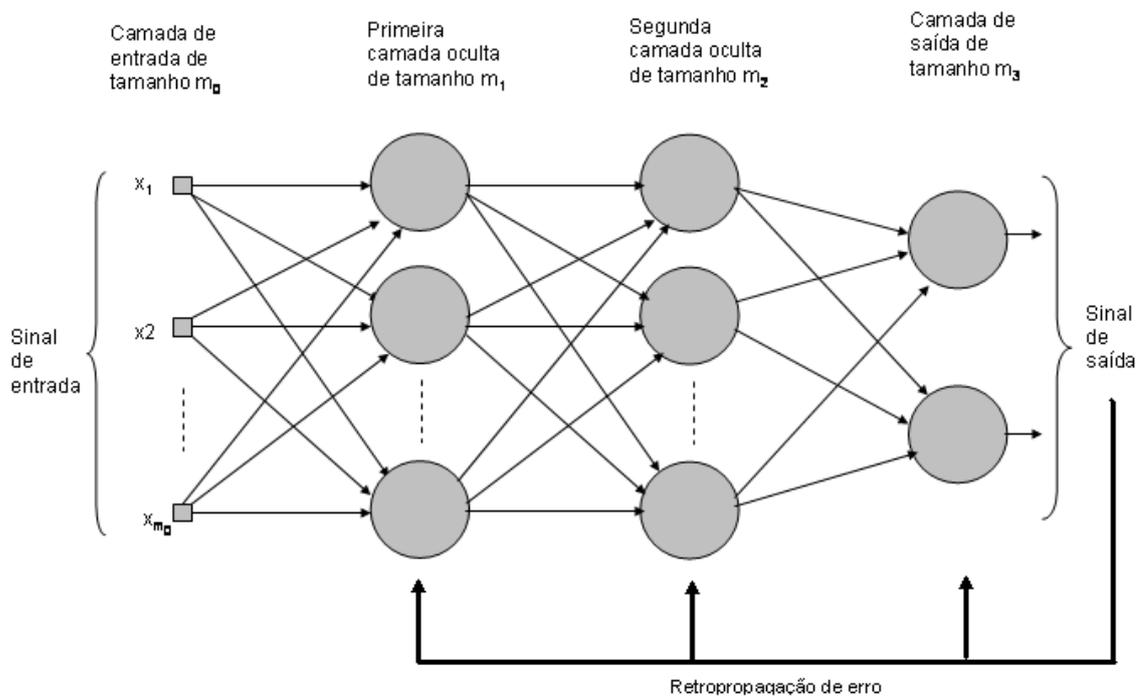


Figura 3.4: Grafo estrutural de um *perceptron* de múltiplas camadas com duas camadas ocultas e retro propagação de erro.

entrada é aplicado aos neurônios da primeira camada e seu efeito se propaga através da rede, camada por camada, utilizando-se pesos sinápticos inicialmente fixos, *bias* e uma função de ativação de neurônio. Após propagação até a camada de saída, passando-se por todas as camadas ocultas intermediárias, tem-se a resposta da rede ao vetor de entrada apresentado. Na segunda passagem - para trás - aplica-se uma regra para cálculo de erro levando-se em consideração a diferença entre a resposta produzida e a desejada, reajusta-se os pesos sinápticos, até então fixos, e repete-se o processo de propagação. Este processo então é executado iterativamente até que a resposta obtida pela rede apresente erro próximo a limiares aceitáveis para a tarefa com uma determinada taxa de aprendizagem η e um momento α .

Em um problema de classificação para duas classes, utilizando uma rede MLP com uma camada oculta, um dado vetor de entrada $x = (x_1, x_2, \dots, x_D)^T$ e um neurônio na camada de saída tem-se:

$$z_j = g_t\left(\sum_{d=1}^D w_{jd}x_d + w_{j0}\right)$$

$$f = h_t\left(\sum_{j=1}^J v_j z_j + v_0\right)$$

Onde:

- z_j , $j=1, 2, \dots, J$ representam as ativações de nós computacionais na camada oculta,
- w_{jd} são os pesos entre a camada de entrada e a camada oculta,
- v_j são os pesos entre a camada oculta e a camada de saída f ,
- w_{j0}, v_0 são os *biases* para as camadas oculta e de saída, respectivamente.
- h_t e g_t são funções de ativação de cada neurônio da rede, funções sigmóides geralmente da forma $\tanh(t)$ ou $1/(1 + e^{-t})$.

Retomando o exemplo da tabela 3.1, pode-se construir uma rede neural artificial de três camadas para aprender o conceito JogarTênis (figura 3.5) através da definição das seguintes características no projeto da rede:

1. Os vetores de entrada da rede serão formados por 4 atributos: Previsão, Temperatura, Umidade e Vento
2. Cada atributo corresponderá a uma unidade de entrada para a rede.

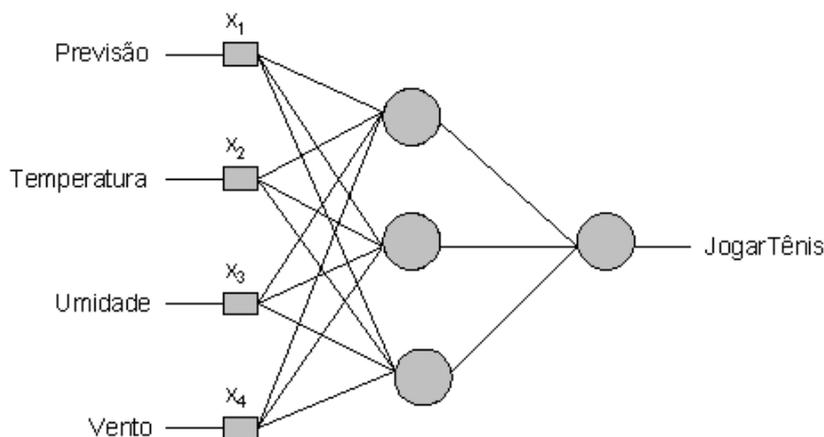


Figura 3.5: Grafo estrutural de uma rede neural artificial de três camadas para representar a decisão JogarTênis.

3. Os valores discretos de cada atributo serão mapeados para valores reais no intervalo entre 0 e 1. Por exemplo, o valor de Previsão Chuvoso será mapeado para 0,1, Nublado para 0,5 e Ensolarado para 0.9.
4. Serão utilizadas 2 camadas de neurônios, sendo uma oculta e uma de saída.
5. A saída da rede será composta por 1 nó computacional que representará os valores finais da função-alvo.
6. Os valores de saída correspondentes a Sim ou Não serão mapeados respectivamente para 0,9 e 0,1.
7. Os pesos sinápticos e *biases* serão atribuídos aleatoriamente em valores pequenos (normalmente no intervalo entre -0,5 e +0,5).
8. A taxa de aprendizagem será, por exemplo, 0,3 e o momento 0,2.
9. As funções de ativação dos neurônios serão funções sigmóides do tipo $1/(1 + e^{-t})$.

No instante inicial de treinamento da rede, os pesos sinápticos são atribuídos aleatoriamente. A partir da segunda iteração, o algoritmo de retropropagação de erros recalculará os pesos de cada camada de neurônios para minimizar o erro médio quadrático obtido através das diferenças entre as saídas produzidas pela rede e os respectivos valores esperados de acordo com os exemplos de treinamento.

3.1.4 *Bagging*

Bagging é uma técnica que se utiliza da combinação de diversos modelos para realizar a classificação, através de um processo semelhante à votação de especialistas, para decidir sobre a classe será atribuída a uma determinada instância (WITTEN; FRANK, 2005).

Para ilustrar o princípio básico desta técnica, tome-se por exemplo um conjunto de *datasets* obtidos aleatoriamente a partir de exemplos de um determinado domínio. Em seguida, induz-se árvores de decisão individualmente para cada *dataset*. Desta forma, cada árvore provavelmente representará uma hipótese diferente. Em seguida apresenta-se para este grupo de árvores de decisão uma instância inédita para classificação como um exemplo positivo ou negativo no domínio em questão. Cada algoritmo apresentará sua provável classificação para esta instância, como membros de um júri. O resultado que obtiver a maior quantidade de votos será então a classe mais provável, e portanto a resposta (veredicto) do grupo classificadores.

Em termos práticos, a técnica de *bagging* se utiliza deste princípio elementar do voto e emprega a repartição do conjunto de dados para efetuar o treinamento. Porém, neste processo não ocorre uma simples divisão dos dados em grupos. Uma re-amostragem do mesmo conjunto é implementada, onde novas instâncias são geradas através da duplicação de exemplos existentes e acrescentadas em alguns dos subconjuntos de dados particionados, enquanto que em outros subconjuntos algumas instâncias são eliminadas. O objetivo desta técnica é reduzir o erro e portanto aumentar a acurácia da tarefa de classificação.

3.2 Classificadores híbridos

Conforme já mencionado, quando a única informação disponível sobre o conceito-alvo é seu valor para cada instância do conjunto de treinamento, o aprendizado é chamado indutivo. Árvores de decisão, Naive-Bayes e redes neurais artificiais são exemplos de algoritmos de aprendizagem indutiva. Entretanto, o aprendizado pode ocorrer também através de mecanismos de inferência sobre uma base de conhecimento, sendo neste caso chamado de analítico. Programas escritos na linguagem PROLOG são exemplos deste tipo de aprendizado, onde o conhecimento é explícito através de um conjunto de sentenças denominadas Cláusulas de Horn, e um mecanismo de inferência é utilizado para deduzir hipóteses. A base de conhecimento é chamada de teoria do domínio.

Quando o aprendizado é construído sobre uma base de conhecimento e um conjunto de exemplos de treinamento, tem-se o aprendizado analítico-indutivo. KBANN (*Knowledge-Based Artificial Neural Network*) é um exemplo de algoritmo analítico-indutivo (MITCHELL, 1997). KBANN é uma rede neural artificial construída a partir de uma base de conhecimento explícita formada por um conjunto de Cláusulas de Horn. Duas etapas podem ser definidas para construção e treinamento deste algoritmo. Na primeira etapa, cada cláusula da teoria do domínio é mapeada para um nó computacional da rede neural

e suas respectivas conjunções são utilizadas para definir os pesos sinápticos para as unidades sigmóides. Na etapa seguinte, apresenta-se à rede os exemplos de treinamento e o aprendizado prossegue como em uma rede neural artificial tradicional alimentada para frente. O algoritmo de retro propagação de erro é então aplicado para refinar os pesos sinápticos da rede, até então fixos.

Esta abordagem analítico-indutiva permite a utilização de teorias incompletas sobre o domínio simultaneamente com conjuntos de treinamento que apresentem ruído, ou seja, que contenham falhas. Diante disso, em diversas tarefas a aplicação de KBANN mostrou-se superior em relação a mecanismos puramente indutivos. Towell et al (TOWELL; SHAVLIK; NOORDEWIER, 1990), por exemplo, obtiveram sucesso no emprego deste classificador híbrido em tarefas de genética molecular ao utilizarem uma teoria do domínio elaborada por um geneticista e um conjunto de 53 exemplos positivos e 53 negativos. Nesta tarefa, que envolveu 106 testes utilizando 105 exemplos cada um, o KBANN teve uma taxa de erros de 4/106 frente uma taxa de 8/106 apresentada por uma rede neural tradicional treinada com retro propagação de erro. Fu (FU, 1993) aplicou uma variação do KBANN sobre os mesmos dados e obteve redução da taxa de erros para 2/106. Ambos relataram que as cláusulas de Horn extraídas dos pesos finais após o treinamento da rede constituíram-se em importante refinamento para a teoria do domínio original. Porém, tal mapeamento contrário dos pesos para a teoria do domínio nem sempre é possível. Outros autores (CRAVEN; SHAVLIK, 1994; CRAVEN, 1996) apresentam técnicas para extrair regras simbólicas a partir de redes treinadas.

3.3 Avaliando a robustez de classificadores

Um método utilizado para avaliação da robustez de classificadores em uma determinada tarefa de reconhecimento de padrões consiste na comparação de áreas sob curvas (AUC - *area under curve*) de *Receiver Operating Characteristic* (ROC). Quanto maior esta área, maior a eficácia do classificador utilizado na tarefa.

Além da utilização na análise da acurácia de testes diagnósticos em medicina, curvas ROC têm sido amplamente aplicadas na avaliação de técnicas de aprendizagem de máquina, sendo que o primeiro relato deste uso foi apresentado por Spackman na comparação e avaliação de diferentes algoritmos de classificação (SPACKMAN, 1989). Ambientes gratuitos de desenvolvimento de experimentos em aprendizagem de máquina e mineração de dados como ORANGE (DEMSAR; ZUPAN; LEBAN, 2004), TANAGRA (RAKOTOMALALA, 2005), WEKA (WITTEN; FRANK, 2005) e YALE (MIERSWA et al., 2006)

apresentam curvas ROC entre suas ferramentas de análise.

A título de exemplo, comparou-se o desempenho dos algoritmos ID3, Naive-Bayes e *Multilayer Perceptrons* na tarefa de classificação do conceito JogarTênis. Para tanto, utilizou-se o ambiente de experimentos em aprendizagem WEKA. Um arquivo no formato ARFF (*Attribute-Relation File Format*), padrão nativo do ambiente, foi gerado com base nos exemplos de treinamento da tabela 3.1. A figura 3.6 exibe o conteúdo deste arquivo onde a estrutura do formato ARFF pode ser identificada.

```
@relation PlayTennis

@attribute Previsão {Ensolarado,Nublado,Chuvoso}
@attribute Temperatura {Quente,Temperado,Frio}
@attribute Umidade {Alta,Normal}
@attribute Vento {Fraco,Forte}
@attribute JogarTênis {Não,Sim}

@data
Ensolarado,Quente,Alta,Fraco,Não
Ensolarado,Quente,Alta,Forte,Não
Nublado,Quente,Alta,Fraco,Sim
Chuvoso,Temperado,Alta,Fraco,Sim
Chuvoso,Frio,Normal,Fraco,Sim
Chuvoso,Frio,Normal,Forte,Não
Nublado,Frio,Normal,Forte,Sim
Ensolarado,Temperado,Alta,Fraco,Não
Ensolarado,Frio,Normal,Fraco,Sim
Chuvoso,Temperado,Normal,Fraco,Sim
Ensolarado,Temperado,Normal,Forte,Sim
Nublado,Temperado,Alta,Forte,Sim
Nublado,Quente,Normal,Fraco,Sim
Chuvoso,Temperado,Alta,Forte,Não
```

Figura 3.6: Conjunto de dados de treinamento do conceito JogarTênis no formato ARFF usado no ambiente de aprendizagem WEKA.

Após a carga do arquivo ARFF como conjunto de dados dos experimentos através do módulo de pré-processamento do WEKA, conduziu-se então a classificação do conceito JogarTênis selecionando-se o tipo de classificador desejado no módulo de classificação. As figuras 3.7, 3.8 e 3.9 apresentam o resumo dos resultados do experimento com os classificadores ID3, Naive-Bayes e *Multilayer Perceptrons*, respectivamente.

Através da análise do percentual de instâncias classificadas corretamente por cada um dos algoritmos pode-se notar que o ID3 apresentou o melhor desempenho, com 85,71% de acertos, contra 71,43% do MLP e apenas 57,14% do Naive-Bayes. Os experimentos foram conduzidos utilizando-se validação cruzada (*10-fold cross validation*).

Ao se construir curvas ROC para os classificadores e calcular-se as respectivas áreas sob as curvas obteve-se 0,8444 para o ID3, 0,7778 para o MLP e 0,5778 para o Naive-Bayes,

```

Scheme:      weka.classifiers.trees.Id3
Relation:    PlayTennis
Instances:   14

== Stratified cross-validation ==
== Summary ==

Correctly Classified Instances      12           85.7143 %
Incorrectly Classified Instances    2            14.2857 %
Kappa statistic                    0.6889
Mean absolute error                 0.1429
Root mean squared error            0.378
Relative absolute error             30 %
Root relative squared error        76.6097 %
Total Number of Instances          14

== Detailed Accuracy By Class ==

TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.8      0.111    0.8        0.8     0.8        Não
0.889    0.2       0.889     0.889  0.889     Sim

== Confusion Matrix ==

a b  <-- classified as
4 1  | a = Não
1 8  | b = Sim

```

Figura 3.7: Saída em formato texto do experimento realizado com o classificador ID3 no WEKA.

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    PlayTennis
Instances:   14

== Stratified cross-validation ==
== Summary ==

Correctly Classified Instances      8           57.1429 %
Incorrectly Classified Instances    6           42.8571 %
Kappa statistic                    -0.0244
Mean absolute error                 0.4374
Root mean squared error            0.4916
Relative absolute error             91.8631 %
Root relative squared error        99.6492 %
Total Number of Instances          14

== Detailed Accuracy By Class ==

TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.2      0.222    0.333     0.2     0.25       Não
0.778    0.8      0.636     0.778  0.7        Sim

== Confusion Matrix ==

a b  <-- classified as
1 4  | a = Não
2 7  | b = Sim

```

Figura 3.8: Saída em formato texto do experimento realizado com o classificador Naive-Bayes no WEKA.

```

Scheme:      weka.classifiers.functions.MultilayerPerceptron -L 0.3
             -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Relation:    PlayTennis
Instances:   14

== Stratified cross-validation ==
== Summary ==

Correctly Classified Instances      10          71.4286 %
Incorrectly Classified Instances    4           28.5714 %
Kappa statistic                     0.3778
Mean absolute error                  0.2871
Root mean squared error              0.5269
Relative absolute error              60.2957 %
Root relative squared error         106.7919 %
Total Number of Instances           14

== Detailed Accuracy By Class ==

TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.6      0.222    0.6        0.6     0.6        Não
0.778    0.4      0.778     0.778  0.778     Sim

