



**Rita Filipa Gomes Figueiredo**  
Licenciatura em Ciências de Engenharia Biomédica

## **Validação de Protocolo em Programas de Rastreio**

Dissertação para obtenção do Grau de Mestre em  
Engenharia Biomédica

Orientador: José Manuel Fonseca, Professor auxiliar, FCT-UNL;  
Co-orientador: Carlos Manta de Oliveira, Retmarker S.A.-  
Coimbra;

Júri:

Presidente: Prof. Doutora Carla Maria Quintão Pereira  
Arguente: Prof. Doutor Pedro Manuel Cardoso Vieira  
Vogal: Prof. Doutor José Manuel da Fonseca



## **Copyright**

Copyright©2014 - Todos os direitos reservados. Rita Filipa Gomes Figueiredo.  
Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



*“By believing passionately in something that still does not exist,  
we create it. The nonexistent is whatever we have not  
sufficiently desired.”*

Franz Kafka



## Agradecimentos

Gostaria de agradecer primeiramente ao Prof. Dr. José Manuel Fonseca não só pela oportunidade em realizar este estudo/projeto, no âmbito da realização da minha Dissertação de Mestrado, sob a sua orientação, bem como pelo grande apoio que me facultou durante a realização do mesmo.

Quero agradecer também ao Eng. Carlos Manta de Oliveira, que além de co-orientar a minha Tese de Mestrado, me prestou também apoio e possibilitou aprendizagem aquando a minha passagem pela na Retmarker-S.A.

Um obrigado à Critical Software S.A., em especial à Retmarker-S.A., pela possibilidade que me ofereceu em realizar um estágio na mesma, e pelos recursos disponibilizados para que eu pudesse realizar o meu estudo. Um grande agradecimento àqueles que não só nas horas mais atarefadas, mas também fora delas, me brindaram com a sua boa disposição, facilitando todo o meu processo de adaptação.

Não posso esquecer os meus amigos e colegas, em especial aqueles que me apoiaram e acarinharam ao longo destes fantásticos anos de estudo em Engenharia Biomédica. Rita, Tiago, Ricardo e todos os outros que de alguma maneira trouxeram algo bom à minha vida, um caloroso obrigado. Estarão sempre no meu coração, e apesar de saber o quão vão gozar comigo pelos sentimentalismos, sei que estarei sempre no vosso também, não mintam! Ao António um muito obrigado pela incansável vontade de ajudar e por todo o ânimo que me passou mesmo nas horas mais desencorajantes.

Por fim, mas de certeza, não menos importante, um enorme obrigado aos meus pais por todo o apoio e encorajamento, em especial à minha mãe, também minha melhor amiga, por tudo o que passou para me fazer aqui chegar. Por me ver de uma maneira que eu não consigo ver, por acreditar sempre em mim mesmo quando eu não acredito, por me apoiar em todos os passos tendo sempre a firme certeza que nunca vou cair, e “se”, se acontecer ela lá estará para me limpar as feridas e lágrimas e para me mostrar de novo o caminho, esquecendo-se das suas próprias feridas. Isto é fruto do meu trabalho, mas acima de tudo é fruto do trabalho da minha Mãe.



## Resumo

A Retinopatia Diabética (RD) é uma patologia ocular, consequência da Diabetes Mellitus, e que pode levar à perda de visão. Dado que nos primeiros estágios da doença, não ocorre manifestação sintomática, é de máxima importância a existência de programas estruturados de rastreio. Para diminuir a carga laboral que estes programas acarretam, pode-se recorrer à utilização de sistemas de deteção e classificação automática de RD que têm vindo a ser desenvolvidos. No entanto, estes demonstram algumas falhas, sendo que na origem destas pode estar a existência de imagens com qualidade não adequada. De maneira a reduzir as falhas destes sistemas, potenciando a sua utilização em programas de rastreio e automatizando assim todo o processo, foi conduzido este estudo, no qual se teve diretamente em consideração a qualidade das imagens, avaliada também de forma automática, utilizando a sua informação como parâmetro de entrada nos classificadores automáticos de RD. A avaliação automática da qualidade é constituída por 4 classificadores das características desta: Cor, Focagem, Contraste e Iluminação, sendo que a combinação destes dá origem a um último classificador, denominado Qualidade Geral. Desta forma, foi efetuada a fusão destes classificadores com os algoritmos de classificação automática de RD para todas as imagens, numa primeira abordagem, e desconsiderando as imagens marcadas pelo classificador Qualidade Geral como inadequadas, numa última abordagem.