== Confusion Matrix ==

a b  <-- classified as
3 2  | a = Não
2 7  | b = Sim

```

Figura 3.9: Saída em formato texto do experimento realizado com o classificador *Multilayer Perceptrons* no WEKA.

evidenciando o melhor desempenho do ID3 na tarefa e o pobre resultado do Naive-Bayes, dado que áreas inferiores a 0,6 indicam falha na classificação.

Estes resultados podem se explicados pela natureza dos atributos e a cardinalidade do conjunto de treinamento face às características de cada algoritmo. Atributos discretos e conceitos-alvo booleanos favorecem a aplicação de algoritmos baseados em árvores de decisão como o ID3, enquanto que conjuntos de treinamento com quantidade muito pequena de instâncias prejudica sensivelmente as estimativas de probabilidades a priori empregadas pelo classificador Bayesiano ingênuo.

3.4 Classificadores para Glaucoma

Vários classificadores baseados em técnicas de aprendizagem de máquina têm sido amplamente empregados em estudos sobre diagnóstico de glaucoma ao longo dos últimos anos. Goldbaum (GOLDBAUM; AL, 2002) avaliou a eficácia de determinados classificadores na tarefa de identificação precoce de perda visual baseando-se em dados obtidos através do exame de perimetria computadorizada padrão pela comparação de áreas sob curvas ROC para os classificadores MLP, *Support Vector Machines* (SVM), *Mixture of Gaussian* (MoG) e *Mixture of Generalized Gaussian* (MGG). Os resultados revelaram que MoG obteve a maior área, sendo portanto o que obteve maior acurácia na tarefa. O mesmo estudo também apontou que os médicos especialistas não obtiveram melhores

resultados ao classificar campos visuais do que os algoritmos classificadores. Nesta mesma tarefa, outro estudo (TUCKER, 2004) propõe o uso de classificadores Bayesianos espaço-temporais, utilizando os dados obtidos através do exame de campo visual e outros dados clínicos como a PIO, coletados e analisados ao longo do tempo.

A avaliação dos classificadores *Linear Discriminant Analysis* (LDA), *Classification Trees* (CTREE) e *Bagged Classification Trees* (Bagged-CTREE) para classificar glaucoma através de dados sobre espessura da camada de fibras nervosas obtidos através da aplicação do método de oftalmoscopia confocal por varredura a laser, indicou o Bagged-CTREE como ótimo para a tarefa (HOTHORN; LAUSEN, 2003). Peters et al (Y HOTHORN T, 2003) igualmente avaliaram o desempenho de classificadores Bagged-CTREE na classificação de glaucoma sobre dados de HRT. Os classificadores SVM analogamente apresentaram resultados satisfatórios na classificação de glaucoma através de parâmetros deste exame (ZANGWILL, 2004).

Burgansky-Eliash et al (BURGANSKY-ELIASH, 2005) avaliaram o emprego dos classificadores LDA, *Generalized Linear Model* (GLM), *Generalized Additive Model* (GAM), SVM e *Recursive Partitioning and Regression Tree* (RPART) na tarefa de classificação de glaucoma baseado em dados de pacientes normais e glaucomatosos coletados com o instrumento Stratus OCT. O melhor classificador foi a SVM utilizando apenas oito parâmetros, apresentando área da curva ROC de 0,981 e acurácia de 0,966.

Estudo recente (HUANG; CHEN; HUANG, 2007) avaliou o uso de técnicas baseadas em redes neurais artificiais e lógica nebulosa (*fuzzy*) e dados de espessura da camada de fibras nervosas obtidas através do instrumento Stratus OCT para classificação de glaucoma. Os resultados indicaram acurácia de 85,6%, área sob a curva ROC de 0,925 e sensibilidades de 0,867 e 0,800 para especificidades de 80% e 90% respectivamente. Utilizando igualmente dados obtidos através de Stratus OCT, Huang et al (HUANG; CHEN; LIN, 2007) avaliaram o emprego de árvores de decisão na classificação de glaucoma. Os resultados foram sensibilidade de 73%, especificidade de 92% e acurácia de 83%. O mesmo estudo apontou as regras de associação extraídas do processo pela árvore de decisão.

Estudo de Bowd et al (BOWD et al., 2008) avaliou a acurácia de algoritmos Bayesianos para classificação de glaucoma a partir de dados estruturais, obtidos através de tomografia de coerência óptica, e funcionais, obtidos através de perimetria computadorizada. Os classificadores foram treinados e testados sobre os dados estruturais e funcionais isoladamente, e em associação. Os resultados revelaram elevação marginal de desempenho quando os dados foram associados em comparação aos mesmos quando utilizados isoladamente.

4 O Gerador GLOR

Como já abordado em capítulos anteriores, para que algoritmos classificadores possam apresentar bom desempenho em tarefas de classificação, conjuntos de dados de qualidade devem ser utilizados como exemplo durante a fase de treinamento. Isto significa que os dados devem ser representativos de populações reais, e disponíveis em quantidade suficiente para que possam expressar a maior variedade de casos possível.

No entanto, uma questão relevante para a condução de estudos de métodos diagnósticos de glaucoma é a dificuldade para coleta de dados, que consome meses de trabalho e envolve custos com equipamentos, materiais e profissionais de saúde.

Diante disso, um dos objetivos deste trabalho foi desenvolver um método para geração artificial de dados representativos de populações reais para emprego no treinamento de classificadores para diagnóstico de glaucoma. Desta forma, um algoritmo inédito foi desenvolvido e denominado GLOR (*Glaucoma Oracle*).

Como ilustrado no diagrama da figura 4.1, o GLOR utiliza como dados de entrada o conjunto de características da amostra populacional que se deseja obter artificialmente como saída. Por exemplo, para obter-se uma amostra contendo os dados Idade e pressão intra-ocular de forma simulada para 1000 indivíduos com glaucoma, deve-se fornecer como parâmetros de entrada para o GLOR a cardinalidade (1000) os atributos que serão gerados (Idade e PIO) e as respectivas médias e desvios padrões para pacientes com glaucoma encontrados em populações reais (por exemplo, 68 ± 10 anos para o atributo Idade, e $27,7 \pm 9,1$ mmHg para a PIO).

O resultado do processamento o GLOR é um conjunto de dados contendo na forma de uma tabela contendo 1000 linhas e 3 colunas (ID, Idade e PIO), onde ID é um identificador numérico sequencial iniciando-se em 1 com incrementos de 1 unidade para cada registro da tabela. Os valores para os atributos Idade e PIO serão representativos da população com glaucoma, de acordo com as respectivas médias e desvios padrões informados.

Este exemplo, simples e reducionista, foi usado apenas para abordar de forma didática

o princípio de funcionamento do algoritmo criado. Na realidade o GLOR possui um nível de complexidade bem maior, sendo capaz de produzir artificialmente valores para 178 atributos utilizados no diagnóstico de glaucoma, sendo agrupados em categorias como: epidemiológicos, clínicos, estruturais e funcionais.

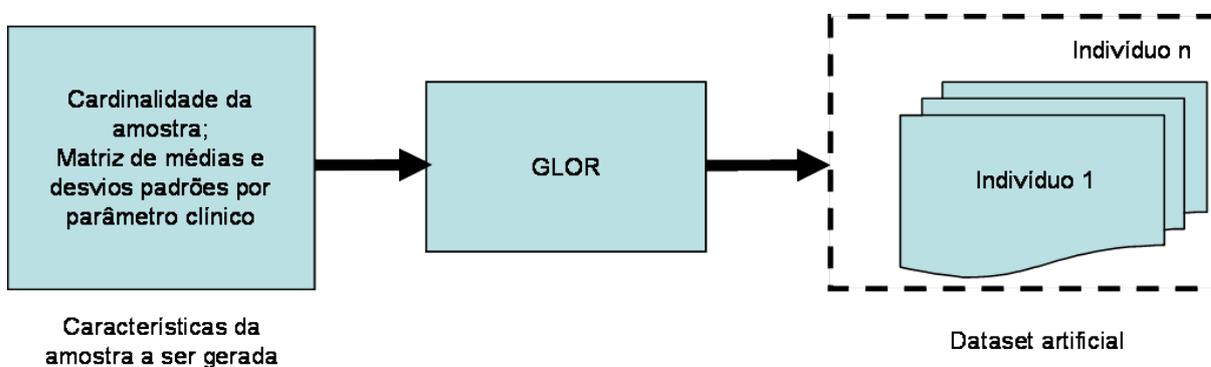


Figura 4.1: Processo de geração de amostras populacionais artificiais pelo GLOR.

O processo de geração de dados através do GLOR envolve 5 etapas:

1. Geração através do método de Monte Carlo (MC) de valores para os atributos Idade, PIO e espessuras da camada de fibras nervosas a cada 30 graus (*Clock-Hour RNFL Thickness*) de acordo com características estatísticas definidas *a priori* para populações de indivíduos normais e de indivíduos com glaucoma;
2. Geração inversa dos seguintes parâmetros de OCT: espessuras médias da camada de fibras nervosas por quadrante, espessura média total e curva de espessuras (*RNFL TSNIT Thickness*) de 256 pontos dispostos ao longo de uma circunferência imaginária com centro no disco óptico;
3. Obtenção dos bancos de dados normativos dos instrumentos OCT e SAP através de resultados de exames de pacientes reais obtidos através dos correspondentes equipamentos;
4. Geração simulada de dados de SAP através do emprego de um modelo de correlação entre estrutura e função e os bancos de dados normativos dos instrumentos OCT e SAP;
5. Produção de uma saída gráfica mimetizando os resultados apresentados pelos equipamentos OCT e SAP, para facilitar a leitura dos dados pelos médicos.

O gerador GLOR foi implementado utilizando-se a linguagem de programação Visual Basic for Applications (VBA) do Microsoft Excel.

4.1 Geração dos Atributos Idade, PIO e Espessuras Setoriais do OCT

Métodos de Monte Carlo (MC) são métodos capazes de gerar valores aleatórios de tal forma que estes valores obedecem determinadas características previamente selecionadas. Amplamente utilizados em simulação computacional, a aplicação de técnicas de Monte Carlo a domínios específicos da medicina é vasta. Canner, Chiang e Javitt (CANNER; CHIANG; JAVITT, 1992) abordaram o uso de simulação MC para estudo de retinopatia diabética. Naquele estudo, MC foi empregado com sucesso para simular probabilidades de mudanças de estados (entre estado inicial saudável para estado final doente) no estudo de uma doença oftalmológica crônica como a retinopatia diabética. Trabalho recente de Cracknell (CRACKNELL; FARNELL; GRIERSON, 2007) sobre o uso de simulação MC para gerar evidências numéricas sobre escurecimento da íris induzido por uso do medicamento *latanoprost*, empregado no tratamento de glaucoma. Neste estudo, técnicas MC foram utilizadas para simulação da propagação da luz na íris. Os resultados demonstraram que MC pode satisfatoriamente ser empregado para simulação da coloração da íris em olhos castanhos e castanho-esverdeados (*hazel*).

Dois exemplos clássicos de aplicações de métodos de Monte Carlo são a aproximação do valor de π e o cálculo de integrais de funções complexas. Para melhor compreensão do método estas 2 aplicações serão abordadas a seguir.

Tome-se por exemplo a figura 4.2, onde um quarto de circunferência de raio r está inscrito em um quadrado. Utilizando-se como referência o plano cartesiano (x,y) , as coordenadas dos vértices do quadrado ABCD na figura serão: A(0,0), B(r,0), C(r,r) e D(0,r). O centro da circunferência inscrita é o ponto A.

Define-se X como uma variável aleatória contínua no intervalo $[0, r]$ para representar valores de x , e Y como uma variável aleatória contínua no intervalo $[0, r]$ para representar valores de y .

Obtendo-se valores para x e y através de um gerador de números aleatórios uniformemente distribuídos no intervalo $[0, r]$, tem-se as coordenadas de um ponto $P(x,y)$. Se $x^2 + y^2 \leq r^2$, então $P(x,y)$ pertencerá à área interna da circunferência inscrita (região azul da figura 4.2), caso contrário pertencerá à área interna do quadrado, mas externa à circunferência.

A razão entre a quantidade total de pontos que pertencerem à região interna da circunferência (P_{circ}) e a quantidade total de pontos gerados (P_{total}) será proporcional à razão

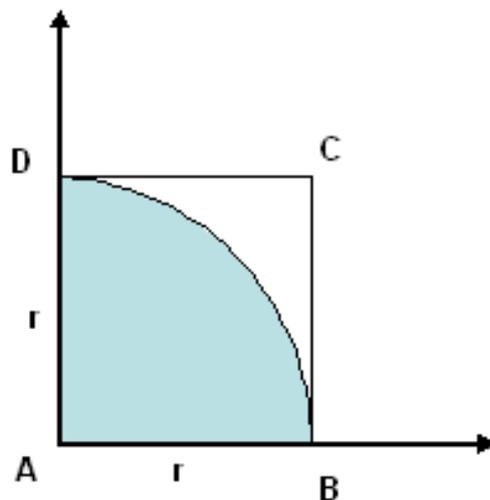


Figura 4.2: Figura de um quadrado contendo uma circunferência inscrita de raio r utilizada na aproximação do valor de π através da aplicação do método Monte Carlo.

entre as áreas das respectivas figuras se as coordenadas dos pontos forem geradas com base em geradores de números randômicos uniformemente distribuídos, ou seja, equiprováveis.

Desta forma, pode-se obter o valor aproximado de π :

$$Area_{circ} = \frac{\pi}{4}r^2$$

$$Area_{quad} = r^2$$

$$\frac{Area_{circ}}{Área_{quad}} = \frac{\pi}{4}$$

$$\frac{Area_{circ}}{Area_{quad}} = \frac{P_{circ}}{P_{total}}$$

$$\pi = 4 \frac{P_{circ}}{P_{total}}$$

A mesma estratégia pode ser empregada para se obter o valor aproximado da integral de uma função $f(x)$ entre dois pontos A(a,0) e B (b,0). Para tanto, deve-se inscrever a função em um quadrado de lado AB e, analogamente ao caso da aproximação do π , deve-se gerar aleatoriamente pontos dentro do quadrado. A quantidade de pontos interiores à região formada pela curva e o eixo X deve ser calculada, assim como a quantidade total

de pontos gerados. A razão entre as duas quantidades corresponde à razão entre as área da figura e a do quadrado. Com base nesta razão de proporcionalidade pode-se então obter de forma aproximada o valor da área sob a curva entre os pontos A e B, ou seja, o valor de $\int_a^b f(x)dx$.

No entanto, para que os valores de π e da integral de $f(x)$ possam ser aproximados com precisão adequada, grandes quantidades de pontos devem ser gerados. Além disso, a premissa de que os números aleatórios gerados devem obedecer determinadas características, expressas em termos de distribuições de probabilidades (no caso dos exemplos acima a distribuição uniforme), é de vital importância para o sucesso do método.

Assim, o presente trabalho utilizou o método Monte Carlo no desenvolvimento do gerador GLOR para simulação de valores para parâmetros utilizados no diagnóstico de glaucoma. A premissa fundamental para o sucesso do método é que os números gerados devem possuir características estatísticas de amostras populacionais reais e distribuições de probabilidade gaussianas. Tais distribuições poderiam a princípio apresentar diversas formas, tais como Rayleigh, Exponencial, Chi-quadrado, Gamma e T-student, porém a distribuição Normal é amplamente utilizada em várias áreas da ciência e da engenharia, bem como em estudos demográficos e sociais (STARK; WOODS, 2002) por representar distribuições adequadas de populações diversas em vários domínios e contextos.

Definiu-se então que um determinado parâmetro k a ser gerado pelo GLOR possui distribuição Normal, com média μ e variância σ^2 , e pode ser representado por uma variável aleatória contínua X cuja função densidade de probabilidade pode ser encontrada na equação 4.1.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (4.1)$$

Deve-se destacar que X (no exemplo a PIO) é uma variável aleatória contínua, e como tal não é possível calcular sua probabilidade exata no ponto x , mas sim a integral da respectiva curva de densidade de probabilidade nas proximidades de x , ou seja, em vez de $P[X = x]$ tem-se $P[a \leq x \leq b]$.

Tome-se como exemplo o parâmetro pressão intra-ocular ($k = PIO$), cuja média e desvio padrão para uma amostra populacional de indivíduos normais, ou seja, sem glaucoma, são 16,9 e 2,5 mmHg, respectivamente. Ao se desenhar um gráfico de frequência, e por consequência, de estimativas de probabilidades de ocorrência, encontrar-se-á uma curva como aquela que pode ser observada na figura 4.3. Para elaborá-la foi empregada

a seguinte função integrante do conjunto de funções estatísticas do Microsoft Excel:

$$DIST.NORM(x, média, desvpadrão, cumulativo) \quad (4.2)$$

Onde x é o número cujo valor da densidade de probabilidade normal ($f(x)$) se deseja obter, $média$ é o valor da média μ de $f(x)$, e $desvpadrão$ corresponde ao desvio padrão σ da distribuição.

Foram empregados como argumentos para a função 59 valores de PIO, variando de 1 a 30 mmHg, média e desvio padrão de 16,9 e 2,5 mmHg, respectivamente, e FALSO como valor para o argumento *cumulativo*. Assim, para cada valor de PIO no intervalo de 1 mmHg a 30 mmHg foi calculada sua probabilidade de ocorrência, supondo-se que a distribuição dos valores deste atributo seja Normal, com média 16,9 mmHg e desvio padrão de 2,5 mmHg.

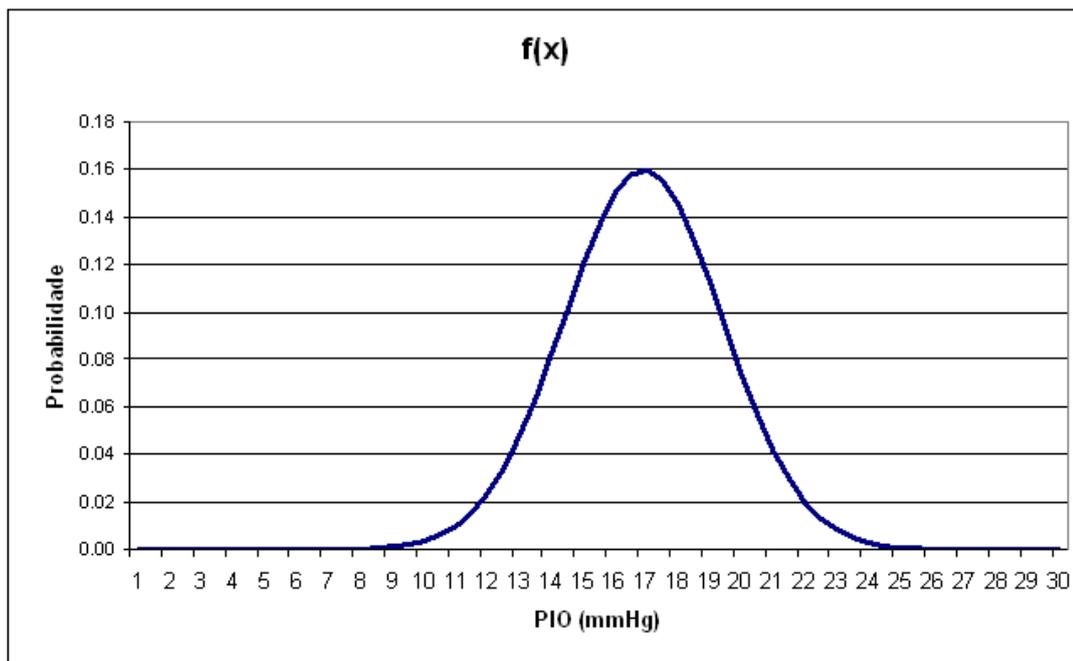


Figura 4.3: Exemplo de $f(x)$ implementada através da função *DIST.NORM*.

Uma outra forma comum de cálculo de probabilidades é o cálculo de probabilidades acumuladas. Em vez de se desejar obter a probabilidade de ocorrência do valor de pressão de 17 mmHg, pode-se desejar calcular o valor da probabilidade da pressão ser inferior a 17 mmHg. Neste caso, o que se deseja obter é a probabilidade $P[X < x] = P[X < 17] \approx 0,5$ (50%). Este valor é obtido através da integração da função densidade de probabilidade entre $-\infty$ e x , como descrito na equação 4.3.

$$F(x) = \int_{-\infty}^x f(x)dx = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (4.3)$$

A função $F(x)$ recebe o nome de distribuição de probabilidade acumulada de X . Desta forma, como X é uma variável aleatória contínua com distribuição gaussiana, para cada valor de X é possível calcular $F(x)$. A figura 4.4 exibe o gráfico de $F(x)$, ou seja, o valor da probabilidade acumulada de X . Para gerar o gráfico foi utilizada novamente a função *DIST.NORM*, porém com o valor VERDADEIRO para o parâmetro *cumulativo*. Pode-se notar no gráfico resultante que, para um dado valor de pressão intra-ocular, existe um único valor de probabilidade acumulada, e para cada valor de probabilidade acumulada existe apenas um valor de pressão correspondente. Diante disso, a função $F(x)$ possui inversa, ou seja, existe a função $F^{-1}(x)$. Esta propriedade pode ser explorada para a geração inversa de valores com distribuição normal a partir de valores de probabilidade.

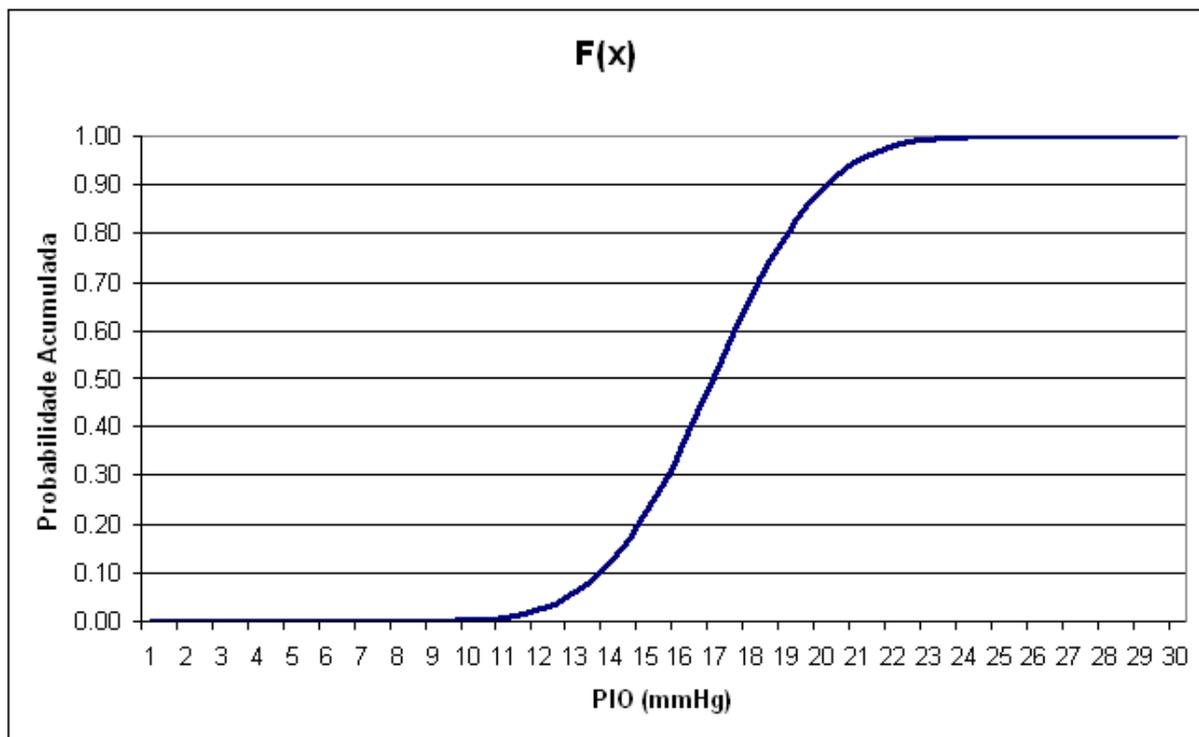


Figura 4.4: Exemplo de $F(x)$ implementada através da função *DIST.NORM*.

A observação da curva Normal dos valores de PIO revela que para um dado valor de pressão existe um valor da estimativa de sua probabilidade de ocorrência na população avaliada. Por exemplo, o valor da probabilidade para a PIO de 17 mmHg é aproximadamente 0,16 (16%), enquanto que a probabilidade do valor 13 mmHg para a pressão é pouco superior a 0,04 (4%). No entanto, diante do formato da curva Normal, para um dado valor de probabilidade existem dois valores de pressão, ambos simétricos em relação

à média da distribuição. Por esta razão, a função gaussiana não possui inversa.

Porém, a questão da geração de dados envolvida no GLOR é obter um conjunto de dados para PIO de forma que a média μ_{PIO} e desvio padrão σ_{PIO} deste conjunto sejam, quando calculados, aproximadamente 16,9 e 2,5 mmHg, respectivamente. Portanto, o problema é justamente inverso ao do cálculo da probabilidade dado um valor para a PIO. O que se deseja agora é obter um valor de PIO que apresente uma dada probabilidade de ocorrência.

Supondo que Y seja uma variável aleatória uniformemente distribuída no intervalo $[0, 1]$, para um dado valor y de Y pode-se obter $F^{-1}(y)$. Como $F^{-1}(y)$ é a inversa de $F(x)$, então ao se calcular $F^{-1}(y)$ obtém-se o valor de x correspondente. No exemplo em questão, a variável aleatória Y representa valores de probabilidade acumulada e, portanto, pertencentes ao intervalo $[0, 1]$. Logo, através do cálculo de $F^{-1}(y)$ obtém-se o valor da pressão intra-ocular (x) cuja probabilidade acumulada é y .

Para se obter tais valores será necessário calcular a função inversa de $F(x)$, expressa na equação 4.3. Tal função inversa pode ser implementada através da função:

$$INV.NORM(y, \text{média}, \text{desvpadrão}) \quad (4.4)$$

Analogamente a *DIST.NORM*, *INV.NORM* também é integrante do conjunto de funções estatísticas do Excel. Entre os argumentos desta função, y representa o valor (entre 0 e 1) da probabilidade acumulada, e *média* e *desvpadrão* representam a média e o desvio padrão da amostra populacional cujo valor correspondente à probabilidade de ocorrência y se deseja calcular.

A título de exemplo, as probabilidades acumuladas obtidas na geração do gráfico da figura 4.4 foram utilizadas como valores para o argumento y em 4.4, juntamente com os valores 16,9 e 2,5 para os argumentos *média* e *desvpadrão*, respectivamente. Desta forma, os valores originais de PIO entre 1 e 30 mmHg puderam ser inversamente obtidos. A figura 4.5 apresenta o gráfico valores de PIO por probabilidade acumulada.

Como se pode notar, a partir de distribuições de probabilidades é possível obter-se inversamente valores para parâmetros com uma aproximação razoável. No exemplo em questão, a média após a geração inversa de 59 valores de PIO foi de 15,5 mmHg e desvio padrão de 8,58 mmHg. Apesar da média ter convergido satisfatoriamente para o valor esperado de 16,9 mmHg, no entanto o desvio padrão apresentou-se elevado frente aos 2,5 mmHg esperados. Porém, quanto maior a cardinalidade da amostra gerada, maior

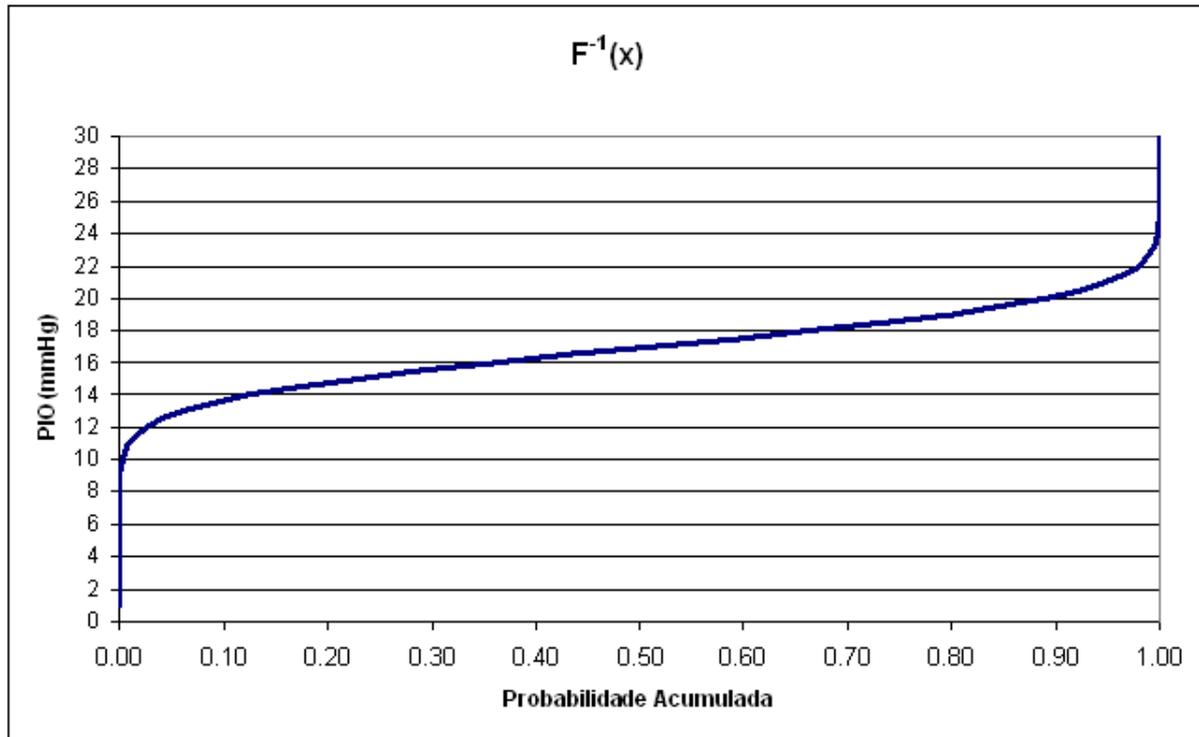


Figura 4.5: Exemplo de $F^{-1}(y)$ implementada através da função *INV.NORM*.

a tendência de convergência. Tal fato poderá ser observado no próximo capítulo deste trabalho, onde conjuntos de dados com número bem maior de instâncias foram gerados através da aplicação deste método.

No exemplo apresentado, as probabilidades acumuladas eram previamente conhecidas, e foram fornecidas como argumentos para a função *INV.NORM*. No entanto, para se gerar randomicamente valores aleatórios de x_k , basta utilizar-se como probabilidades acumuladas números pseudo-aleatórios y_{RAND} com uniformemente distribuídos, pertencentes ao intervalo $0 < y < 1$. Para esta geração de números pseudo-aleatórios foi empregada a função *RAND* do Excel.

Sendo assim, para se gerar um conjunto de dados $D = \{x_1, x_2, \dots, x_m\}$ com m atributos e n tuplas, ou seja, com grau m e cardinalidade n , deve-se aplicar a função *INV.NORM* para cálculo dos atributos $x_{i,j}$ para cada uma das n tuplas de D . Formalmente tem-se a expressão 4.5.

$$x_{i,j} = F^{-1}(y_{i,j}) \quad (4.5)$$

Onde i ($1 \leq i \leq n$) representa as tuplas (linhas), j ($1 \leq j \leq m$) representa os atributos (colunas) e $y_{i,j}$ é um número aleatório com distribuição uniforme no intervalo

$$0 < y_{i,j} < 1.$$

Neste trabalho, o método Monte Carlo ora descrito foi utilizado para gerar um conjunto de dados contendo 14 (m) parâmetros ($k_1 \dots k_{14}$), a saber: Idade, PIO, e *clock-hours RNFL Thickness*, de acordo com médias e desvios padrões desejados para cada atributo. Os valores das médias ($\mu_{k_1 \dots \mu_{k_{14}}}$) e desvios padrões ($\sigma_{k_1 \dots \sigma_{k_{14}}}$) para indivíduos normais e com glaucoma foram obtidos na literatura médica, através dos estudos de Leske (LESKE, 1994) e Medeiros (MEDEIROS, 2004). Tais valores são apresentados na tabela 4.1 e foram utilizados para se obter dois conjuntos de dados: um contendo somente dados de indivíduos normais e outro contendo somente dados de indivíduos com glaucoma. Ambos foram unidos para formar um único conjunto ao final do processo de geração.

Tabela 4.1: Lista de atributos e correspondentes médias e desvios padrões em população normal e com glaucoma.

Atributo	Normal ($\mu \pm \sigma$)	Glaucoma ($\mu \pm \sigma$)
Idade	65,0 \pm 8,0 ^[1]	68,0 \pm 10,0 ^[1]
PIO	16,9 \pm 2,5 ^[1]	27,7 \pm 9,1 ^[1]
Espessura à 1 hora de relógio	109,4 \pm 21,5 ^[2]	87,1 \pm 28,2 ^[2]
Espessura às 2 horas de relógio	92,9 \pm 25,1 ^[2]	70,5 \pm 23,6 ^[2]
Espessura às 3 horas de relógio	64,3 \pm 19,3 ^[2]	48,7 \pm 12,2 ^[2]
Espessura às 4 horas de relógio	76,6 \pm 19,9 ^[2]	58,9 \pm 17,4 ^[2]
Espessura às 5 horas de relógio	106,1 \pm 25,6 ^[2]	77,3 \pm 24,5 ^[2]
Espessura às 6 horas de relógio	134,7 \pm 23,7 ^[2]	89,2 \pm 29,3 ^[2]
Espessura às 7 horas de relógio	132,2 \pm 21,4 ^[2]	88,2 \pm 32,9 ^[2]
Espessura às 8 horas de relógio	69,3 \pm 18,5 ^[2]	62,4 \pm 18,5 ^[2]
Espessura às 9 horas de relógio	52,8 \pm 11,5 ^[2]	52,9 \pm 12,7 ^[2]
Espessura às 10 horas de relógio	81,8 \pm 19,0 ^[2]	70,6 \pm 21,0 ^[2]
Espessura às 11 horas de relógio	127,7 \pm 17,9 ^[2]	97,1 \pm 31,1 ^[2]
Espessura às 12 horas de relógio	122,0 \pm 26,5 ^[2]	95,4 \pm 29,9 ^[2]

Legenda: [1] Leske (LESKE, 1994); [2] Medeiros (MEDEIROS, 2004)

Iniciando-se com características estatísticas de uma população de indivíduos normais, o valor do primeiro atributo (Idade) para a primeira instância do conjunto representativo da população desejada, μ_{Idade} e σ_{Idade} , foram fornecidos como argumentos para uma função matemática que implementa $F^{-1}(x)$. A função $INV.NORM(y, média, desvpadrão)$ foi empregada para gerar um número que pertence a uma distribuição normal acumulada, com média $média$ e desvio padrão $desvpadrão$. Um número aleatório uniformemente distribuído foi obtido através da função $RAND$ e fornecido como o argumento y para $INV.NORM$. Assim, $x_{1,1}$ pode ser calculado, ou seja, o valor do primeiro atributo da primeira instância da amostra.

Este procedimento deve ser repetido outras 13 vezes para obter-se os valores para

os demais atributos $(x_{1,2}...x_{1,14})$ e assim finalizar o cálculo dos 14 parâmetros para a primeira instância do conjunto. Uma vez obtida esta primeira instância, o processo deve ser repetido outras $n - 1$ vezes para finalizar a geração da primeira amostra, contendo dados representativos de indivíduos normais. Em seguida, o método deve novamente ser empregado para gerar a amostra populacional com dados representativos de indivíduos com glaucoma, utilizando-se para tal agora as características estatísticas da população com glaucoma.

Uma vez obtidos os conjuntos de dados contendo os 14 atributos já descritos, em seguida pode-se iniciar a fase de geração dos demais parâmetros do OCT.

4.2 Geração dos Demais Parâmetros de OCT

Após a geração dos parâmetros Idade, PIO e as espessuras médias da camada de fibras nervosas a cada setor de 30 graus (*Clock-Hour thickness*) procede-se à geração dos seguintes dados:

1. Espessuras médias por quadrante.
2. Espessura média geral.
3. Espessuras por ponto ao redor de uma circunferência com centro no disco óptico.

Na figura 4.6 pode-se identificar as espessuras médias por setor (*Clock-Hours*), quadrantes Temporal, Superior, Nasal e Inferior (*Quadrants*) e espessura média total (*Thickness Average*) obtidas através da saída gráfica do instrumento Cirrus HD-OCT. As cores vermelho, amarelo, verde e branco que preenchem as imagens representam intervalos de confiança dos dados, e serão abordados mais adiante, na seção sobre bancos de dados normativos.

No processo empregado para geração destes dados pelo instrumento Cirrus HD-OCT, primeiramente são realizadas 200 varreduras verticais e 200 horizontais em torno de um quadrado imaginário com centro no disco óptico para obtenção das medidas de espessura da camada de fibras nervosas. A partir daí então são gerados os dados para compor o gráfico de espessuras. Neste gráfico, para cada um dos 256 pontos avaliados é plotado no espaço bidimensional o valor da espessura correspondente. Cada ponto representa a espessura de um setor angular de aproximadamente $1^{\circ}24'$ de uma circunferência imaginária com centro no disco óptico.

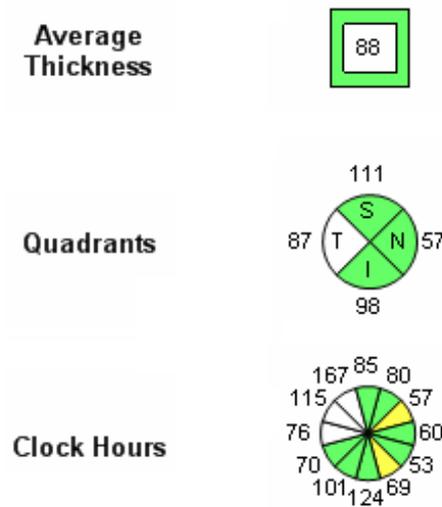


Figura 4.6: Espessuras médias por setor, quadrantes e total obtidas através da saída gráfica do instrumento Cirrus HD-OCT.

A partir da obtenção das medidas por ponto, pode-se calcular a espessura média por setor de 30° (*Clock-Hour*) através da média das espessuras correspondentes aos pontos que integram cada setor.

Analogamente, uma vez obtida a espessura por setor, pode-se calcular as espessuras por quadrante. São 4 os quadrantes de interesse: Temporal (T), Superior (S), Nasal (N) e Inferior (I). Assim, a média entre os setores correspondentes às 8, 9 e 10 horas de relógio equivale à média do quadrante T. A média das espessuras em 11, 12 e 1 horas correspondem à média do quadrante S. A média dos setores de 2, 3 e 4 horas corresponde à média do setor N, enquanto que a média das espessuras dos setores 5, 6 e 7 horas representa a média do quadrante I. Na figura 4.7 pode-se observar graficamente estas correspondências.

Finalizando, para se obter a espessura média total calcula-se a média entre os quadrantes. Formulando matematicamente estes cálculos tem-se:

$$T_T = (T_{8h} + T_{9h} + T_{10h})/3$$

$$T_S = (T_{11h} + T_{12h} + T_{1h})/3$$

$$T_N = (T_{2h} + T_{3h} + T_{4h})/3$$

$$T_I = (T_{5h} + T_{6h} + T_{7h})/3$$

$$T_{AVG} = (T_T + T_S + T_N + T_I)/4$$

Onde $T_{1h}...T_{12h}$ são as espessuras a cada hora, T_T, T_S, T_N, T_I são as espessuras por quadrante e T_{AVG} é a espessura média total.

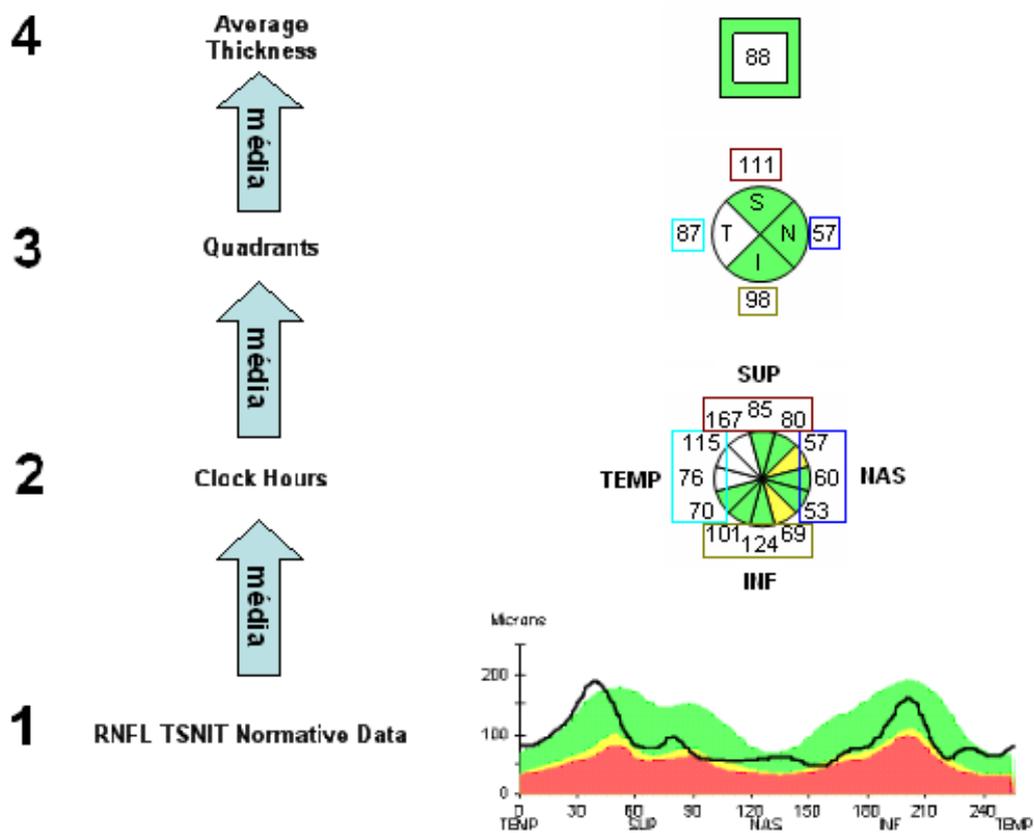


Figura 4.7: Processo de obtenção das médias por setor, quadrantes e total a partir dos dados discretos que compõem o gráfico de espessuras da camada de fibras nervosas.

No GLOR o processo de geração inicia-se pela obtenção inversa dos valores das espessuras por setor de acordo com respectivas médias e desvios padrões identificados na literatura. Em seguida as espessuras por quadrante e a espessura média total podem ser obtidas. A figura 4.8 ilustra este processo.

Para obter os dados da curva de espessuras é necessário uma interpolação linear para que o gráfico possa ser construído, uma vez que somente estão disponíveis 12 valores que

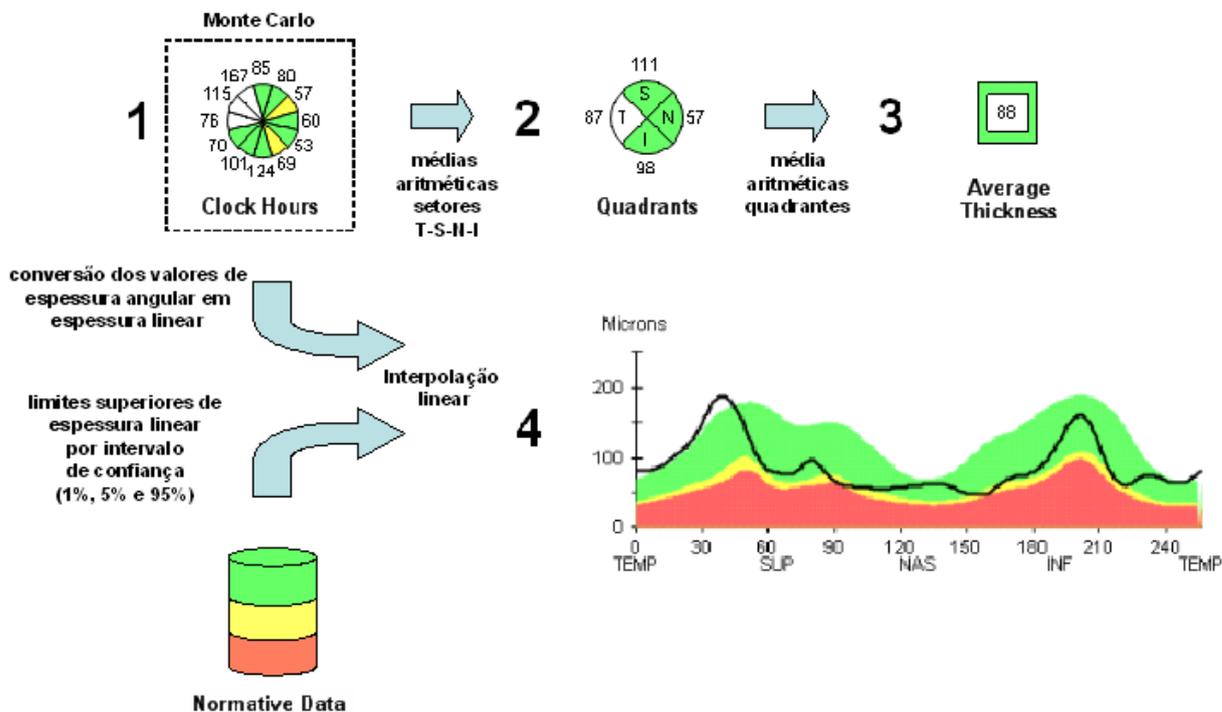


Figura 4.8: Resumo das 4 etapas envolvidas na geração artificial inversa dos dados de Cirrus HD-OCT.

correspondem artificialmente às espessuras médias por setor de 30 graus. Como este é o menor nível de granularidade de espessuras que pode ser gerado, daí a necessidade da interpolação para efeitos de exibição da respectiva curva.

4.3 Bancos de Dados Normativos dos Instrumentos OCT e SAP

Por questões de sigilo industrial, as empresas fabricantes dos instrumentos, não disponibilizam livremente os bancos de dados normativos por faixa etária de cada um de seus instrumentos, o que ocorre com o Cirrus HD-OCT e Humphrey HFAII do fabricante Carl Zeiss Meditec, Inc. Desta forma, fez-se necessário desenvolver alternativas para obtenção destes dados.

4.3.1 Reconstrução do Banco de Dados Normativo do Cirrus HD-OCT

Dados reais de um indivíduo voluntário foram adquiridos utilizando-se o instrumento Cirrus HD-OCT para se obter os valores normativos de espessuras por intervalo de confiança e faixa etária. Os dados foram extraídos a partir das saídas gráficas (*printouts*) do equi-

pamento. Na figura 4.9 tem-se destacado o gráfico de espessuras, que foi utilizado na obtenção dos dados normativos.

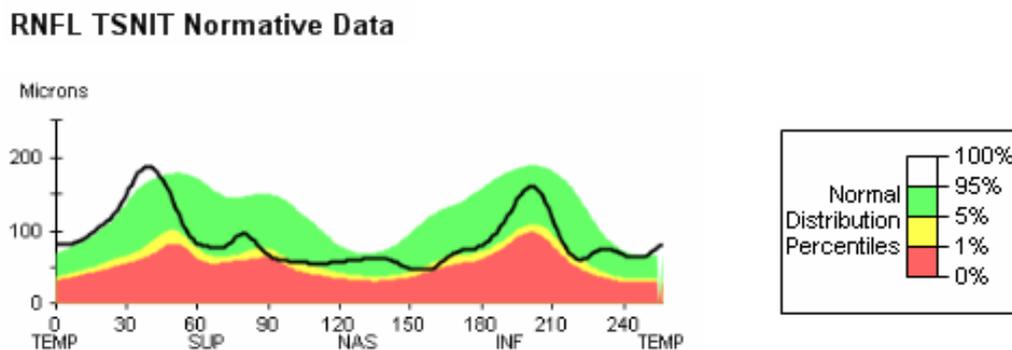


Figura 4.9: Gráfico normativo das espessuras da camada de fibras nervosas (*RNFL Thickness Normative Data*) de um indivíduo voluntário obtido com o instrumento Cirrus HD-OCT.

Deve-se notar que neste gráfico existem 4 curvas diferenciadas por cores. A curva em preto é a que apresenta as espessuras em microns para cada um dos 256 pontos obtidos ao longo de uma circunferência imaginária com centro no disco óptico do paciente que foi avaliado. As outras 3 curvas, exibidas em vermelho, amarelo e verde, representam os intervalos de confiança de 1%, 5% e 95%, respectivamente. A função destas curvas é possibilitar ao oftalmologista a avaliação da intersecção de pontos da curva de espessuras do paciente com as curvas que indicam espessuras encontradas em 1%, 5% ou 95% da população normal. Na condição de uma retina saudável, ou seja, sem perdas significativas além das naturais, ocorridas com a idade, a curva preta deve possuir todos os seus pontos dentro da curva verde ou acima dela (área branca). Por outro lado, se a curva de espessuras do paciente possuir muitos pontos de intersecção com as curvas amarela e vermelha, aumentara a probabilidade do oftalmologista classificar os resultados do exame como positivos para glaucoma.

Para se obter através do gráfico os valores limiares dos intervalos de confiança, pode-se desenhar uma linha vertical, paralela ao eixo das espessuras, passando pelo ponto desejado no eixo horizontal e obter então os pontos de intersecção desta reta com as curvas vermelha, amarela e verde. Na figura 4.10 é possível observar os pontos A, B e C que indicam as intersecções da reta vertical que passa pelo ponto 60 com as curvas normativas dos intervalos de confiança de 1%, 5% e 95% respectivamente. Em cada ponto de intersecção obtém-se uma reta ortogonal à primeira, gerando-se então 3 novas retas. A intersecção destas retas ortogonais com o eixo das espessuras determina 3 novos pontos

E_A , E_B e E_C . A leitura destes valores na escala do gráfico produzirá então os valores dos limiares desejados.

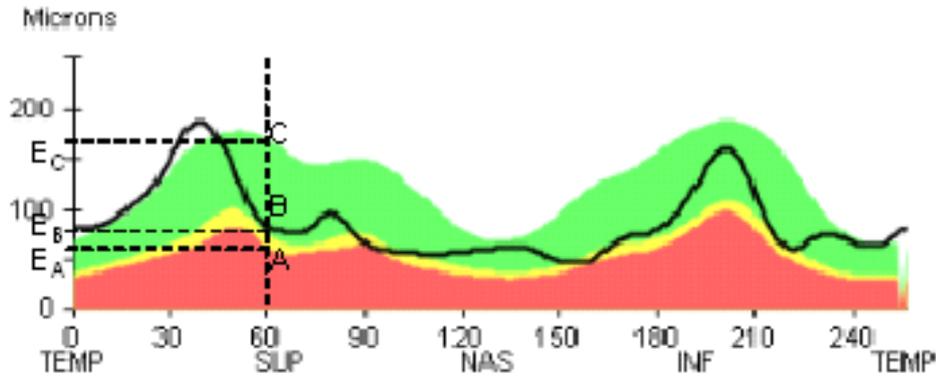


Figura 4.10: Método para identificação dos limiares de espessura por intervalo de confiança.

No presente trabalho, as imagens das saídas gráficas foram digitalizadas através de uma função específica do software do Cirrus HD-OCT e exportadas para arquivos no formato TIFF (*Tagged Image File Format*). Em seguida, com o auxílio do software editor de imagens Photoshop, as imagens foram individualmente carregadas. Uma vez identificado o gráfico normativo de espessuras da camada de fibras nervosas (*RNFL Thickness Normative Data*), obteve-se a leitura da medida em pixels da altura do segmento de reta iniciando no ponto 0 e finalizando no ponto limiar da primeira faixa do intervalo de confiança de 1% (curva vermelha). Registrou-se então o comprimento deste segmento em quantidade de pixels em uma planilha. Em seguida obteve-se igualmente a leitura em pixels do comprimento do segmento de reta que inicia-se no ponto 0 e encerra-se no ponto que representa o limiar do intervalo de confiança de 5% (curva amarela). Esta quantidade de pixels foi igualmente registrada. Depois disso, obteve-se o comprimento em quantidade de pixels do segmento de reta que inicia-se no ponto 0 e encerra-se no ponto limiar do intervalo de confiança de 95%, registrando-se o resultado.

Uma vez encontrados os valores dos limiares dos intervalos de confiança para o ponto 0, expressos em quantidade de pixels, repete-se o processo para outros 26 pontos, de 10 em 10, finalizando-o com o ponto 255. Uma vez obtidos os valores dos limiares em números de pixels, um fator de conversão deve ser obtido para transformá-los em microns. Para tanto foi obtido o comprimento em pixels do segmento de reta determinado pelas coordenadas (0,0) e (0,200). Como o comprimento deste segmento de reta em pixels corresponde a