Os resultados obtidos mostraram uma melhoria no conjunto dos indicadores sensibilidade e especificidade dos métodos de classificação automática de RD quando se procedeu à combinação destes com a Qualidade Geral. Mostraram ainda uma melhoria mais acentuada quando se usou informação de cada um dos classificadores das características que contribuem para a avaliação da Qualidade, em vez de uma informação única sobre qualidade. A presença de imagens de má qualidade mostrou ser prejudicial para o desempenho dos classificadores automáticos de RD, não tendo tanta influência no classificador resultante da fusão que produziu os melhores resultados do conjunto sensibilidade e especificidade.

Palavras-chave: retinopatia diabética, sistemas de classificação automática, avaliação automática de qualidade, fusão de classificadores



## Abstract

Diabetic retinopathy (DR) is an eye disease caused by Diabetes Mellitus, and can lead to vision loss. In the early stages of the disease symptomatic manifestation does not occur, therefore it is of the utmost importance the existence of structured screening programs. To reduce the work load these programs carry, automatic DR detection and classification systems have been developed. However, they show some faults, and can be due to the existence of inadequate quality images. In order to reduce failures of these systems, enhancing their use in screening programs and thus automating the entire process, this study was undertaken. In which it had in direct account the quality of the images, also evaluated automatically, using its information as an input parameter in DR automated classification systems. Automated quality assessment consists of four classifiers of the characteristics: Color, Focus, Contrast and Brightness, and the combination of these gives rise to a final classifier, called the Overall Quality. Thus, the fusion of these classifiers with the algorithms for DR automated classification for all images was performed in a first approach, and for the final approach, disregarding the images marked by the classifier Overall Quality as inadequate.

The results showed an improvement in the sensitivity of the methods of DR automated classification when we assessed the combination of these with the Overall Quality, and an even greater improvement when, instead of using the Overall Quality, used the four classifiers of Quality characteristics. The presence of low-quality images proved to be detrimental to the performance of DR automated classification systems, not having much influence on the classifier resulting from the merger that produced the best results of sensitivity.

Keywords: diabetic retinopathy, automated classification system, automated quality assessment, classifier fusion



## Índice de Conteúdos

Agradecimentos.....	vii
Resumo.....	ix
Abstract .....	xi
Índice de Conteúdos.....	xiii
Índice de Figuras .....	xv
Índice de Tabelas.....	xxiii
Lista de Abreviaturas .....	xxv
1 Introdução .....	1
1.1 Motivação.....	1
1.2 Estado da arte .....	3
2 Conceitos teóricos .....	11
2.1 Classificador automático .....	11
2.2 Qualidade de um Classificador.....	11
2.3 Indicadores de Performance .....	15
2.4 Árvores de decisão .....	16
2.4.1 Introdução.....	16
2.4.2 Indução de árvores de decisão .....	18
2.4.3 Algoritmo CART ( <i>Classification and Regression Trees</i> ).....	23
2.4.4 Extração de Regras de Classificação .....	24
3 Métodos.....	27
3.1 Base de dados .....	27
3.2 Processamento .....	30
3.3 Construção dos classificadores .....	30
3.3.1 Conjunto de treino .....	30
3.3.2 Material .....	32
4 Resultados .....	33
4.1 Apresentação de Resultados .....	33
Base de dados nº 1 .....	34
Base de dados nº 2.....	35
Base de dados nº 3.....	36
4.2 Análise de resultados.....	36
5 Conclusão e trabalho futuro .....	45
Referências.....	51

A . Anexo 1: Árvores de decisão obtidas .....	55
A.1 Base de Dados nº 1 .....	55
A.2 Base de Dados nº 2 .....	73
A.3 Base de Dados nº 3 .....	91

## Índice de Figuras

Figura 1.1.1 – Exemplo de imagem com qualidade adequada.....	2
Figura 1.1.2 – Exemplo de imagem com qualidade inadequada.....	2
Figura 1.2.1 – Diagrama de blocos do projeto.....	9
Figura 2.4.1 – Exemplo de árvore de decisão.....	16
Figura 2.4.2 – Indução por profundidade (HALMENSCHLAGER, 2002).....	18
Figura 2.4.3 – Indução por largura (HALMENSCHLAGER, 2002).....	19
Figura 3.3.1 – Diagrama da combinação de classificadores para a primeira abordagem.....	31
Figura 3.3.2 – Diagrama da combinação de classificadores para segunda abordagem.....	31
Figura A.1.1- árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=1 e FP =1.....	55
Figura A.1.2 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=4 e FP =1.....	55
Figura A.1.3 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=8 e FP =1.....	56
Figura A.1.4 -árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=4 e FP =1.....	57
Figura A.1.5 -árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=8 e FP =1.....	57
Figura A.1.6 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=4 e FP =1.....	58
Figura A.1.7 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=8 e FP =1.....	58
Figura A.1.8 -árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=1 e FP =1.....	59
Figura A.1.9 -árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=4 e FP =1.....	59
Figura A.1.10 -árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=8 e FP =1.....	59