200 microns de acordo com a escala do gráfico, o fator de conversão pode ser calculado. Supondo, por exemplo, que o comprimento para este segmento de reta seja 400 pixels, então o fator de conversão utilizado no gráfico em questão será de $200/400$, ou seja, 0,5 micron/pixel.

Para se reproduzir o banco de dados normativos para diversas idades, a seguinte abordagem foi empregada:

1. Adquiriu-se as medidas de espessura para um indivíduo voluntário.
2. Alterou-se o campo data de nascimento de forma a simular inicialmente a idade de 25 anos (início da primeira faixa etária alvo).
3. Acionou-se o algoritmo de análise do software do instrumento.
4. Exportou-se a imagem do resultado gerado pela análise para um arquivo em formato TIFF.
5. Repetiu-se os passos 2,3 e 4 para idades simuladas de 30 a 90 anos, em intervalos de 5 anos.

Ao final do processo acima, 14 arquivos foram gerados contendo os dados de um mesmo indivíduo, porém avaliados pelo software do instrumento para idades variando de 25 a 90 anos. Desta forma, o banco de dados normativo do Cirrus HD-OCT pode ser indiretamente reproduzido através das curvas normativas observadas nos resultados de cada análise refletem as curvas normativas do instrumento para as idades desejadas.

Na figura 4.11 tem-se uma imagem-exemplo da planilha com os valores normativos para faixas etárias de 25 a 29 anos, de 30 a 34 anos, de 35 a 39 anos e de 40 a 44 anos.

4.3.2 Obtenção do Banco de Dados Normativo do HFA-II

Assim como no método desenvolvido para obter o banco de dados normativo do Cirrus HD-OCT, o banco de dados normativo do instrumento HFA-II também foi obtido através de resultados de exames. Porém neste caso foram utilizados dados de 82 pacientes normais coletados com o instrumento Humphrey Field Analyzer HFA-II (Carl Zeiss Meditec, Inc). O programa de testes empregado foi C30-2 *full-threshold* e o software STATPAC para as análises. Na figura 4.12 pode-se observar um exemplo de saída gráfica de um resultado de exame de campo visual computadorizado ou perimetria acromática padrão. Os dados destacados em vermelho foram aqueles utilizados para compor o banco de dados normativo do SAP para o GLOR.

Ponto	Faixas Etárias											
	25			30			35			40		
0	33.33	36.36	69.70	30.30	36.36	69.70	33.33	36.36	69.70	30.30	36.36	69.70
10	39.39	45.45	84.85	39.39	45.45	84.85	39.39	45.45	84.85	39.39	42.42	84.85
20	48.48	57.58	106.06	48.48	57.58	109.09	45.45	57.58	103.03	45.45	54.55	106.06
30	57.58	66.67	142.42	57.58	66.67	142.42	54.55	69.70	145.45	54.55	66.67	139.39
40	69.70	84.85	175.76	69.70	84.85	175.76	69.70	84.85	175.76	66.67	84.85	175.76
50	90.91	106.06	187.88	87.88	106.06	184.85	84.85	103.03	184.85	84.85	103.03	184.85
60	66.67	81.82	181.82	66.67	81.82	178.79	66.67	78.79	175.76	63.64	78.79	175.76
70	60.61	69.70	127.27	60.61	69.70	154.55	60.61	66.67	154.55	57.58	66.67	154.55
80	63.64	72.73	154.55	63.64	72.73	154.55	60.61	72.73	151.52	60.61	69.70	151.52
90	66.67	81.82	157.58	66.67	78.79	157.58	66.67	78.79	154.55	66.67	75.76	154.55
100	54.55	66.67	145.45	51.52	63.64	142.42	51.52	63.64	142.42	51.52	60.61	139.39
110	45.45	51.52	115.15	42.42	51.52	115.15	42.42	54.55	115.15	39.39	51.52	115.15
120	33.33	45.45	84.85	30.30	45.45	84.85	33.33	45.45	84.85	33.33	42.42	81.82
130	30.30	36.36	66.67	33.33	45.45	84.85	30.30	36.36	66.67	33.33	39.39	69.70
140	27.27	36.36	72.73	27.27	39.39	72.73	30.30	36.36	72.73	33.33	36.36	72.73
150	33.33	42.42	93.94	36.36	45.45	93.94	33.33	42.42	90.91	36.36	42.42	93.94
160	45.45	57.58	127.27	42.42	54.55	124.24	45.45	57.58	124.24	45.45	57.58	124.24
170	60.61	72.73	142.42	57.58	66.67	142.42	54.55	66.67	139.39	57.58	66.67	139.39
180	63.64	75.76	166.67	63.64	75.76	163.64	60.61	75.76	254.55	60.61	72.73	160.61
190	81.82	90.91	184.85	81.82	90.91	184.85	78.79	90.91	181.82	75.76	84.85	178.79
200	103.03	115.15	200.00	103.03	112.12	200.00	100.00	109.09	196.97	100.00	109.09	193.94
210	93.94	109.09	196.97	90.91	106.06	193.94	96.97	106.06	193.94	87.88	103.03	190.91
220	60.61	72.73	169.70	57.58	72.73	163.64	57.58	72.73	163.64	54.55	69.70	163.64
230	42.42	51.52	115.15	39.39	51.52	112.12	36.36	48.48	109.09	39.39	48.48	109.09
240	30.30	36.36	75.76	30.30	36.36	78.79	27.27	36.36	78.79	30.30	36.36	75.76
250	27.27	33.33	60.61	27.27	33.33	66.67	30.30	33.33	66.67	30.30	33.33	66.67
255	27.27	33.33	66.67	27.27	33.33	66.67	30.30	33.33	69.70	30.30	33.33	66.67

Figura 4.11: Valores normativos (μm) por intervalo de confiança e faixa etária para 27 pontos obtidos através da digitalização das saídas gráficas do Cirrus HD-OCT.

Os seguintes dados foram obtidos:

1. Idade (item A da figura 4.12);
2. 74 valores em dB de sensibilidade relativa do mapa de campo visual (item B da figura 4.12);
3. 74 valores em dB do gráfico *total deviation* (item C da figura 4.12);
4. 74 valores em dB do gráfico *pattern deviation* (item D da figura 4.12);
5. Valor do MD (item E da figura 4.12) e
6. Valor do PSD (item E da figura 4.12).

Os dados de cada um dos 82 indivíduos foram digitados em uma planilha. Os resultados foram então agrupados por faixa etária, de 30 a 80 anos, em intervalos de 10 anos. A média e desvio-padrão para cada um dos valores foi calculada para cada faixa etária.

Todavia, no método GLOR optou-se por gerar dados de SAP que reproduzissem a estratégia de teste C24-2 *SITA Standard*, mais comum e rápida por testar menos pontos do campo visual e não realizar medidas em duplicidade. Esta estratégia se utiliza da medição de sensibilidades em 52 posições (54 pontos menos 2 relativos ao ponto cego ou *blind spot*) do campo visual contra 74 (76 pontos menos 2 relativos ao ponto cego) do programa C30-2. Desta forma os pontos mais periféricos testados pela estratégia C30-2 e não utilizados pela estratégia C24-2, indicados na figura 4.13, serão desconsiderados no processo de geração de dados conduzida pelo GLOR. Pode-se observar nesta figura a indicação dos pontos periféricos (em vermelho) que não são avaliados na estratégia C24-2. Na figura também é possível identificar (em azul) o ponto cego (*blind spot*). A presença de uma segunda medida, exibida entre parênteses, apenas em alguns pontos, é uma característica da estratégia *full-threshold*, onde estes valores adicionais representam uma segunda medida repetida no ponto determinado.

A diferença entre um valor de sensibilidade relativa em um determinado ponto do mapa de campo visual e o valor normal para a idade no mesmo ponto é exibida no gráfico *total deviation*. Retomando o exemplo da figura 4.12, o valor de sensibilidade no mapa de campo visual (item B da figura) do primeiro ponto mais periférico no quadrante superior esquerdo obtém-se o valor de 23 dB. Ao se observar o ponto correspondente a este, ou seja, o primeiro ponto mais periférico no quadrante superior esquerdo, no gráfico *total deviation* (item C da figura) o valor de 1 dB será identificado. Isto significa que naquele dado ponto do campo visual do indivíduo em questão, seu limiar de sensibilidade relativa foi de 23 dB, sendo que este valor está 1dB acima do esperado para um indivíduo da mesma faixa etária,

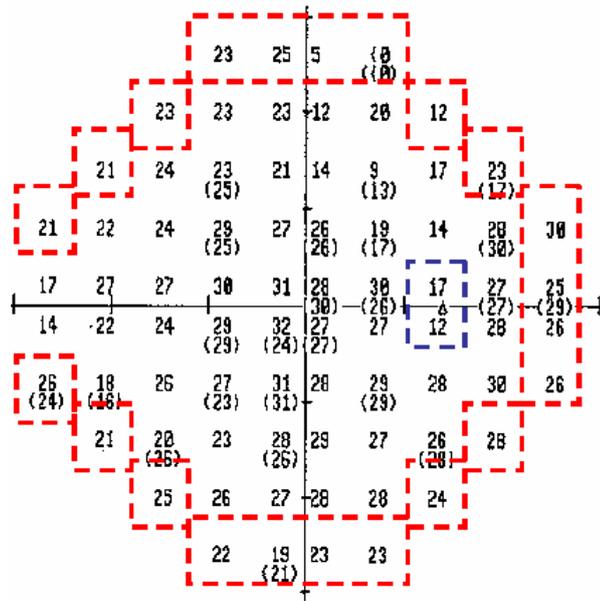


Figura 4.13: Mapa de campo visual obtido através do programa C30-2 *full threshold*.

de acordo com o banco de dados normativo do instrumento. Observando como exemplo agora o primeiro ponto da segunda linha do mapa de campo visual, imediatamente abaixo do ponto anterior, o valor de 23 dB também será obtido para este novo ponto. Porém, ao identificar-se o ponto correspondente no gráfico *total deviation*, o valor de -2 dB será encontrado. Isto significa que neste novo ponto a sensibilidade relativa do indivíduo é 23 dB, porém em comparação com um indivíduo na mesma faixa etária este número se encontra 2 dB abaixo do esperado.

Assim, o gráfico *total deviation* apresenta os valores em dB das diferenças ponto a ponto entre o valor de sensibilidade medido e o esperado de acordo com o banco de dados normativo de acordo com a idade.

Já os valores do gráfico *pattern deviation* são obtidos através de um processo de normalização. Neste processo, inicialmente ocorre a ordenação decrescente de todos os valores do gráfico *total deviation*, formando um vetor de pontos. Em seguida o valor do sétimo elemento deste vetor é usado como valor de referência para normalização. Este valor de referência é então subtraído de cada ponto do gráfico *total deviation* para gerar o seu correspondente no *pattern deviation*.

Os índices globais *Mean Deviation* (MD) e *Pattern Standard Deviation* (PSD) são obtidos através dos gráficos *total deviation* e *pattern deviation*, respectivamente. O MD corresponde à média aritmética entre os pontos do *total deviation*, representando então o desvio médio de sensibilidade em relação à população normal na mesma idade do in-

divíduo avaliado. Por sua vez o PSD corresponde ao desvio padrão entre os valores do gráfico *pattern deviation*, representando então o desvio padrão apresentado pelos dados normalizados.

No cálculo do MD, os pontos mais centrais do gráfico *total deviation* recebem mais peso do que os pontos mais periféricos. A função deste índice é indicar a redução geral de sensibilidade. Números próximos a zero indicam um exame normal, enquanto que números negativos indicam redução geral de sensibilidade. Já a função do índice PSD é oferecer ao especialista a medida de variabilidade da sensibilidade, indicando a possível existência de defeitos localizados no campo visual.

Como os valores discretos ponto a ponto dos gráficos de probabilidades não são exibidos na saída gráfica de um resultado de SAP, tais valores não puderam ser extraídos dos *printouts* para emprego no GLOR. Para resolver esta questão, as médias e desvios padrões por faixa etária para cada ponto dos gráficos *total deviation* e *pattern deviation* foram calculadas para obtenção da probabilidade. Para tanto a distribuição de probabilidade foi assumida Normal e pode ser então calculada. Na implementação do GLOR, realizada utilizando o software Excel da Microsoft, utilizou-se a função *DIST.NORM*(x , *média*, *desvpadrão*, *cumulativo*). Esta função utiliza-se de 4 parâmetros, a saber: o número x cuja probabilidade se deseja obter, a média *média*, o desvio padrão *desvpadrão* da população e *cumulativo*, parâmetro booleano cujo valor VERDADEIRO indicará o retorno da probabilidade acumulada e FALSO retornará o valor da densidade de probabilidade. No GLOR, como se deseja a probabilidade acumulada, forneceu-se VERDADEIRO para o parâmetro *cumulativo* para cálculo de probabilidades para os gráficos *total deviation probability* e *pattern deviation probability*.

Uma vez reproduzidos os bancos de dados normativos do OCT e SAP, a próxima fase do processo envolve a correlação matemática entre as espessuras da camada de fibras nervosas (estrutura) e pontos do mapa de campo visual (função). Desta forma poder-se-á simular dados de SAP com base nos dados gerados de OCT.

4.4 A Geração de Dados de SAP a partir de Dados de OCT

Como já foi abordado em capítulo anterior deste trabalho, os danos às estruturas oculares precedem os danos no campo visual no processo de evolução do glaucoma.

De acordo com o estudo de Hood et al (HOOD et al., 2007), o relacionamento entre

perda de espessura da camada de fibras nervosas e a correspondente redução de sensibilidade relativa no campo visual é logarítmica, embora o relacionamento entre espessura e sensibilidade relativa seja linear.

Diante disso, é possível estimar matematicamente a redução de sensibilidade relativa de um determinado ponto do campo visual a partir de uma taxa de redução de espessura do setor correspondente no disco óptico.

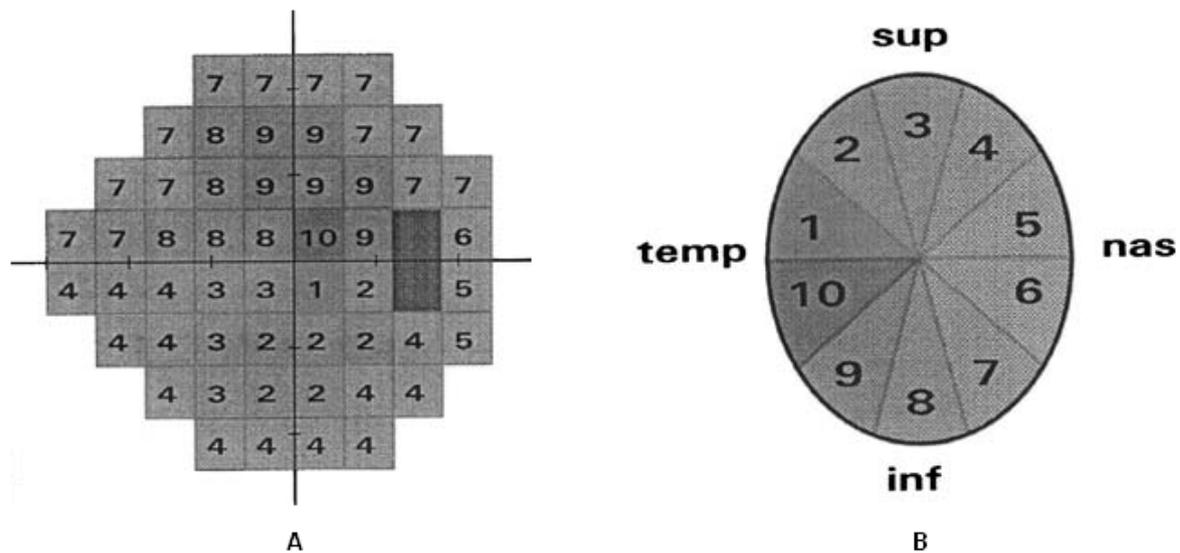


Figura 4.14: Mapeamento entre pontos do campo visual (A) e setores da cabeça do nervo óptico (B). Adaptado de (HARWERTH; WHEAT, 2008)

Estudo de Harwerth e Wheat (HARWERTH; WHEAT, 2008) apontou o relacionamento entre setores da cabeça do nervo óptico e pontos no mapa de campo visual. Tal mapeamento do tipo um-para-vários é apresentado graficamente na figura 4.14. Observando-se a imagem é possível notar que, enquanto alguns setores do nervo óptico se associam a apenas um ponto do campo visual (como os setores 1 e 10 da figura), outros são relacionados a diversos pontos (setores 4 e 7, por exemplo).

Assim sendo, é coerente supor que uma determinada perda percentual em um destes setores que se relacionam com mais de um ponto do mapa de campo visual deva ser igualmente distribuída por todos os pontos a ele relacionados. Em outras palavras, se o setor 7 do nervo óptico no modelo de Harwerth e Wheat apresentar redução de, por exemplo, 10% em sua espessura média, todos os pontos do campo visual que estão diretamente relacionados a ele apresentarão redução de 1 dB cada.

Desta forma, o método GLOR se utiliza de um modelo de correlação entre estrutura e função que se baseia nestes princípios, e pode ser assim formalmente descrito:

1. Dada a idade I de um indivíduo;
2. Dado um setor da cabeça do nervo óptico N ;
3. Dado o valor normativo $T_{NORMAL}(N, I)$ da espessura no setor N de acordo a idade I , definido como a mediana entre os limites dos intervalos de confiança de 5% e 95%;
4. Dado o valor de $T(N)$ da espessura em N ;
5. Dado a razão entre espessuras $T(N)/T_{NORMAL}(N, I)$;
6. Dado um ponto P do mapa de campo visual;
7. Dado o valor normativo $S_{NORMAL}(P, I)$ de sensibilidade normativo em P de acordo com a idade;
8. Dada a função de mapeamento $M(P)$ que retorna o valor N para um dado ponto P .

Pode-se então calcular a sensibilidade $S(P)$ esperada em P através das equações 4.6 e 4.7.

$$S_R(P) = 10 \log\left(1 + \frac{T(M(P))}{T_{NORMAL}(M(P), I)}\right) \quad (4.6)$$

$$S(P) = S_{NORMAL}(P, I) + S_R(P) \quad (4.7)$$

Uma vez obtidas as sensibilidades relativas $S(P)$ para cada um dos 52 pontos do campo visual baseados nas perdas ou acréscimos de espessura nos setores correspondentes do nervo óptico, os demais dados do SAP podem ser calculados:

- **Gráfico da sensibilidade da perimetria padrão.** O valor esperado para a espessura de um determinado setor da RNFL é extraído diretamente do banco de dados normativo do OCT. Calcula-se a relação entre o valor de espessura obtido aleatoriamente e o valor esperado. Com base neste valor e no banco de dados normativo do SAP, obtém-se a sensibilidade em dB para um determinado ponto p do campo visual através da equação 4.7.
- **Gráfico *Total Deviation*.** Para cada ponto do gráfico de sensibilidade subtrai-se o respectivo valor normal de acordo com o valor do parâmetro idade.
- **Gráfico *Pattern Deviation*.** Através de ordenação decrescente dos valores do gráfico *Total Deviation*, o sétimo maior valor pode ser encontrado. Subtrai-se então este número de cada um dos pontos do gráfico *Total Deviation* para obtenção do valor correspondente no gráfico *Pattern Deviation*.

- **Gráfico *Total Deviation Probability*.** Para cada ponto do gráfico *Total Deviation* o valor da probabilidade de normalidade é obtido com base na distribuição Normal do banco de dados normativo do SAP, cujas médias e desvios padrões por faixa etária foram previamente calculados.
- **Gráfico *Pattern Deviation Probability*.** Para cada ponto do gráfico *Pattern Deviation* o valor da probabilidade de normalidade é obtido com base na distribuição Normal do banco de dados normativo do SAP, cujas médias e desvios padrões por faixa etária foram previamente calculados.
- **Mean deviation (MD).** O MD é calculado através da média aritmética dos valores do gráfico *Total Deviation*.
- **Pattern Standard Deviation (PSD).** O PSD pode ser calculado através da média dos valores do gráfico *pattern deviation*.

Uma vez obtidos todos os dados necessários pode-se então proceder à fase final do processo, que promove a geração de uma saída gráfica dos resultados obtidos para facilitar a legibilidade e conseqüentemente a compreensão pelos médicos especialistas.

4.5 A Saída Gráfica do GLOR

Os oftalmologistas especialistas em glaucoma estão familiarizados com a visualização dos resultados dos exames na forma gráfica, de acordo com o padrão de saída apresentado pelos instrumentos empregados nos testes diagnósticos.

Para avaliação da capacidade do presente método em gerar dados representativos de populações reais é de vital importância o parecer médico. Sendo assim, avaliar apenas números discretos, sem a correspondente informação gráfica, dificultaria sensivelmente a compreensão dos resultados do GLOR.