Figura A.1.11 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=1 e FP =1 .....	60
Figura A.1.12 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=4 e FP =1 .....	60
Figura A.1.13 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=8 e FP =1 .....	61
Figura A.1.14 -árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=1 e FP =1 .....	62
Figura A.1.15 -árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=4 e FP =1 .....	62
Figura A.1.16 -árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=8 e FP =1 .....	63
Figura A.1.17 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=1 e FP =1 .....	64
Figura A.1.18 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=4 e FP =1 .....	65
Figura A.1.19 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=8 e FP =1 .....	65
Figura A.1.20 -árvore de decisão resultante da fusão: Retmarker + DMUC desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 .....	66
Figura A.1.21 -árvore de decisão resultante da fusão: Retmarker + DMUC desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 .....	66
Figura A.1.22 -árvore de decisão resultante da fusão: Retmarker + DMUC desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 .....	66
Figura A.1.23 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 .....	67
Figura A.1.24 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 .....	67
Figura A.1.25 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 .....	68
Figura A.1.26 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 .....	68

Figura A.1.27 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 ..... 69

Figura A.1.28 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1; Parte 1..... 70

Figura A.1.29 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1; Parte 2..... 71

Figura A.1.30 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1; Parte 3..... 71

Figura A.1.31 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 ..... 72

Figura A.1.32 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 ..... 72

Figura A.2.1 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=1 e FP =1 ..... 73

Figura A.2.2 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=4 e FP =1 ..... 73

Figura A.2.3 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=8 e FP =1 ..... 73

Figura A.2.4 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=1 e FP =1 ..... 74

Figura A.2.5 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=4 e FP =1 ..... 74

Figura A.2.6 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=8 e FP =1 ..... 74

Figura A.2.7 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=1 e FP =1 ..... 75

Figura A.2.8 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=4 e FP =1 ..... 75

Figura A.2.9 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=8 e FP =1 ..... 76

Figura A.2.10 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=1 e FP =1 .....	77
Figura A.2.11 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=4 e FP =1 .....	77
Figura A.2.12 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=8 e FP =1 .....	77
Figura A.2.13 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=1 e FP =1 .....	78
Figura A.2.14 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=4 e FP =1 .....	78
Figura A.2.15 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=8 e FP =1 .....	79
Figura A.2.16 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=1 e FP =1 .....	80
Figura A.2.17 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=4 e FP =1 .....	80
Figura A.2.18 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=8 e FP =1 .....	81
Figura A.2.19 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=1 e FP =1 .....	82
Figura A.2.20 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=4 e FP =1 .....	82
Figura A.2.21 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=8 e FP =1 .....	83
Figura A.2.22 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 .....	84
Figura A.2.23 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 .....	84
Figura A.2.24 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 .....	84
Figura A.2.25 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 .....	85

Figura A.2.26 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 ..... 85

Figura A.2.27 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 ..... 86

Figura A.2.28 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 ..... 87

Figura A.2.29 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 ..... 87

Figura A.2.30 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 ..... 88

Figura A.2.31 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 . 89

Figura A.2.32 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 . 89

Figura A.2.33 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 . 90

Figura A.3.1 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=1 e FP =1 ..... 91

Figura A.3.2 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=4 e FP =1 ..... 91

Figura A.3.3 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=8 e FP =1 ..... 92

Figura A.3.4 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=1 e FP =1 ..... 93

Figura A.3.5 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=4 e FP =1 ..... 93

Figura A.3.6 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=1 e FP =1 ..... 94

Figura A.3.7 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=4 e FP =1 ..... 94

Figura A.3.8 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=8 e FP =1 ..... 95

Figura A.3.9 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=1 e FP =1 .....	96
Figura A.3.10 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=4 e FP =1 .....	96
Figura A.3.11 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=8 e FP =1 .....	96
Figura A.3.12 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=1 e FP =1 .....	97
Figura A.3.13 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=4 e FP =1 .....	97
Figura A.3.14 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=8 e FP =1 .....	98
Figura A.3.15 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=1 e FP =1 .....	99
Figura A.3.16 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=4 e FP =1 .....	99
Figura A.3.17 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=8 e FP =1 .....	100
Figura A.3.18 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=1 e FP =1 .....	101
Figura A.3.19 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=4 e FP =1 .....	101
Figura A.3.20 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=8 e FP =1 .....	102
Figura A.3.21 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 .....	103
Figura A.3.22 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 .....	103
Figura A.3.23 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 .....	104
Figura A.3.24 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 ...	105