Desta forma, desenvolveu-se uma saída gráfica que simula a forma de apresentação de resultados dos instrumentos Cirrus HD-OCT e HFA-II. Além de exibir os dados de idade e PIO, os dados e gráficos do HD-OCT e SAP são dispostos em uma mesma página, facilitando ainda mais a legibilidade e estabelecimento visual de intercorrelações entre as informações estruturais e funcionais. Um modelo desta saída gráfica do GLOR pode ser observada na íntegra através da figura 4.15.

Na figura 4.16 pode-se observar a mesma saída gráfica, porém destacando-se os atributos Idade (*AGE*) e PIO (*IOP*) em A, espessura total média (*Thickness Average*) da

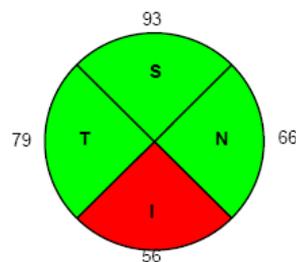
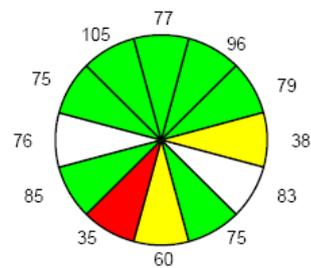
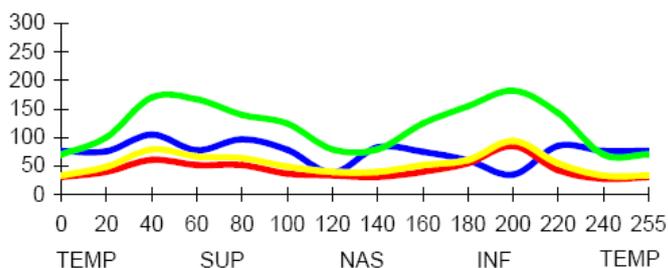
GLOR

glaucoma oracle

Set# 38

Age 80
IOP 25
Thick_Avg 74

G []
N []
R []



		26	26	26	25				
		27	25	24	24	28	27		
		27	29	27	26	26	25	29	28
		26	28	28	28	28	32	26	29
		26	28	30	30	30	33	30	28
		28	30	29	31	30	30	30	28
		29	27	30	28	30	29		
		28	28	29	29				

		0	0	0	-1				
		-1	-3	-5	-5	0	0		
		-1	0	-2	-5	-4	-5	0	0
		0	0	-2	-3	-3	1	-4	0
		0	0	-1	-1	-2	1	-1	-1
		0	0	-2	-1	-1	-1	0	-1
		0	-2	-1	-1	0	-1		
		0	0	0	0				

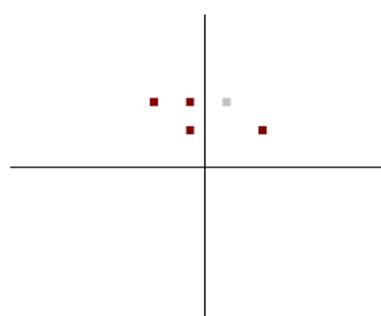
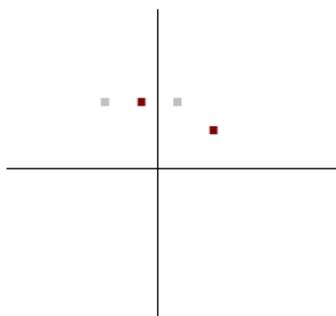
		0	0	0	-1				
		-1	-3	-5	-5	0	0		
		-1	0	-2	-5	-4	-5	0	0
		0	0	-2	-3	-3	1	-4	0
		0	0	-1	-1	-2	1	-1	-1
		0	0	-2	-1	-1	-1	0	-1
		0	-2	-1	-1	0	-1		
		0	0	0	0				

Total Deviation

Pattern Deviation

MD -1,13 dB

PSD 1,58 dB



- <5%
- <2%
- <1%
- <0.5%

Figura 4.15: Exemplo de saída gráfica do GLOR para uma determinada instância obtida artificialmente.

RNFL em B, o gráfico normativo de espessuras (*RNFL TSNIT Normative Data*) em C, as espessuras médias a cada hora de relógio (*Clock-hour Thickness*) em D, a espessura média por quadrante (*Quadrants*), gráfico de sensibilidade relativa do campo visual em F, gráfico *Total Deviation* em G, gráfico *Total Deviation Probability Plot* em H, gráfico *Pattern Deviation* em I e gráfico *Pattern Deviation Probability Plot* em J, além dos índices globais MD e PSD em K.

Esta saída gráfica do GLOR possui ainda no canto superior direito o espaço reservado para avaliação qualitativa do médico. Para atender a esta finalidade, uma lista contendo 3 possíveis notas foi concebida. Nesta lista, o oftalmologista especialista em glaucoma pode assinalar com um X a opção que reflete o resultado de sua classificação para um determinado conjunto de dados artificiais gerados para um paciente virtual. As opções disponíveis são: G para indicar que o resultado é qualitativamente representativo de um indivíduo real e condiz com o diagnóstico de glaucoma, N para indicar que o resultado é qualitativamente representativo de um indivíduo real normal, ou seja, sem glaucoma e R para indicar a rejeição do resultado pelo mesmo não apresentar satisfatoriamente indícios de que o resultado possui características de um indivíduo real.

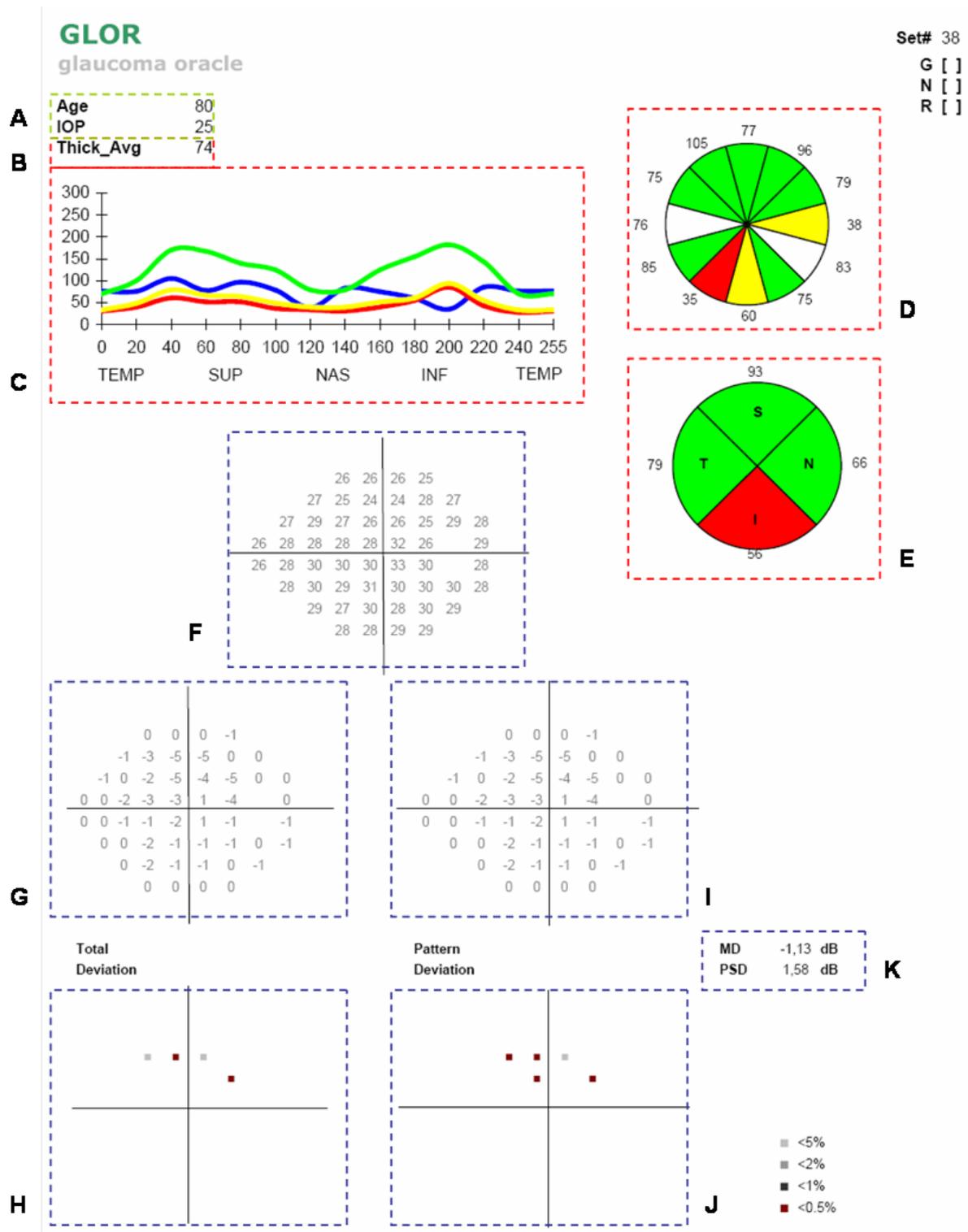


Figura 4.16: Saída gráfica do GLOR destacando os parâmetros epidemiológicos e clínicos (em verde), estruturais (em vermelho) e funcionais (em azul) gerados pelo GLOR.

5 Avaliação dos Resultados

Com o objetivo de avaliar o desempenho do GLOR ao gerar amostras populacionais artificiais representativas de populações reais para treinamento de classificadores de glaucoma, primeiramente foi conduzida uma avaliação qualitativa dos resultados com o auxílio de médicos especialistas. Em seguida um novo *dataset* foi gerado para análise qualitativa dos resultados obtidos frente os esperados. Nas duas etapas seguintes classificadores foram treinados e testados com dados artificiais. Finalmente, os classificadores treinados com dados gerados pelo GLOR, foram avaliados na tarefa de classificação de glaucoma através de dados de pacientes reais, previamente classificados por oftalmologistas.

5.1 Avaliação Qualitativa do GLOR

O critério empregado para avaliação da capacidade do método desenvolvido em gerar artificialmente dados que representem populações reais de indivíduos normais e glaucomatosos foi submeter os dados a médicos especialistas em glaucoma para que realizassem uma análise qualitativa dos resultados apresentados.

Apesar de ser subjetivo, pois depende da avaliação individual de cada médico, este critério de avaliação foi escolhido baseando-se na premissa de que, caso os oftalmologistas não conseguissem distinguir dados que foram gerados artificialmente de dados de indivíduos reais, então o método poderia ser considerado bem sucedido.

Para tanto, um *dataset* de baixa cardinalidade (40 instâncias) foi gerado. A opção por gerar uma pequena quantidade de exemplos foi motivada pelo tempo necessário para que os oftalmologistas efetuassem suas análises, o que seria bastante dificultado caso grandes conjuntos de dados tivessem sido empregados. Entre as 40 instâncias geradas, 20 foram obtidas através das características estatísticas de uma população normal e as outras 20 de acordo com as características de uma população de indivíduos com glaucoma.

As saídas gráficas foram então submetidas à avaliação de 2 médicos especialistas em

glaucoma para que avaliassem a pertinência dos dados gerados artificialmente como dados representativos de indivíduos reais. Das 40 instâncias avaliadas, o consenso observado ao se comparar as classificações dos especialistas frente a classe definida a priori, houveram 36 acertos e 4 erros (taxa de erro de 10%). Em outras palavras, dos 40 resultados avaliados isoladamente pelos especialistas, 36 das classificações apontadas como glaucoma ou normal pelos médicos coincidiram com a classe definida para cada instância a priori. É importante destacar que os médicos desconheciam tanto a classificação prévia quanto a proporção de glaucomatosos e normais na amostra.

No entanto, das 80 avaliações efetuadas (40 de cada médico), 13 foram rejeitadas por apenas um deles, o que representa uma taxa de ocorrência de erros de 16,25%. No entanto, nenhuma das instâncias avaliadas foi rejeitada por ambos os especialistas.

Como ambas as taxas de erro não foram significativas, os resultados foram considerados satisfatórios, fazendo com que o método de avaliação prosseguisse para a próxima etapa, que abordará o desempenho quantitativo do GLOR ao gerar grandes conjuntos de dados.

5.2 Avaliação Quantitativa do GLOR

Para avaliar a capacidade do método em gerar populações que obedecessem às características estatísticas definidas a priori calculou-se as respectivas médias e desvios padrões para cada um dos 14 atributos gerados artificialmente e comparou-as com os valores esperados.

Os resultados desta comparação encontram-se na tabela 5.1, onde os atributos $T_{1h} \dots T_{12h}$ indicam as espessuras setoriais da camada de fibras nervosas. Nesta tabela estão dispostos os valores das médias (μ) e desvios padrões (σ) esperados e os obtidos pelo GLOR para indivíduos normais e com glaucoma para cada atributo gerado. Para medir a diferença entre as médias dentro de cada grupo (normal e glaucoma) foi calculado um índice ($\Delta\mu$) que reflete o erro percentual entre a média obtida frente a esperada. Analogamente, $\Delta\sigma$ representa o erro percentual dos desvios padrões.

Avaliando-se os valores de $\Delta\mu$ pode-se identificar que o maior desvio apresentado foi 0,85% no grupo que representa indivíduos normais e 2,11% para o grupo com glaucoma. A média entre as variações, positivas e negativas, de $\Delta\mu$ foi de 0,13% para o primeiro grupo e 0,00% para o segundo. Já para os valores de $\Delta\sigma$, o maior erro relativo apresentado foi -2,02% para os normais e -6,07% para os glaucomatosos, sendo que as médias entre os

Tabela 5.1: Médias e desvios padrões por atributos gerados artificialmente e respectivos erros percentuais em relação aos valores esperados definidos a priori.

Atrib.	Normal				Glaucoma			
	Real(A)	GLOR(B)	(B-A)/A(%)		Real(C)	GLOR(D)	(D-C)/C(%)	
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\Delta\mu$	$\Delta\sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\Delta\mu$	$\Delta\sigma$
Idade	65,0±8,0	65,1 ± 8,0	0,08	0,11	68,0±10,0	68,5±10,2	0,66	1,60
PIO	16,9±2,5	16,9 ± 2,5	0,21	-0,95	27,7± 9,1	28,3± 9,4	2,11	3,32
T _{1h}	109,4±21,5	109,6±21,1	0,18	-2,02	87,1±28,2	88,8±29,1	1,97	3,15
T _{2h}	92,9±25,1	92,7 ±25,0	-0,26	-0,39	70,5±23,6	71,1±24,3	0,89	3,01
T _{3h}	64,3±19,3	64,4 ±19,4	0,13	0,38	48,7±12,2	48,4±12,6	-0,60	2,93
T _{4h}	76,6±19,9	76,9 ±20,0	0,38	0,48	58,9±17,4	58,5±17,8	-0,66	2,28
T _{5h}	106,1±25,6	107,0±26,1	0,85	1,87	77,3±24,5	76,5±23,9	-1,08	-2,55
T _{6h}	134,7±23,7	134,7±23,7	-0,01	0,03	89,2±29,3	87,5±28,5	-1,87	-2,90
T _{7h}	132,2±21,4	132,3±21,3	0,04	-0,53	88,2±32,9	87,9±32,8	-0,39	-0,18
T _{8h}	69,3±18,5	69,3 ±18,8	-0,02	1,81	62,4±18,5	62,3±18,1	-0,12	-1,90
T _{9h}	52,8±11,5	52,8 ±11,5	0,07	-0,23	52,9±12,7	52,6±11,9	-0,52	-6,07
T _{10h}	81,8±19,0	81,9 ±19,0	0,07	0,02	70,6±21,0	70,5±21,1	-0,18	0,52
T _{11h}	127,7±17,9	127,7±17,9	0,03	0,19	97,1±31,1	95,4±30,4	-1,78	-2,26
T _{12h}	122,0±26,5	122,1±26,5	0,07	-0,06	95,4±29,9	96,9±30,5	1,62	1,96
μ			0,13	0,05			0,00	-0,21

desvios em cada grupo foram de 0,05% e -0,21%, respectivamente.

Ao avaliar-se graficamente a frequência de ocorrência dos valores apresentados pelos parâmetros, observou-se a distribuição Normal com as médias e desvios padrões esperados. Nas figuras 5.1 e 5.2 tem-se, respectivamente, os gráficos de distribuições de frequência de ocorrência para o conjunto de 4500 instâncias geradas de acordo com as características de uma população normal, e a mesma distribuição para a amostra de 500 instâncias geradas de acordo com as características de uma população com glaucoma. Em ambas as curvas o formato de sino que caracteriza distribuições gaussianas pode ser percebido, sendo mais evidente na população normal. Isto se deve basicamente a dois fatores: a amostra de pacientes normais é sensivelmente maior do que a de indivíduos com glaucoma, e o valor da medida da variância (9.1) do atributo PIO utilizado para gerar o grupo de glaucomatosos é elevado em relação à média (27.7), causando um efeito de espalhamento mais acentuado na curva.

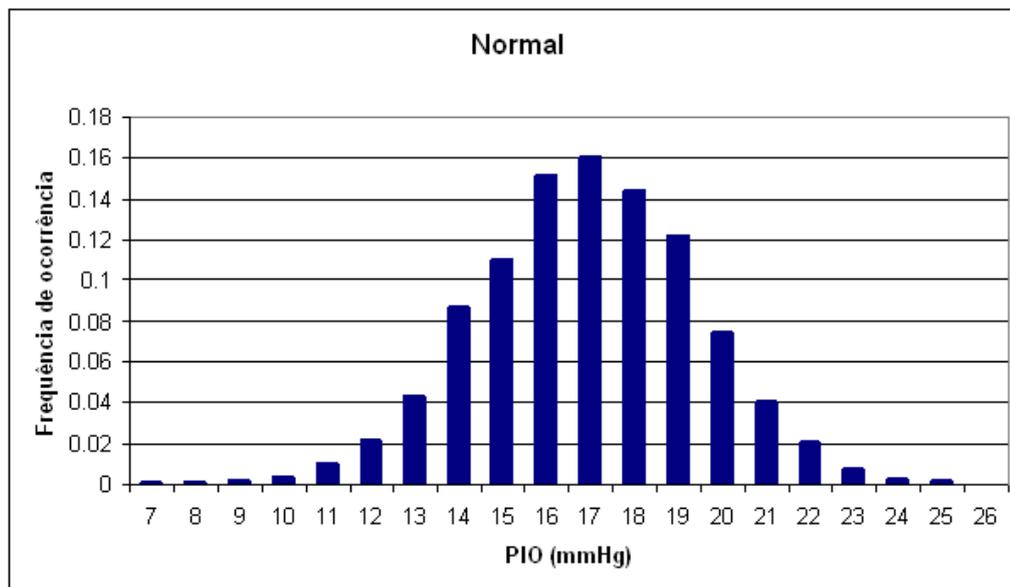


Figura 5.1: Distribuição de frequências de ocorrência para valores de pressão intra-ocular gerados pelo GLOR de acordo com características de uma população normal (n=4500).

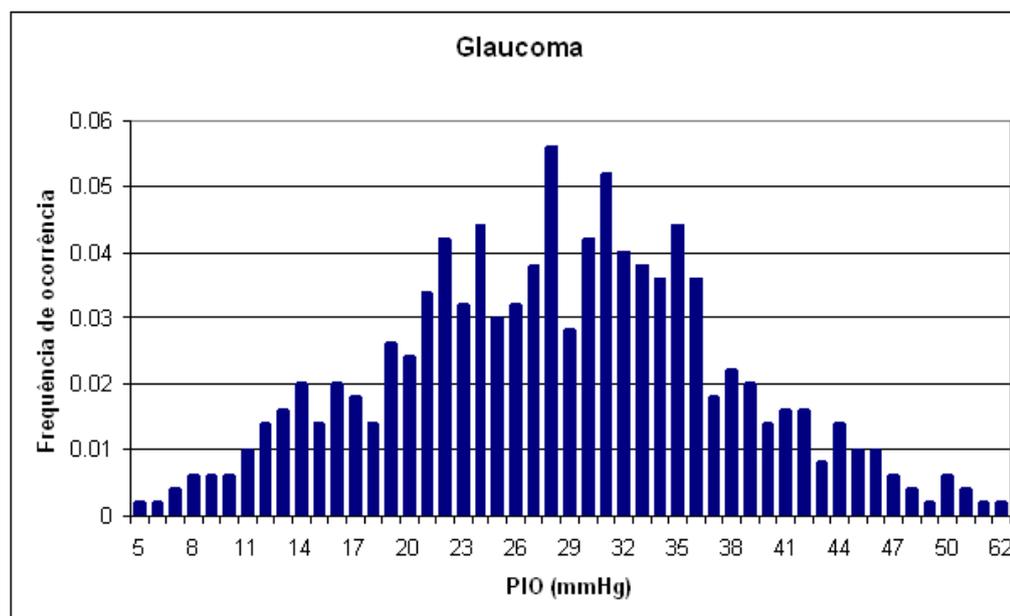


Figura 5.2: Distribuição de frequências de ocorrência para valores de pressão intra-ocular gerados pelo GLOR de acordo com características de uma população com glaucoma (n=500).

5.3 Avaliação dos Classificadores com Dados Artificiais

Gerou-se uma amostra populacional contendo 4500 instâncias representativas de indivíduos normais e 500 instâncias representativas de indivíduos com glaucoma. Além dos parâmetros numéricos da tabela 5.2, cada conjunto possui adicionalmente um atributo CLASSE. Este atributo foi definido a priori com o valor nominal N para cada um dos 4500 elementos do subconjunto gerado de acordo com as características de uma população de indivíduos normais. Analogamente, em cada uma das 500 instâncias pertencentes ao subconjunto gerado de acordo com as características de uma população de indivíduos com glaucoma, o atributo CLASSE foi definido a priori com o valor nominal G. Em seguida, ambas as amostras foram unidas, de modo a constituir um único conjunto de dados (*dataset*) com 178 atributos e 5000 instâncias.

Tabela 5.2: Lista de atributos numéricos gerados pelo GLOR e utilizados no treinamento e teste dos classificadores, agrupados por categoria.

Categoria	Atributo
Dados epidemiológicos	Idade
Dados clínicos	PIO
Dados de OCT	Espessura Média Total Espessura Quadrante Temporal Espessura Quadrante Superior Espessura Quadrante Nasal Espessura Quadrante Inferior Espessura a 1 hora de relógio Espessura às 2 horas de relógio Espessura às 3 horas de relógio Espessura às 4 horas de relógio Espessura às 5 horas de relógio Espessura às 6 horas de relógio Espessura às 7 horas de relógio Espessura às 8 horas de relógio Espessura às 9 horas de relógio Espessura às 10 horas de relógio Espessura às 11 horas de relógio Espessura às 12 horas de relógio
Dados de SAP	52 pontos de sensibilidade relativa do mapa de campo visual 52 pontos do gráfico Total Deviation 52 pontos do gráfico Total Deviation MD PSD
Total	177 atributos

Uma vez finalizada a geração do *dataset*, o ambiente de aprendizagem WEKA (WITTEN; FRANK, 2005) foi empregado no treinamento e testes dos classificadores. Para tanto, um arquivo contendo as instâncias de treinamento foi exportado para um arquivo no formato CSV (*Comma Separated Values*) ou valores separados por vírgula, e convertidos para o formato ARFF (*Attribute-Relation File Format*), padrão nativo do WEKA .

Como algoritmos de classificação foram selecionados:

1. Naive-Bayes
2. MLP
3. Árvore de decisão (CTREE)
4. Árvore de decisão com técnica de *Bagging* (Bagged-CTREE)

A estratégia de testes de cada algoritmo foi a validação cruzada através de 10 grupos (*10-fold cross-validation*). Nesta estratégia, o conjunto original contendo 5000 instâncias é segmentado em 10 subconjuntos com aproximadamente a mesma quantidade de instâncias, ou seja, 500, de forma que em cada um destes subconjuntos a proporção de exemplos positivos (CLASSE=G) e negativos (CLASSE=N) seja aproximadamente a mesma encontrada na amostra total. Assim, as 5000 instâncias são aleatoriamente distribuídas em 10 subconjuntos de 500 elementos, onde 50 pertencem á classe G e 450 à classe N, reproduzindo então a taxa de 10% de instâncias positivas em cada grupo. Em seguida, dos 10 subconjuntos gerados, 9 são empregados como dados de treinamento e 1 para testes e mensuração de erros. Este processo de treinamento e teste é então repetido outras 9 vezes, de tal forma que cada um dos 10 subconjuntos tenha sido utilizado como conjunto de testes. A taxa de erro final corresponderá à média das taxas de erro obtidas pelo classificador em cada uma das 10 etapas de treinamento e testes. Entre as diversas estratégias para estimativa de erros de classificadores, a validação cruzada é considerada padrão para validação de algoritmos de aprendizagem sobre um *dataset* em particular (WITTEN; FRANK, 2005).

Para cada classificador foram obtidos e registrados as seguintes medidas de desempenho:

1. Acurácia total;
2. Sensibilidade;
3. Especificidade;
4. Área sob a curva ROC.

Após o treinamento e testes dos 4 classificadores, com o registro dos respectivos desempenhos, o processo foi repetido outras 5 vezes, mantendo-se a cardinalidade das amostras (5000 instâncias), porém variando-se agora os subconjuntos de atributos utilizados:

1. Idade + PIO + dados de OCT (*dataset* IOP + AGE + OCT);
2. Idade + PIO + dados de SAP (*dataset* IOP + AGE + SAP);
3. Idade + PIO + dados de OCT + dados de SAP (*dataset* IOP + AGE + OCT + SAP);
4. Idade + dados de OCT (*dataset* AGE + OCT);
5. Idade + dados de SAP (*dataset* AGE + SAP);
6. Idade + dados de OCT + dados de SAP (*dataset* AGE + OCT + SAP).

O objetivo de empregar-se estes subconjuntos no treinamento e testes dos classificadores foi avaliar quais tipos de dados podem proporcionar o melhor desempenho na tarefa de classificação de glaucoma.

Para avaliação da acurácia dos classificadores quando treinados e testados com dados artificialmente gerados pelo GLOR, inicialmente foi gerado um *dataset* com 5000 instâncias e 178 atributos, conforme descrito no capítulo anterior. Em seguida 4 classificadores foram treinados e testados com o ambiente de aprendizagem WEKA. Como estratégia de testes utilizou-se a validação cruzada (*10-fold cross-validation*). Os resultados apresentados encontram-se dispostos na tabela 5.3, organizados por classificador e *dataset*.

Na tabela 5.4 tem-se os resultados de acordo com o critério de acurácia total, apresentando como melhor desempenho o classificador *Bagged-CTREE* sobre os dados dos *datasets* IOP+AGE+OCT e IOP+AGE+OCT+SAP (ambos com acurácia de 99,0% e com áreas sob a curva ROC praticamente idênticas (0,998 e 0,997). O classificador com pior desempenho neste quesito acurácia total foi o Naive-Bayes, mesmo assim com elevados índices de 95,7% de acurácia total e 0,982 de área sob a curva ROC.

Já na tabela 5.5 tem-se os resultados dispostos de acordo com o critério de área sob a curva ROC. Através destes dados é possível identificar que todos os classificadores apresentaram resultados aparentemente excelentes, sendo que o de melhor desempenho foi o CTREE sobre os dados idade, pressão intra-ocular e SAP, com área sob a curva de 0.999, enquanto que o pior foi o CTREE sobre atributos idade e SAP, com área de 0.927. Estes valores elevados que todos os classificadores apresentaram pode ser um sinal de

Tabela 5.3: Resultados dos testes dos classificadores com dados artificiais organizados por classificador e *dataset*.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	BAG	AGE+OCT	98,2%	88,2%	99,4%	0,945
2	BAG	AGE+OCT+SAP	98,2%	87,8%	99,3%	0,994
3	BAG	AGE+SAP	97,8%	85,6%	99,2%	0,938
4	BAG	IOP+AGE+OCT	99,0%	92,6%	99,7%	0,998
5	BAG	IOP+AGE+OCT+SAP	99,0%	92,6%	99,7%	0,997
6	BAG	IOP+AGE+SAP	98,4%	87,8%	99,6%	0,954
7	CTR	AGE+OCT	98,2%	89,6%	99,1%	0,993
8	CTR	AGE+OCT+SAP	98,2%	89,6%	99,1%	0,963
9	CTR	AGE+SAP	96,9%	83,4%	98,4%	0,927
10	CTR	IOP+AGE+OCT	98,6%	93,0%	99,2%	0,993
11	CTR	IOP+AGE+OCT+SAP	98,7%	92,6%	99,4%	0,996
12	CTR	IOP+AGE+SAP	98,4%	89,8%	99,4%	0,999
13	MLP	AGE+OCT	97,9%	87,0%	99,1%	0,994
14	MLP	AGE+SAP	98,1%	88,0%	99,2%	0,997
15	MLP	IOP+AGE+OCT	98,6%	92,2%	99,4%	0,962
16	MLP	IOP+AGE+OCT+SAP	98,7%	92,0%	99,5%	0,996
17	MLP	IOP+AGE+SAP	98,7%	91,8%	99,4%	0,997
18	MLP	AGE+OCT+SAP	98,1%	90,6%	99,0%	0,995
19	NB	AGE+OCT	98,2%	88,2%	99,4%	0,990
20	NB	AGE+OCT+SAP	96,2%	93,0%	96,6%	0,986
21	NB	AGE+SAP	95,7%	92,4%	96,1%	0,982
22	NB	IOP+AGE+OCT	98,6%	97,2%	98,8%	0,984
23	NB	IOP+AGE+OCT+SAP	96,5%	94,4%	96,8%	0,989
24	NB	IOP+AGE+SAP	96,1%	94,4%	96,3%	0,986

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

sobre-ajuste (*overfitting*). O sobre-ajuste é caracterizado por um desempenho a princípio ótimo com os dados de treinamento, mas que reduz-se muitas vezes significativamente quando dados inéditos (ou seja, não utilizados de nenhuma forma durante a etapa de treinamento) são apresentados ao classificador.

Uma possível fonte de sobre-ajuste é a presença de ruído randômico no conjunto de treinamento. Este ruído pode fazer com que o classificador encontre uma hipótese no espaço de hipóteses que consiga representar o conceito-alvo adequadamente frente aos dados de treinamento, porém falhe ao tentar classificar exemplos que são cobertos por uma outra hipótese melhor que a anterior, porém não identificada pelo algoritmo. Isto acontece porque o algoritmo refina sua busca pelo espaço de hipóteses para que os exemplos ruidosos possam ser classificados, gerando com isso hipóteses mais complexas, mas que poderão falhar na tarefa de classificação de novos dados sem ruído.

Outra fonte de sobre-ajuste é a quantidade baixa de exemplos de treinamento, ainda

Tabela 5.4: Resultados dos testes dos classificadores com dados artificiais de acordo com o critério de acurácia total.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	BAG	IOP+AGE+OCT	99,0%	92,6%	99,7%	0,998
2	BAG	IOP+AGE+OCT+SAP	99,0%	92,6%	99,7%	0,997
3	MLP	IOP+AGE+OCT+SAP	98,7%	92,0%	99,5%	0,996
4	CTR	IOP+AGE+OCT+SAP	98,7%	92,6%	99,4%	0,996
5	MLP	IOP+AGE+SAP	98,7%	91,8%	99,4%	0,997
6	MLP	IOP+AGE+OCT	98,6%	92,2%	99,4%	0,962
7	NB	IOP+AGE+OCT	98,6%	97,2%	98,8%	0,984
8	CTR	IOP+AGE+OCT	98,6%	93,0%	99,2%	0,993
9	CTR	IOP+AGE+SAP	98,4%	89,8%	99,4%	0,999
10	BAG	IOP+AGE+SAP	98,4%	87,8%	99,6%	0,954
11	NB	AGE+OCT	98,2%	88,2%	99,4%	0,990
12	BAG	AGE+OCT	98,2%	88,2%	99,4%	0,945
13	BAG	AGE+OCT+SAP	98,2%	87,8%	99,3%	0,994
14	CTR	AGE+OCT+SAP	98,2%	89,6%	99,1%	0,963
15	CTR	AGE+OCT	98,2%	89,6%	99,1%	0,993
16	MLP	AGE+OCT+SAP	98,1%	90,6%	99,0%	0,995
17	MLP	AGE+SAP	98,1%	88,0%	99,2%	0,997
18	MLP	AGE+OCT	97,9%	87,0%	99,1%	0,994
19	BAG	AGE+SAP	97,8%	85,6%	99,2%	0,938
20	CTR	AGE+SAP	96,9%	83,4%	98,4%	0,927
21	NB	IOP+AGE+OCT+SAP	96,5%	94,4%	96,8%	0,989
22	NB	AGE+OCT+SAP	96,2%	93,0%	96,6%	0,986
23	NB	IOP+AGE+SAP	96,1%	94,4%	96,3%	0,986
24	NB	AGE+SAP	95,7%	92,4%	96,1%	0,982

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

que eventualmente sejam 100% livres de ruído. No caso específico do GLOR, é mais provável que eventual sobre-ajuste ocorra por presença de ruído nos dados, graças às características randômicas do processo de geração artificial de dados, do que relativo à baixa cardinalidade dos conjuntos de treinamento, uma vez que grandes quantidades de dados foram gerados e empregados como exemplos de treinamento.

Tabela 5.5: Resultados dos testes dos classificadores com dados artificiais de acordo com o critério de área sob a curva ROC.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	CTR	IOP+AGE+SAP	98,4%	89,8%	99,4%	0,999
2	BAG	IOP+AGE+OCT	99,0%	92,6%	99,7%	0,998
3	MLP	IOP+AGE+SAP	98,7%	91,8%	99,4%	0,997
4	MLP	AGE+SAP	98,1%	88,0%	99,2%	0,997
5	BAG	IOP+AGE+OCT+SAP	99,0%	92,6%	99,7%	0,997
6	MLP	IOP+AGE+OCT+SAP	98,7%	92,0%	99,5%	0,996
7	CTR	IOP+AGE+OCT+SAP	98,7%	92,6%	99,4%	0,996
8	MLP	AGE+OCT+SAP	98,1%	90,6%	99,0%	0,995
9	BAG	AGE+OCT+SAP	98,2%	87,8%	99,3%	0,994
10	MLP	AGE+OCT	97,9%	87,0%	99,1%	0,994
11	CTR	AGE+OCT	98,2%	89,6%	99,1%	0,993
12	CTR	IOP+AGE+OCT	98,6%	93,0%	99,2%	0,993
13	NB	AGE+OCT	98,2%	88,2%	99,4%	0,990
14	NB	IOP+AGE+OCT+SAP	96,5%	94,4%	96,8%	0,989
15	NB	AGE+OCT+SAP	96,2%	93,0%	96,6%	0,986
16	NB	IOP+AGE+SAP	96,1%	94,4%	96,3%	0,986
17	NB	IOP+AGE+OCT	98,6%	97,2%	98,8%	0,984
18	NB	AGE+SAP	95,7%	92,4%	96,1%	0,982
19	CTR	AGE+OCT+SAP	98,2%	89,6%	99,1%	0,963
20	MLP	IOP+AGE+OCT	98,6%	92,2%	99,4%	0,962
21	BAG	IOP+AGE+SAP	98,4%	87,8%	99,6%	0,954
22	BAG	AGE+OCT	98,2%	88,2%	99,4%	0,945
23	BAG	AGE+SAP	97,8%	85,6%	99,2%	0,938
24	CTR	AGE+SAP	96,9%	83,4%	98,4%	0,927

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

5.4 Avaliação dos Classificadores com Dados Reais

Para avaliar o desempenho dos classificadores treinados com dados artificiais gerados pelo GLOR, efetuou-se uma série de 24 testes envolvendo os 4 classificadores treinados sobre 6 *datasets* artificiais, porém agora utilizando-se 6 *datasets* com os mesmos atributos, porém contendo dados reais inéditos, ou seja, jamais apresentados aos classificadores.

Cada *dataset* real continha 70 instâncias, sendo 44 delas pertencentes à classe N (normal) e 26 pertencentes à classe G (glaucoma). Tais dados foram coletados ao longo de 5 meses pelo departamento de oftalmologia do Hospital das Clínicas da Universidade de Campinas (UNICAMP). Todos os pacientes assinaram um Termo de Consentimento Livre e Esclarecido sobre o uso dos seus dados no estudo para fins da pesquisa científica previamente aprovada pelo Comitê de Ética em Pesquisa (CEP) da Faculdade de Ciências Médicas (FCM) da UNICAMP (COSTA, 2007). Os respectivos critérios de inclusão e exclusão de pacientes neste estudo encontram-se no Apêndice A deste trabalho. Médicos

especialistas em glaucoma fizeram a classificação de cada paciente quanto à presença ou não de glaucoma de acordo com anamnese e avaliação clínica, a medida da pressão intra-ocular, resultados de exames de perimetria acromática padrão (SAP), obtidos com o instrumento Humphrey HFA-II, e resultados de tomografia de coerência óptica de alta definição obtidos com o instrumento Cirrus HD-OCT.

Pode-se observar na tabela 5.6 os resultados ordenados por classificador e *dataset*. Nela são exibidos os dados da acurácia total, que corresponde à porcentagem total de instâncias corretamente classificadas, os índices de sensibilidade e especificidade, além da área sob a curva ROC de cada classificador.

Tabela 5.6: Resultados dos testes dos classificadores com dados reais por classificador e *dataset*.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	BAG	AGE+OCT	85,7%	76,9%	90,9%	0,905
2	BAG	AGE+OCT+SAP	85,7%	76,9%	90,9%	0,900
3	BAG	AGE+SAP	55,7%	84,6%	38,6%	0,761
4	BAG	IOP+AGE+OCT	84,3%	73,1%	90,9%	0,865
5	BAG	IOP+AGE+OCT+SAP	82,9%	73,1%	88,6%	0,842
6	BAG	IOP+AGE+SAP	54,3%	84,6%	36,4%	0,627
7	CTR	AGE+OCT	88,6%	80,8%	93,2%	0,910
8	CTR	AGE+OCT+SAP	70,0%	88,5%	59,1%	0,619
9	CTR	AGE+SAP	64,3%	84,6%	52,3%	0,724
10	CTR	IOP+AGE+OCT	88,6%	76,9%	95,5%	0,818
11	CTR	IOP+AGE+OCT+SAP	72,9%	88,5%	63,6%	0,628
12	CTR	IOP+AGE+SAP	58,6%	76,9%	47,7%	0,568
13	MLP	AGE+OCT	87,1%	80,8%	90,9%	0,941
14	MLP	AGE+OCT+SAP	77,1%	80,8%	75,0%	0,855
15	MLP	AGE+SAP	54,3%	76,9%	40,9%	0,680
16	MLP	IOP+AGE+OCT	80,0%	69,2%	86,4%	0,804
17	MLP	IOP+AGE+OCT+SAP	75,7%	73,1%	77,3%	0,844
18	MLP	IOP+AGE+SAP	55,7%	76,9%	43,2%	0,658
19	NB	AGE+OCT	85,7%	80,8%	88,6%	0,932
20	NB	AGE+OCT+SAP	48,6%	96,2%	20,5%	0,621
21	NB	AGE+SAP	45,7%	92,3%	18,2%	0,572
22	NB	IOP+AGE+OCT	85,7%	80,8%	88,6%	0,932
23	NB	IOP+AGE+OCT+SAP	48,6%	96,2%	20,5%	0,617
24	NB	IOP+AGE+SAP	45,7%	92,3%	18,2%	0,572

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

Na tabela 5.7 tem-se os resultados de acordo com o critério de acurácia total, apresentando como melhor desempenho o classificador CTREE sobre os dados dos *datasets* AGE + OCT e IOP + AGE + OCT (ambos com acurácia de 88,57%. Porém o desempenho com o *dataset* sem o atributo da pressão intra-ocular apresentou área sob a curva ROC de 0,910 contra 0,818 do mesmo conjunto de atributos com a PIO. Apesar de ter apresentado

acurácia ligeiramente inferior ao CTREE (87,57%), o classificador MLP obteve a maior área sob a curva ROC (0,941) quando utilizado para classificar o *dataset* contendo apenas atributos idade e dados de OCT, apresentando excelente desempenho nesta tarefa, com sensibilidade e especificidade de 80,8% e 90,9%, respectivamente. O pior resultado de acordo com o critério de acurácia total, com apenas 45,71% de acertos, foi apresentado pelo classificador bayesiano sobre os dados de idade, pressão intra-ocular e SAP.

Tabela 5.7: Resultados dos testes dos classificadores com dados reais de acordo com o critério de acurácia total.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	CTR	AGE+OCT	88,6%	80,8%	93,2%	0,910
2	CTR	IOP+AGE+OCT	88,6%	76,9%	95,5%	0,818
3	MLP	AGE+OCT	87,1%	80,8%	90,9%	0,941
4	NB	AGE+OCT	85,7%	80,8%	88,6%	0,932
5	NB	IOP+AGE+OCT	85,7%	80,8%	88,6%	0,932
6	BAG	AGE+OCT	85,7%	76,9%	90,9%	0,905
7	BAG	AGE+OCT+SAP	85,7%	76,9%	90,9%	0,900
8	BAG	IOP+AGE+OCT	84,3%	73,1%	90,9%	0,865
9	BAG	IOP+AGE+OCT+SAP	82,9%	73,1%	88,6%	0,842
10	MLP	IOP+AGE+OCT	80,0%	69,2%	86,4%	0,804
11	MLP	AGE+OCT+SAP	77,1%	80,8%	75,0%	0,855
12	MLP	IOP+AGE+OCT+SAP	75,7%	73,1%	77,3%	0,844
13	CTR	IOP+AGE+OCT+SAP	72,9%	88,5%	63,6%	0,628
14	CTR	AGE+OCT+SAP	70,0%	88,5%	59,1%	0,619
15	CTR	AGE+SAP	64,3%	84,6%	52,3%	0,724
16	CTR	IOP+AGE+SAP	58,6%	76,9%	47,7%	0,568
17	BAG	AGE+SAP	55,7%	84,6%	38,6%	0,761
18	MLP	IOP+AGE+SAP	55,7%	76,9%	43,2%	0,658
19	MLP	AGE+SAP	54,3%	76,9%	40,9%	0,680
20	BAG	IOP+AGE+SAP	54,3%	84,6%	36,4%	0,627
21	NB	AGE+OCT+SAP	48,6%	96,2%	20,5%	0,621
22	NB	IOP+AGE+OCT+SAP	48,6%	96,2%	20,5%	0,617
23	NB	AGE+SAP	45,7%	92,3%	18,2%	0,572
24	NB	IOP+AGE+SAP	45,7%	92,3%	18,2%	0,572

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

Na tabela 5.8 tem-se os resultados dispostos de acordo com o critério de valor área sob a curva ROC, que é o principal critério para avaliação de sensibilidade e especificidade de forma ponderada, uma vez que, como já foi abordado neste trabalho, aumentos de sensibilidade são seguidos de reduções de especificidade e vice-versa.

Observando os dados desta tabela é possível identificar que o índice de sensibilidade (80,8%) e especificidade (90,9%) apresentados pelo classificador MLP sobre o *dataset* contendo apenas os atributos idade e as espessuras das camadas de fibras nervosas, são conjuntamente boas, revelando um equilíbrio entre os índices, o que é evidenciado pelo

valor da área sob a curva ROC obtida. Deve-se notar que na última colocação nesta lista encontra-se o classificador Bayesiano com o *dataset* contendo os atributos idade, pressão intra-ocular e dados de SAP. A área sob a curva ROC de apenas 0,568 evidencia o péssimo desempenho na classificação dos dados reais, lembrando que valores entre 0,50 e 0,60 indicam que o teste falhou em sua tarefa.

Tabela 5.8: Resultados dos testes dos classificadores com dados reais de acordo com o critério de área sob a curva ROC.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	MLP	AGE+OCT	87,1%	80,8%	90,9%	0,941
2	NB	AGE+OCT	85,7%	80,8%	88,6%	0,932
3	NB	IOP+AGE+OCT	85,7%	80,8%	88,6%	0,932
4	CTR	AGE+OCT	88,6%	80,8%	93,2%	0,910
5	BAG	AGE+OCT	85,7%	76,9%	90,9%	0,905
6	BAG	AGE+OCT+SAP	85,7%	76,9%	90,9%	0,900
7	BAG	IOP+AGE+OCT	84,3%	73,1%	90,9%	0,865
8	MLP	AGE+OCT+SAP	77,1%	80,8%	75,0%	0,855
9	MLP	IOP+AGE+OCT+SAP	75,7%	73,1%	77,3%	0,844
10	BAG	IOP+AGE+OCT+SAP	82,9%	73,1%	88,6%	0,842
11	CTR	IOP+AGE+OCT	88,6%	76,9%	95,5%	0,818
12	MLP	IOP+AGE+OCT	80,0%	69,2%	86,4%	0,804
13	BAG	AGE+SAP	55,7%	84,6%	38,6%	0,761
14	CTR	AGE+SAP	64,3%	84,6%	52,3%	0,724
15	MLP	AGE+SAP	54,3%	76,9%	40,9%	0,680
16	MLP	IOP+AGE+SAP	55,7%	76,9%	43,2%	0,658
17	CTR	IOP+AGE+OCT+SAP	72,9%	88,5%	63,6%	0,628
18	BAG	IOP+AGE+SAP	54,3%	84,6%	36,4%	0,627
19	NB	AGE+OCT+SAP	48,6%	96,2%	20,5%	0,621
20	CTR	AGE+OCT+SAP	70,0%	88,5%	59,1%	0,619
21	NB	IOP+AGE+OCT+SAP	48,6%	96,2%	20,5%	0,617
22	NB	AGE+SAP	45,7%	92,3%	18,2%	0,572
23	NB	IOP+AGE+SAP	45,7%	92,3%	18,2%	0,572
24	CTR	IOP+AGE+SAP	58,6%	76,9%	47,7%	0,568

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

As tabelas 5.9 e 5.10 apresentam os resultados de acordo com os índices de sensibilidade e especificidade, respectivamente. Estas tabelas demonstram que excelentes desempenhos no critério de sensibilidade são acompanhados por péssimos resultados de especificidade, e vice-versa. Deve-se notar as baixas áreas sob a curva ROC, como era esperado na avaliação que emprega os critérios de sensibilidade e especificidade de forma isolada.

Na tabela 5.11 tem-se a os resultados dos testes segmentados por tipo de *dataset* empregado. O critério utilizado para desempenho foi o valor da área sob a curva ROC. A observação das áreas médias para cada *dataset* revela que os melhores desempenhos

Tabela 5.9: Resultados dos testes dos classificadores com dados reais de acordo com o critério de sensibilidade.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	NB	AGE+OCT+SAP	48,6%	96,2%	20,5%	0,621
2	NB	IOP+AGE+OCT+SAP	48,6%	96,2%	20,5%	0,617
3	NB	AGE+SAP	45,7%	92,3%	18,2%	0,572
4	NB	IOP+AGE+SAP	45,7%	92,3%	18,2%	0,572
5	CTR	IOP+AGE+OCT+SAP	72,9%	88,5%	63,6%	0,628
6	CTR	AGE+OCT+SAP	70,0%	88,5%	59,1%	0,619
7	BAG	AGE+SAP	55,7%	84,6%	38,6%	0,761
8	CTR	AGE+SAP	64,3%	84,6%	52,3%	0,724
9	BAG	IOP+AGE+SAP	54,3%	84,6%	36,4%	0,627
10	MLP	AGE+OCT	87,1%	80,8%	90,9%	0,941
11	NB	AGE+OCT	85,7%	80,8%	88,6%	0,932
12	NB	IOP+AGE+OCT	85,7%	80,8%	88,6%	0,932
13	CTR	AGE+OCT	88,6%	80,8%	93,2%	0,910
14	MLP	AGE+OCT+SAP	77,1%	80,8%	75,0%	0,855
15	BAG	AGE+OCT	85,7%	76,9%	90,9%	0,905
16	BAG	AGE+OCT+SAP	85,7%	76,9%	90,9%	0,900
17	CTR	IOP+AGE+OCT	88,6%	76,9%	95,5%	0,818
18	MLP	AGE+SAP	54,3%	76,9%	40,9%	0,680
19	MLP	IOP+AGE+SAP	55,7%	76,9%	43,2%	0,658
20	CTR	IOP+AGE+SAP	58,6%	76,9%	47,7%	0,568
21	BAG	IOP+AGE+OCT	84,3%	73,1%	90,9%	0,865
22	MLP	IOP+AGE+OCT+SAP	75,7%	73,1%	77,3%	0,844
23	BAG	IOP+AGE+OCT+SAP	82,9%	73,1%	88,6%	0,842
24	MLP	IOP+AGE+OCT	80,0%	69,2%	86,4%	0,804

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

foram apresentados pelos conjuntos de dados que não continham dados funcionais (SAP), enquanto que os piores resultados foram aqueles dos *datasets* que não continham dados estruturais (OCT). Os *datasets* que continham dados estruturais e funcionais apresentaram bom desempenho, sendo que no melhor caso, o classificador Bagged-CTREE obteve área sob a curva ROC de 0,900, ocupando a sexta colocação na tabela 5.8 de acordo com este critério.

Tabela 5.10: Resultados dos testes dos classificadores com dados reais de acordo com o critério de especificidade.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	CTR	IOP+AGE+OCT	88,6%	76,9%	95,5%	0,818
2	CTR	AGE+OCT	88,6%	80,8%	93,2%	0,910
3	MLP	AGE+OCT	87,1%	80,8%	90,9%	0,941
4	BAG	AGE+OCT	85,7%	76,9%	90,9%	0,905
5	BAG	AGE+OCT+SAP	85,7%	76,9%	90,9%	0,900
6	BAG	IOP+AGE+OCT	84,3%	73,1%	90,9%	0,865
7	NB	AGE+OCT	85,7%	80,8%	88,6%	0,932
8	NB	IOP+AGE+OCT	85,7%	80,8%	88,6%	0,932
9	BAG	IOP+AGE+OCT+SAP	82,9%	73,1%	88,6%	0,842
10	MLP	IOP+AGE+OCT	80,0%	69,2%	86,4%	0,804
11	MLP	IOP+AGE+OCT+SAP	75,7%	73,1%	77,3%	0,844
12	MLP	AGE+OCT+SAP	77,1%	80,8%	75,0%	0,855
13	CTR	IOP+AGE+OCT+SAP	72,9%	88,5%	63,6%	0,628
14	CTR	AGE+OCT+SAP	70,0%	88,5%	59,1%	0,619
15	CTR	AGE+SAP	64,3%	84,6%	52,3%	0,724
16	CTR	IOP+AGE+SAP	58,6%	76,9%	47,7%	0,568
17	MLP	IOP+AGE+SAP	55,7%	76,9%	43,2%	0,658
18	MLP	AGE+SAP	54,3%	76,9%	40,9%	0,680
19	BAG	AGE+SAP	55,7%	84,6%	38,6%	0,761
20	BAG	IOP+AGE+SAP	54,3%	84,6%	36,4%	0,627
21	NB	AGE+OCT+SAP	48,6%	96,2%	20,5%	0,621
22	NB	IOP+AGE+OCT+SAP	48,6%	96,2%	20,5%	0,617
23	NB	AGE+SAP	45,7%	92,3%	18,2%	0,572
24	NB	IOP+AGE+SAP	45,7%	92,3%	18,2%	0,572

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

Tabela 5.11: Resultados dos testes dos classificadores com dados reais segmentados por *dataset*.

	CLS	Dataset	Acurácia	Sensib.	Espec.	AROC
1	MLP	AGE+OCT	87,1%	80,8%	90,9%	0,941
2	NB	AGE+OCT	85,7%	80,8%	88,6%	0,932
3	CTR	AGE+OCT	88,6%	80,8%	93,2%	0,910
4	BAG	AGE+OCT	85,7%	76,9%	90,9%	0,905
		Média	86,8%	79,8%	90,9%	0,922
1	NB	IOP+AGE+OCT	85,7%	80,8%	88,6%	0,932
2	BAG	IOP+AGE+OCT	84,3%	73,1%	90,9%	0,865
3	CTR	IOP+AGE+OCT	88,6%	76,9%	95,5%	0,818
4	MLP	IOP+AGE+OCT	80,0%	69,2%	86,4%	0,804
		Média	84,6%	75,0%	90,3%	0,855
1	BAG	AGE+OCT+SAP	85,7%	76,9%	90,9%	0,900
2	MLP	AGE+OCT+SAP	77,1%	80,8%	75,0%	0,855
3	NB	AGE+OCT+SAP	48,6%	96,2%	20,5%	0,621
4	CTR	AGE+OCT+SAP	70,0%	88,5%	59,1%	0,619
		Média	70,4%	85,6%	61,4%	0,749
1	MLP	IOP+AGE+OCT+SAP	75,7%	73,1%	77,3%	0,844
2	BAG	IOP+AGE+OCT+SAP	82,9%	73,1%	88,6%	0,842
3	CTR	IOP+AGE+OCT+SAP	72,9%	88,5%	63,6%	0,628
4	NB	IOP+AGE+OCT+SAP	48,6%	96,2%	20,5%	0,617
		Média	70,0%	82,7%	62,5%	0,732
1	BAG	AGE+SAP	55,7%	84,6%	38,6%	0,761
2	CTR	AGE+SAP	64,3%	84,6%	52,3%	0,724
3	MLP	AGE+SAP	54,3%	76,9%	40,9%	0,680
4	NB	AGE+SAP	45,7%	92,3%	18,2%	0,572
		Média	55,0%	84,6%	37,5%	0,684
1	MLP	IOP+AGE+SAP	55,7%	76,9%	43,2%	0,658
2	BAG	IOP+AGE+SAP	54,3%	84,6%	36,4%	0,627
3	NB	IOP+AGE+SAP	45,7%	92,3%	18,2%	0,572
4	CTR	IOP+AGE+SAP	58,6%	76,9%	47,7%	0,568
		Média	53,6%	82,7%	36,4%	0,606

Legenda: CLS = Classificador, AROC = Área sob a curva ROC.

6 Considerações Finais

Neste trabalho de pesquisa foi desenvolvido um gerador de populações artificiais, denominado GLOR, com o objetivo de prover dados para treinamento de classificadores para uso no diagnóstico do glaucoma. Nos experimentos realizados, foram treinados classificadores dos tipos Naive-Bayes, *Multilayer Perceptrons*, Árvores de Decisão e a Árvores de Decisão com *Bagging*, e a avaliação realizada com dados de populações reais mostrou que é possível melhorar o desempenho desses classificadores e consequentemente do diagnóstico do Glaucoma.

A avaliação qualitativa dos resultados do GLOR implementada por médicos especialistas, aprovou o método de geração por indicar que os dados obtidos simulavam adequadamente resultados de exames reais. Contudo, algumas questões foram levantadas.

Um dos problemas identificados pelos oftalmologistas foi a carência de ponderação das reduções de espessuras da camada de fibras nervosas para os pontos do campo visual correspondentes. Segundo a experiência clínica, apesar da correspondência entre áreas afetadas estar correta no modelo matemático desenvolvido para correlação entre estrutura e função, as reduções de sensibilidade no campo visual esperadas para determinadas taxas de redução de espessuras da camada de fibras nervosas eram maiores nas áreas centrais do campo visual, e menores nos pontos periféricos. Isto se deve ao fato do modelo distribuir uniformemente a redução em dB da espessura da camada de fibras nervosas por todos os correspondentes pontos do campo visual associados através do mapeamento espacial empregado para correlacionar setores da cabeça do nervo óptico com pontos do campo visual. Para resolver este problema, uma regra de ponderação precisa ser desenvolvida para conferir mais peso aos pontos centrais e menos peso aos pontos periféricos do campo visual no cálculo da distribuição das reduções de sensibilidade.

Outra crítica efetuada pelos médicos foi dirigida às faixas de probabilidades dos gráficos *total deviation probability* e *pattern deviation probability*. Segundo a experiência clínica tais valores estavam aquém do esperado em alguns casos. Um dos motivos que pode ter gerado este problema é eventualmente baixa quantidade de exemplos de pacientes

normais no banco de dados normativo do SAP que foi reconstruído a partir dos resultados de 82 exames. Em algumas faixas etárias, por exemplo, menos de 5 exemplos estavam disponíveis, o que certamente reflete em estimativas de probabilidade que não condizem com a prática clínica. À medida que uma maior quantidade de resultados de exames de pacientes normais estiverem disponíveis para alimentar o banco de dados normativo do GLOR esta questão deve ser resolvida.

Estas duas questões apontadas pelos médicos serão incorporadas ao conjunto de requisitos de melhorias da próxima versão do GLOR, visando aprimorar o processo de geração de dados de SAP.

Já a avaliação quantitativa dos dados gerados pelo GLOR, conduzida através da comparação das características das populações esperadas frente às características das populações obtidas, apontaram uma significativa convergência entre as médias e desvios padrões tanto no grupo de indivíduos normais quanto nos indivíduos com glaucoma.

A avaliação gráfica das distribuições de frequência nas populações artificiais apresenta indícios claros de que tais distribuições seguem de forma aproximada as distribuições gaussianas esperadas.

Tais fatos indicam que quantitativamente o método Monte Carlo desenvolvido para o GLOR foi capaz de gerar um conjunto de dados que obedecem as características definidas a priori para suas distribuições estimadas de probabilidade. Tais características foram representadas por um conjunto de parâmetros empregados no diagnóstico de glaucoma, e definidas em termos de suas respectivas médias e desvios padrões para populações representativas de indivíduos normais e glaucomatosos.

As tarefas de treinamento dos classificadores baseados em Bayes, *Multilayer Perceptrons*, Árvores de Decisão e Árvores de Decisão com técnica de *Bagging*, foram conduzidas com 5000 instâncias geradas pelo GLOR, o que constitui um conjunto de dados de alta cardinalidade para o domínio do diagnóstico de glaucoma. Na prática, muitos anos de pesquisas seriam necessários para coletar esta quantidade de dados utilizando-se pacientes reais. Além do tempo elevado, os custos associados seriam bastante significativos. Este conjunto de dados foi formado inicialmente por 178 atributos como idade, pressão intra-ocular, dados do exame de tomografia de coerência óptica de alta definição, dados de perimetria computadorizada padrão e a classe correspondente da instância (G para glaucoma e N para normal). Em seguida foi subdividido de acordo com subconjuntos de atributos previamente selecionados, dando origem a 6 datasets.

Como cada um dos 4 classificadores foi treinado com todos os datasets, 24 testes foram

executados utilizando a estratégia de validação cruzada. A análise dos resultados destes testes revelou um desempenho elevado e generalizado em relação a todos os 4 critérios empregados: acurácia total, sensibilidade, especificidade e área sob curvas ROC. Este fato gerou a suspeita de sobre-ajuste (*overfitting*) dos classificadores, causada pela eventual presença de ruído nos dados de treinamento.

Quando os classificadores foram submetidos a novo processo de testes, porém com dados inéditos de 70 pacientes reais, a suspeita de sobre-ajuste foi confirmada. A análise dos resultados dos testes evidenciou que em determinadas combinações de classificador e atributos empregados no treinamento, o desempenho ficou muito aquém dos resultados dos testes com dados artificiais. Das 24 combinações entre classificador e dataset empregados nos testes, 3 falharam por apresentarem áreas sob a curva ROC inferiores a 0,60. O classificador utilizado em 2 destes 3 testes foi o classificador Bayesiano, enquanto que os datasets possuíam predominantemente dados de SAP, e nenhum atributo estrutural (OCT). Das 12 situações de testes que obtiveram áreas sob curvas ROC inferiores a 0,80, somente duas continham dados de OCT, mas mesmo assim em datasets com a presença dos atributos de SAP.

Os resultados satisfatórios dos classificadores *Bagged-CTREE* e MLP com datasets formados pela integração dos dados anatômicos e funcionais (AGE + OCT + SAP e IOP + AGE + OCT + SAP), onde o melhor classificador (*Bagged-CTREE* sobre dataset AGE + OCT + SAP) obteve área sob a curva ROC de 0,900, acurácia total de 85,7%, sensibilidade de 76,9% e especificidade de 90,9%, resultados relativamente próximos aos do melhor classificador entre os avaliados (MLP sobre dataset AGE + OCT). Desta forma, é possível supor que com o refinamento previsto do modelo matemático de relacionamento entre estrutura e função do gerador possa produzir dados de SAP com menos ruídos, representando com mais acurácia o que ocorre no mundo real. Assim, dados de melhor qualidade poderão ser empregados no treinamento de classificadores. Assim, é provável que a sensibilidade e especificidade destes classificadores cresçam no futuro.

Já os valores das áreas sob curvas ROC obtidos pelos 4 classificadores: MLP (0,941), Naive-Bayes (0,932), CTREE (0,910) e *Bagged-CTREE* (0,905), quando avaliados sobre o conjunto de dados contendo os somente atributos idade e espessuras da camada de fibras nervosas (dataset AGE+OCT), podem ser classificados como excelentes, uma vez que superaram o valor limiar de 0,90.

Tal fato por si só constitui forte evidência de que, apesar de apresentar alguns problemas para geração de dados de SAP, o gerador GLOR foi capaz de produzir artificialmente

dados representativos de populações reais de indivíduos normais e com glaucoma para treinamento de classificadores no diagnóstico de glaucoma.

Diante do exposto, pode-se concluir que o GLOR pode ser empregado satisfatoriamente para simular dados epidemiológicos, clínicos, anatômicos e funcionais com qualidade e quantidade necessária para o treinamento de classificadores.

Desta forma, o GLOR pode ser empregado satisfatoriamente em experimentos para elevar a sensibilidade e especificidade de testes diagnósticos de glaucoma através da aplicação de técnicas de aprendizagem de máquina em tarefas de classificação, reduzindo significativamente o tempo e custos associados á coleta de dados de pacientes reais.

6.1 Contribuições deste Trabalho

Entre as contribuições apresentadas por este trabalho é possível destacar:

1. Criação de um método Monte Carlo inédito para geração artificial de dados epidemiológicos, anatômicos e funcionais para treinamento de classificadores no diagnóstico de glaucoma, reduzindo significativamente o tempo e custos associados á coleta de dados de pacientes reais.
2. Geração de banco de dados normativo público do Instrumento Cirrus HD-OCT, indisponível na literatura;
3. Geração de um banco de dados normativo público para SAP, para cálculo de probabilidades nos gráficos de *total deviation probability* e *pattern standard deviation probability*;
4. Criação de uma saída gráfica que une dados epidemiológicos, estruturais e funcionais em um só documento, facilitando a interpretação das informações pelos médicos especialistas;
5. Desenvolvimento de um modelo matemático de correlação entre espessura de camada de fibras nervosas e pontos do mapa de campo visual (modelo de correlação entre estrutura e função).

6.2 Trabalhos Futuros

Face aos enormes desafios em aberto que o diagnóstico médico apresenta, é possível sugerir a continuidade da presente pesquisa através dos seguintes trabalhos:

1. Adicionar mais fontes de informações médicas obtidas através de outros métodos ao GLOR, tais como SWAP e FDT, para diagnóstico funcional, e HRT e GDx-VCC, para diagnóstico anatômico;
2. Incorporar ao GLOR a possibilidade de utilização de distribuições de probabilidade diferentes da Normal no processo de geração inversa dos dados;
3. Empregar os classificadores treinados artificialmente em estudos longitudinais em pacientes reais para avaliação do seu desempenho;
4. Aprimorar o modelo matemático de correlação entre estrutura e função do gerador;
5. Avaliar o desempenho de outros classificadores no diagnóstico de glaucoma, como os híbridos, mediante treinamento baseado em dados artificiais;
6. Incorporar possíveis modelos matemáticos de progressão do glaucoma ao GLOR para que cenários de evolução da doença possam ser simulados computacionalmente;
7. Desenvolver classificadores capazes de prever o grau de risco de progressão da doença ao longo do tempo baseados em experiência;
8. Adaptar o GLOR para gerar dados para diagnóstico de outras doenças crônicas como a retinopatia diabética e a degeneração macular.

Referências

- ANDERSON, D. R.; DRANCE, S. M.; SCHULZER, M. Natural history of normal-tension glaucoma. *Ophthalmology*, v. 108(2), p. 247–53, 2001.
- AYDIN, A.; WOLLSTEIN, G.; PRICE, L. L.; FUJIMOTO, J. G.; SCHUMAN, J. S. Optical coherence tomography assessment of retinal nerve fiber layer thickness changes after glaucoma surgery. *Ophthalmology*, v. 110(8), p. 1506–11, 2003.
- AZUARA-BLANCO, A.; COSTA, V. P.; WILSON, R. P. *Handbook of Glaucoma*. [S.l.]: Martin Dunitz, 2002.
- BANK, I. *The World Bank: World Development Report 1993: Investing in health*. [S.l.]: Oxford University, 1993.
- BARBOZA, W. L.; JR, R. S. Estudo de valores estereométricos em olhos glaucomatosos e normais pela tomografia neuro-retiniana: projeto piloto. *Revista Brasileira de Oftalmologia*, v. 57(3), p. 187–92, 1998.
- BOWD, C.; HAO, J.; TAVARES, I. M.; MEDEIROS, F. A.; ZANGWILL, L. M.; LEE, T. W.; SAMPLE, P. A.; WEINREB, R. N.; GOLDBAUM, M. H. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Investigative Ophthalmology & Visual Science*, March , Vol. 49, No. 3, v. 49, p. 945–953, 2008.
- BOWD, C.; WEINREB, R. N.; WILLIAMS, J. M.; ZANGWILL, L. M. The retinal nerve fiber layer thickness in ocular hypertensive, normal, and glaucomatous eyes with optical coherence tomography. *Archives of Ophthalmology*, v. 118, p. 22–6, 2000.
- BRANDT, J. D.; BEISER, J. A.; KASS, M. A.; GORDON, M. O. Central corneal thickness in the ocular hypertension treatment study (ohts). ; :. *Ophthalmology*, v. 108(10), p. 1779–88, 2001.
- BRENTON, R. S.; PHELPS, C. D. The normal visual field on the humphrey field analyser. *Ophthalmologica*, v. 193, p. 56–74, 1986.
- BURGANSKY-ELIASH, S. e. a. Optical coherence tomography machine learning classifiers for glaucoma detection: A preliminary study. *Investigative in Ophthalmology and Visual Science*, v. 46, p. 4147–4152, 2005.
- CANNER, J. K.; CHIANG, Y. P.; JAVITT, J. C. A monte carlo based simulation network model for a chronic progressive disease: The case of diabetic retinopathy. In: *Proceedings of the 199.2 Winter Simulation Conference*. [S.l.: s.n.], 1992.
- CHAUHAN, B. C.; MCCORMICK, T. A.; NICOLELA, M. T.; LEBLANC, R. P. Optic disc and visual field changes in a prospective longitudinal study of patients with glaucoma: comparison of scanning laser tomography with conventional perimetry and optic disc photography. *Archives of Ophthalmology*, v. 119(10), p. 1492–9, 2001.
- CHAUVIN, Y.; RUMELHART, D. *Backpropagation: Theory, Architectures, and Applications*. [S.l.]: Lawrence Erlbaum Associates, 1995.

- CLARK, P.; NIBLETT, R. The cn2 induction algorithm. *Machine Learning*, v. 3, p. 261-284, 1989.
- COSTA, V. P. *Diagnóstico precoce e de progressão do glaucoma baseados em sistemas de aprendizagem de classificadores híbridos*. May 2007. Proposta de projeto de pesquisa à Fapesp - projeto número 2007/51281-9.
- CRACKNELL, K. P. B.; FARNELL, D. J. J.; GRIERSON, I. Monte carlo simulation of latanoprost induced iris darkening. *Computer Methods and Programs in Biomedicine*, v. 87, p. 93-103, 2007.
- CRAVEN, M. W. *Extracting comprehensible models from trained neural networks*. Tese (Doutorado) — Department of Computer Science: University of Wisconsin-Madison, 1996.
- CRAVEN, M. W.; SHAVLIK, J. W. Using sampling and queries to extract rules from trained neural networks. In: *Proceedings of the 11th International Conference on Machine Learning*. [S.l.]: Morgan-Kaufmann, 1994. p. 37-45.
- DELGADO, M. F.; NGUYEN, N. T.; COX, T. A.; SINGH, K.; LEE, D. A.; DUEKER, D. K. e. a. Automated perimetry: a report by the american academy of ophthalmology. *Ophthalmology*, v. 109(12), p. 2362-74, 2002.
- DEMSAR, J.; ZUPAN, B.; LEBAN, G. *Orange: From Experimental Machine Learning to Interactive Data Mining*,. 2004. White Paper (www.aillab.si/orange), Faculty of Computer and Information Science, University of Ljubljana.
- DIAS, J. P.; LIMA, F. E. *Glaucoma*. [S.l.: s.n.], 1998.
- DIELEMANS, I. e. a. The prevalence of primary open-angle glaucoma in a population-based study in the netherlands. *Ophthalmology*, v. 101, p. 1851, 1994.
- DUARTE, F. B.; PARANHOS, A.; GALHARDO, R. A.; ABDANUR, L. R. A.; PRATA, J. A.; MELLO, P. A. A. Reprodutividade do exame de medida da camada de fibras nervosas da retina por meio da polarimetria de varredura a laser. *Arquivo Brasileiro de Oftalmologia*, v. 62(5), p. 555-60, 1999.
- FAWCETT, T. *ROC Graphs: Notes and Practical Considerations for Researchers*. [S.l.], 2004.
- FAYYAD, U. M.; IRANI, K. B. Multi-interval discretization of continuous-valued attributes for classification learning. In: BAJESY, R. (Ed.). *Proceedings of the 13th International Joint Conference on Artificial Intelligence*. [S.l.]: Morgan-Kaufmann, 1993. p. 1022-1027.
- FEUER, W. J.; PARRISH, R. K.; SCHIFFMAN, J. C.; ANDERSON, D. R.; BUDENZ, D. L.; WELLS, M. C. e. a. The ocular hypertension treatment study: reproducibility of cup/disk ratio measurements over time at an optic disc reading center. *American Journal of Ophthalmology*, v. 133(1), p. 19-28, 2002.
- FU, L. M. Knowledge-based connectionism for revising domain theories. *IEEE Transactions on System Management and Cybernetics*, v. 23(1), p. 173-182, 1993.
- GOLDBAUM, M. H.; AL et. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Investigative Ophthalmology and Visual Science*, v. 43, p. 162-169, 2002.
- GREANEY, M. J.; HOFFMAN, D. C.; GARWAY-HEATH, D. F.; NAKLA, M.; COLEMAN, A. L.; CAPRIOLI, J. Comparison of optic nerve imaging methods to

- distinguish normal eyes from those with glaucoma. *Investigative Ophthalmology Visual Science*, v. 43(1), p. 140–5, 2002.
- GREEN, D. M.; SWETS, J. M. *Signal detection theory and psychophysics*. [S.l.: s.n.], 1966.
- GREENFIELD, D. S.; KNIGHTON, R. W.; FEUER, W. J.; SCHIFFMAN, J. C.; ZANGWILL, L.; WEINREB, R. N. Correction for corneal polarization axis improves the discriminating power of scanning laser polarimetry. *American Journal Ophthalmology*, v. 134(1), p. 27–33, 2002.
- HARWERTH, R. S.; WHEAT, J. L. Modeling the effects of aging on retinal ganglion cell density and nerve fiber layer thickness. *Graefes Arch Clin Exp Ophthalmol*, v. 246, p. 305314, 2008.
- HAYKIN, S. *Neural Networks: a comprehensive foundation*. 2nd. ed. [S.l.: s.n.], 1999.
- HECKERMAN, D.; GEIGER, D.; CHICKERING, D. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, v. 3, p. 197243, 1995.
- HEIDELBERG-ENGINEERING. 2009. Disponível em <http://www.heidelbergengineering.com/products/hrt-glaucoma-module/product-literature>. Último acesso em 30/01/2009.
- HEIJL, A.; LESKE, M. C.; BENGTTSSON, B.; HYMAN, L.; HUSSEIN, M. Reduction of intraocular pressure and glaucoma progression: results from the early manifest glaucoma trial. ;. *Archives of Ophthalmology*, v. 120(10), p. 1268–79, 2002.
- HEIJL, A.; LESKE, M. C.; BENGTTSSON, B.; HUSSEIN, M. Measuring visual field progression in the early manifest glaucoma trial. *Acta Ophthalmol Scand*, v. 81(3), p. 286–93, 2003.
- HOOD, D. C.; ANDERSON, S. C.; WALL, M.; KARDON, R. H. R. H. Structure versus function in glaucoma: An application of a linear model. *Investigative Ophthalmology & Visual Science*, v. 48, p. 3662–3668, 2007.
- HOTHORN, H.; LAUSEN, B. Bagging tree classifiers for laser scanning images: a data- and simulation-based strategy. *Artificial Intelligence in Medicine*, vol.27, n. 1, p. 65–79, 2003.
- HUANG, M. L.; CHEN, H. Y.; HUANG, J. J. Glaucoma detection using adaptive neuro-fuzzy inference system. *Expert Systems with Applications*, v. 32, p. 458468, 2007.
- HUANG, M. L.; CHEN, H. Y.; LIN, J. C. Rule extraction for glaucoma detection with summary data from stratusoct. *Investigative Ophthalmology & Visual Science*, v. 48, p. 244–250, 2007.
- JAFFE, G. J.; CAPRIOLI, J. Optical coherence tomography to detect and manage retinal disease and glaucoma. *American Journal of Ophthalmology*, v. 137, p. 156–69, 2004.
- JOHNSON, C. A.; ADAMS, A. J.; LEWIS, R. A. Patterns of early glaucomatous visual field loss. *Archives of Ophthalmology*, v. 111, p. 654–650, 1993.
- KASS, M. A.; HEUER, D. K.; HIGGINBOTHAM, E. J.; JOHNSON, C. A.; KELTNER, J. L.; MILLER, J. P. e. a. The ocular hypertension treatment study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Archives of Ophthalmology*, v. 120(6), p. 701–13; discussion 829–30, 2002.

- KINGMAN, S. Glaucoma is second leading cause of blindness globally. *Bull World Health Organ*, v. 82(11), p. 887–888, 2004.
- KLEIN, B. E. K. e. a. Prevalence of glaucoma: The beaver dam eye study. *Ophthalmology*, v. 99, p. 1449, 1992.
- KONONENKO, I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, v. 23, p. 89–109, 2001.
- LANDERS, J. A.; GOLDBERG, I.; GRAHAM, S. L. Detection of early visual field loss in glaucoma using frequency-doubling perimetry and short-wavelength automated perimetry. *Archives of Ophthalmology*, v. 121(12), p. 1705–10, 2003.
- LAUANDE-PIMENTEL, R.; CARVALHO, R. A.; OLIVEIRA, H. C.; GONÇALVES, D. C.; SILVA, L. M.; COSTA, V. P. Discrimination between normal and glaucomatous eyes with visual field and scanning laser polarimetry measurements. *Br J Ophthalmology*, v. 85(5), p. 586–91, 2001.
- LESKE, M. C. e. a. The barbados eye study: prevalence of open angle glaucoma. *Archives of Ophthalmology*, v. 112, p. 821, 1994.
- LIU, J. H.; ZHANG, X.; KRIPKE, D. F.; WEINREB, R. N. Twenty-four-hour intraocular pressure pattern associated with early glaucomatous changes. *Investive in Ophthalmology and Visual Science*, v. 44(4), p. 1586–90, 2003.
- MASON, R. P. e. a. National survey of the prevalence and risk factors of glaucoma in st. lucia, west indies. i prevalence findings. *Ophthalmology*, v. 65, p. 1363, 1989.
- MEDEIROS, F. A.; SAMPLE, P. A.; WEINREB, R. N. Frequency doubling technology perimetry abnormalities as predictors of glaucomatous visual field loss. *American Journal of Ophthalmology*, v. 137(5), p. 863–71, 2004.
- MEDEIROS, F. A.; WEINREB, R. N. Medical backgrounders: glaucoma. *Drugs Today (Barc)*, v. 38(8), p. 563–70, 2002.
- MEDEIROS, F. A. e. a. Comparision of the gdx vcc scanning laser polarimeter, hrt ii confocal scanning laser ophthalmoscope, ans stratus oct optical coherence tomograph for the detection of glaucoma. *Archives of Ophthalmology*, vol 122, p. 827–837, 2004.
- MERCK & CO. INC. *Manual Merck de Informação Médica - Saúde Para a Família*. Disponível em http://www.msd-brazil.com/msd43/mm_anual/mm_sec2026.htm, último acesso em 08/05/2007., 2004.
- MIERSWA, I.; WURST, M.; KLINKENBERG, R.; SCHOLZ, M.; EULER, T. Yale: Rapid prototyping for complex data mining tasks. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06)*. [S.l.: s.n.], 2006.
- MITCHELL, T. M. *Machine Learning*. [S.l.]: McGraw-Hill, 1997.
- MITCHELL, T. M. Machine learning and data mining. *Commun. ACM*, ACM Press, New York, NY, USA, v. 42, n. 11, p. 30–36, 1999. ISSN 0001-0782.
- MOK, K. H.; LEE, V. W. Nerve fiber analyser and short-wavelength automated perimetry in glaucoma suspects. *Ophthalmology*, v. 107, p. 2101–2104, 2000.
- MOURA, C. R.; MELLO, P. A. A.; PRATA, J. A. Scanning laser polarimetry in glaucomatous and nonglaucomatous brazilian subjects. *Arquivo Brasileiro de Oftalmologia*, v. 66(6), p. 749–754, 2003.

- MUGGLETON, S. Bayesian inductive logic programming. In: *COLT '94: Proceedings of the seventh annual conference on Computational learning theory*. New York, NY, USA: ACM Press, 1994. p. 3–11. ISBN 0-89791-655-7.
- MURTHY, S. K.; KASIF, S.; SALZBERG, S. A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, v. 2, p. 1–33, 1994.
- OBUCHOWSKI, N. A. Receiver operating characteristic curves and their use in radiology. *Radiology*, v. 229 (1), p. 38, 2003.
- PACHECO-CUTILLAS, M.; SAHRAIE, A.; EDGAR, D. F. Acquired colour vision defects in glaucoma their detection and clinical significance. ; :. *British Journal of Ophthalmology*, v. 83, p. 1396–1402, 1999.
- PACZKA, J. A.; FRIEDMAN, D. S.; QUIGLEY, H. A.; BARRON, Y.; VITALE, S. Diagnostic capabilities of frequency-doubling technology, scanning laser polarimetry, and nerve fiber layer photographs to distinguish glaucomatous damage. *American Journal of Ophthalmology*, v. 131(2), p. 188–97, 2001.
- PEPE, M. S. *The statistical evaluation of medical tests for classification and prediction*. [S.l.: s.n.], 2003.
- POLO, V.; LARROSA, J. M.; PABLO, L. E. e. Correlation of functional and structural measurements in eyes suspected of having glaucoma. *Journal of Glaucoma*, v. 8, p. 172–176, 1999.
- PRATA, J. A.; MOURA, C. R.; MANDELO, M. H.; MAIA, M. B.; GALHARDO, R. A.; MELLO, P. A. A. Análise da camada de fibras nervosas da retina com gdx: validade diagnóstica em glaucoma primário de ângulo aberto e suspeitos de glaucoma. *Arquivo Brasileiro de Oftalmologia*, v. 64(4), p. 337–340, 2001.
- QUIGLEY, H. A.; HOGMAN, R. M.; ADDICKS, E. M.; et al. Morphologic changes in the lamina cribrosa correlated with neural loss in open-angle glaucoma. *American Journal of Ophthalmology*, v. 96, p. 673–691, 1983.
- QUINLAN, J. *C4.5: Programs for Machine Learning*. [S.l.]: Morgan Kaufmann, 1993.
- QUINLAN, J. R. Induction of decision trees. *Machine Learning*, I (1), p. 81–106, 1983.
- RAKOTOMALALA, R. Tanagra: a free software for research and academic purposes. In: *Proceedings of EGC'2005, RNTI-E-3*. [S.l.: s.n.], 2005. v. 2, p. 697–702. (in French).
- SANCHEZ-GALEANA, C.; BOWD, C.; BLUMENTHAL, E. Z.; GOKHALE, P. A.; ZANGWILL, L. M.; WEINREB, R. N. Using optical imaging summary data to detect glaucoma. *Ophthalmology*, v. 108(10), p. 1812–8, 2001.
- SCHUMAN, J. S.; HEE, M. R.; PULIAFITO, C. A.; WONG, C.; PEDUT-KLOIZMAN, T.; LIN, C. P. e. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. *Archives of Ophthalmology*, v. 113, p. 586–96, 1995.
- SOMMER, A.; KATZ, J.; QUIGLEY, H. A.; MILLER, N. R.; ROBIN, A. L.; RICHTER, R. C. e. a. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Archives of Ophthalmology*, v. 109(1), p. 77–83, 1991.
- SPACKMAN, K. A. Signal detection theory: Valuable tools for evaluating inductive learning. In: MATEO, C. M. K. S. (Ed.). *Proceedings of the Sixth International Workshop on Machine Learning*. [S.l.: s.n.], 1989. p. 160163.

- STARK, H.; WOODS, J. W. *Probability and Random Process with Applications to Signal Processing*. [S.l.]: Pearson Education, 2002.
- SUNG, V. C.; BHAN, A.; VERNON, S. A. Agreement in assessing optic discs with a digital stereoscopic optic disc camera (discam) and heidelberg retina tomograph. *British Journal of Ophthalmology*, v. 86(2), p. 196–202, 2002.
- TANITO, M.; ITAI, N.; OHIRA, A.; CHIHARA, E. Reduction of posterior pole retinal thickness in glaucoma detected using the retinal thickness analyzer. *Ophthalmology*, v. 111(2), p. 265–75, 2004.
- TAPE, T. G. *Interpreting Diagnostic Tests*. [S.l.], 2006. Disponível on-line em <http://gim.unmc.edu/dxtests/Default.htm> (último acesso em 02/05/2007).
- TAVARES, I. M.; JR, L. A. S. M.; MOTTA, M. M. S.; HIGA, F. S.; JR, A. P.; MELLO, P. A. A. Tonometria de não-contato: variação com medidas repetidas com e sem colírio anestésico. *Arq. Bras. Oftalmologia*, v. 66(6), p. 835–8, 2003.
- TIELSH, J. M. e. a. Racial variation in the prevalence of primary open-angle glaucoma. the baltimore eye survey. *JAMA*, v. 266, p. 369, 1991.
- TOWELL, G.; SHAVLIK, J.; NOORDEWIJER, M. Refinement of approximate domain theories by knowledge-based neural networks. In: *Proceeding of the Eighth National Conference on Artificial Intelligence*. [S.l.]: MIT Press, 1990. p. 861–866.
- TUCKER, A. e. a. A spatio-temporal bayesian network classifier for understanding visual field deterioration. *Artificial Intelligence in Medicine*, available on line, p. 1–15, 2004.
- UTGOFF P E ANS BRODLEY, C. E. *Linear machine decision trees*. [S.l.], 1991. (COINS Technical Report).
- WEINREB, R. N.; KHAW, P. T. Primary open-angle glaucoma. *Lancet*, v. 363(9422), p. 711–20, 2004.
- WEINREB, R. N.; ZANGWILL, L.; BERRY, C. C.; BATHIJA, R.; SAMPLE, P. A. Detection of glaucoma with scanning laser polarimetry. *Archives of Ophthalmology*, v. 116, p. 1583–1589, 1998.
- WHITACRE, M. M.; STEIN, R. Sources of error with use of goldmann-type tonometers. *Surv Ophthalmology*, v. 38(1), p. 1–30, 1993.
- WITTEN, I. H.; FRANK, E. *Data Mining: Practical machine learning tools and techniques*. 2nd. ed. [S.l.: s.n.], 2005.
- Y HOTHORN T, P. A. e. a. M. C. New glaucoma classification method based on standard heidelberg retina tomograph parameters by bagging classification tree. *J Glaucoma*, v. 12, p. 340346, 2003.
- ZANGWILL, L. M.; BOWD, C.; BERRY, C. C.; WILLIAMS, J.; BLUMENTHAL, E. Z.; SANCHEZ-GALEANA, C. A. e. a. Discriminating between normal and glaucomatous eyes using the heidelberg retina tomograph, gdx nerve fiber analyzer, and optical coherence tomograph. 001;. *Archives of Ophthalmology*, v. 119(7), p. 985–93, 2001.
- ZANGWILL, L. M. e. a. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Investigative Ophthalmology and Visual Science*, v. 45(9), p. 3144–3151, 2004.
- ZEISS, C. 2009. Disponível em <http://www.meditec.zeiss.com>. Último acesso em 30/01/2009.

ZWEIG, M. H.; CAMPBELL, G. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, v. 39 (8), p. 561577, 1993.

Apêndice A – Projeto FAPESP: Critérios de Inclusão e Exclusão

A seguir são apresentados os critérios de inclusão e exclusão de pacientes do estudo denominado **Diagnóstico precoce e de progressão do glaucoma baseados em sistemas de aprendizagem de classificadores híbridos - Projeto FAPESP número 2007/51281-9**, com aprovação sem restrições pelo Comitê de Ética em Pesquisa (CEP) da Faculdade de Ciências Médicas (FCM) da Universidade de Campinas (UNICAMP) conforme parecer número 406/2005. Todos os pacientes participaram mediante assinatura do Termo de Consentimento Livre e Esclarecido.

A.1 Grupo 1 - Indivíduos Normais

A.1.1 Critérios de Inclusão

O grupo de indivíduos normais será constituído por 300 indivíduos membros da comunidade universitária, amigos ou cônjuges de indivíduos glaucomatosos, indivíduos acompanhados em outros setores do Serviço de Oftalmologia do Hospital de Clínicas da Universidade Estadual de Campinas, desde que preencham os seguintes critérios de inclusão: acuidade visual com melhor correção = 20/30; equivalente esférico = ± 5 dioptrias; idade > 40 anos (sem limite superior); 2 campos visuais confiáveis (definidos como tendo perdas de fixação < 20%, respostas falso positivas < 33% e respostas falso negativas < 33%) e normais (sem qualquer alteração descrita nos critérios de inclusão do Grupo 2 - Indivíduos Glaucomatosos); exame de disco óptico normal; gonioscopia com ângulo aberto; PIO < 21 mmHg.

A.1.2 Critérios de Exclusão

Não poderão ser incluídos no grupo de indivíduos normais aqueles que apresentem qualquer um dos seguintes critérios: história de doença sistêmica ou ocular que possa interferir

nos resultados dos exames (SAP, FDT, OCT) (e.g. retinopatia diabética, compressão das vias ópticas, doenças desmielinizantes, coriorretinite, neurites); pseudofacia ou afacia; catarata com opacidade significativa de acordo com os critérios do Lens Opacification Classification System III, definida como opacidade máxima nuclear (NC3, NO3), cortical (C3) e subcapsular (P3); disco óptico suspeito de glaucoma (ver descrição nos critérios de inclusão do Grupo 2 - Indivíduos Glaucomatosos); história familiar de glaucoma.

A.2 Grupo 2 - Indivíduos Glaucomatosos

A.2.1 Critérios de Inclusão

Um grupo de 300 indivíduos glaucomatosos será constituído por indivíduos em acompanhamento no Setor de Glaucoma do Serviço de Oftalmologia do Hospital de Clínicas da Universidade Estadual de Campinas, que preencham os seguintes critérios de inclusão: diagnóstico clínico de glaucoma primário de ângulo aberto (GPAA); histórico de duas ou mais medidas de pressão ocular acima de 21 mmHg; acuidade visual com melhor correção = 20/30 em pelo menos um olho; equivalente esférico = ± 5 dioptrias; idade > 40 anos (sem limite superior); disco óptico tipicamente glaucomatoso (e.g. escavação maior que 0,6, defeitos focais da rima tipo notch, fossetas adquiridas do nervo óptico, hemorragias peripapilares, vasos em passarela ou desnudamento de vaso circumlineal e atrofia peripapilares); 2 perimetria acromática confiáveis (definidos como tendo perdas de fixação < 20%, respostas falso positivas < 33% e respostas falso negativas < 33%) e reprodutíveis apresentando defeito de campo visual glaucomatoso, definido como Glaucoma Hemifield Test fora dos limites normais, 3 pontos adjacentes, não periféricos, com $p < 0.05$ no gráfico Pattern Deviation, com, pelo menos, um ponto em $p < 0,01$, Pattern Standard Deviation < 0,05. Os pacientes serão classificados como apresentando dano precoce, moderado ou avançado de acordo com os seguintes critérios: a) dano precoce: MD > -6 dB; b) dano moderado: MD entre -6 e -15 dB; c) dano avançado: MD = -15 dB.

A.2.2 Critérios de Exclusão

Indivíduos glaucomatosos que preencherem um ou mais dos seguintes critérios serão excluídos do estudo: história de doença sistêmica ou ocular que possa interferir nos resultados dos exames (SAP, FDT, OCT) (e.g. retinopatia diabética, compressão das vias ópticas, doenças desmielinizantes, coriorretinite, neurites); pseudofacia ou afacia; catarata com opacidade significativa de acordo com os critérios do Lens Opacification Classifica-

tion System III, definida como opacidade máxima nuclear (NC3, NO3), cortical (C3) e subcapsular (P3).

A.3 Grupo 3 - Hipertensos Oculares

Um grupo de 100 hipertensos oculares será constituído por indivíduos em acompanhamento no Setor de Glaucoma do Serviço de Oftalmologia do Hospital de Clínicas da Universidade Estadual de Campinas. Os critérios de inclusão para este grupo são idênticos aos já descritos para indivíduos normais, com exceção do valor da PIO, que deverá ser maior que 22 mmHg em pelo menos 2 ocasiões. Os critérios de exclusão são idênticos aos já descritos para os indivíduos normais.