Figura A.3.25 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 ... 105

Figura A.3.26 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 ... 106

Figura A.3.27 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1 ..... 107

Figura A.3.28 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1 ..... 107

Figura A.3.29 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1 ..... 108

Figura A.3.30 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1109

Figura A.3.31 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1109

Figura A.3.32 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1110



## **Índice de Tabelas**

Tabela 2.1 – Exemplo de uma matriz de custos para um problema com 3 classes .....	12
Tabela 2.2 – Exemplo de uma matriz confusão para um problema com 3 classes .....	13
Tabela 3.1 – Descrição sumária das bases de dados .....	29
Tabela 4.1 – Resultados obtidos utilizando a base de dados n° 2.....	34
Tabela 4.2 – Resultados obtidos utilizando a base de dados n° 2.....	35
Tabela 4.3 – Resultados obtidos utilizando a base de dados n° 3.....	36



## Lista de Abreviaturas

RD	Retinopatia Diabética
DR	<i>Diabetic Retinopathy</i>
AUC	<i>Area Under ROC Curve</i>
ROC	<i>Receiver Operating Characteristic</i>
MA	Microaneurisma
ISC	<i>Image Structure Clustering</i>
FOV	<i>Field Of View</i>
FN	Falso Negativo
FP	Falso Positivo
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
NPDR	<i>Non Progressive Diabetic Retinopathy</i>
PDR	<i>Progressive Diabetic Retinopathy</i>



# 1 Introdução

A Retinopatia Diabética (RD) é uma patologia ocular causada pela Diabetes sendo uma das principais causas de cegueira na população trabalhadora na União Europeia e nos Estados Unidos [2].

Nos primeiros estágios da doença não há manifestação de sintomas, havendo portanto a necessidade de se utilizar sistemas estruturados de rastreio que facilitem o diagnóstico precoce. Estes sistemas de rastreio fazem uso da imagem médica, nomeadamente a Retinografia (na qual se fotografa o fundo do olho), que pode ser usada para o diagnóstico da RD bem como da Degeneração Macular relacionada com o avanço da idade. Estas imagens são normalmente arquivadas e transmitidas para serem posteriormente classificadas manualmente por especialistas, geralmente oftalmologistas ou optometristas, o que torna todo este processo dispendioso e demorado.

## 1.1 Motivação

Diversos estudos [1][2] demonstraram ser possível detetar automaticamente os primeiros sinais de RD com base na deteção de lesões bem conhecidas e caracterizadas, visíveis nas imagens do fundo do olho. O uso de sistemas de avaliação automática de retinografias pode assim reduzir a sobrecarga laboral, aumentando a relação custo-eficácia de todo o processo.

Hoje em dia estão em desenvolvimento aplicações informáticas que visam implementar sistemas automáticos, como é o caso da família de produtos Retmarker, constituindo uma ferramenta de análise de imagem com capacidades de auxílio ao diagnóstico. No entanto, a deteção de lesões características de Retinopatia Diabética não é suficiente para permitir a automatização de todo o processo de classificação da imagem. Para tal, é necessário, no mínimo, garantir-se que a imagem tem uma clareza adequada e que inclui a porção correta da retina, isto é, possui uma qualidade adequada, como exemplificado na Figura 1.1.1. Imagens desfocadas, por exemplo, podem disfarçar lesões, fazendo com que um olho com patologia seja erroneamente classificado como normal. Para assegurar um diagnóstico fiável, idealmente dever-se-á ter em consideração a qualidade da imagem. Imagens de fraca qualidade, das quais é

exemplo a Figura 1.1.2, não deverão ser usadas para avaliação de diagnóstico, sendo necessária a reaqisição das mesmas, o que representa um problema, dado que geralmente o local e momento de aquisição é diferente daquele em que o especialista faz a avaliação da imagem, o que torna esta reaqisição um processo economicamente dispendioso e lento. É portanto muito vantajoso utilizar a avaliação da qualidade de imagem como parâmetro na classificação automática da doença.

As imagens que mostrem sinais de retinopatia ou possuam uma qualidade inadequada, insuficiente para serem classificadas como saudáveis, idealmente, deverão ambas ser transmitidas pelos sistemas de classificação automática para os especialistas de classificação manual. Ou seja, apenas as imagens com qualidade suficiente e em que não são detetáveis lesões dispensam atenção urgente de um especialista humano.

O que se pretende com este projeto é, através da fusão dos classificadores automáticos de RD com classificadores automáticos de avaliação de Qualidade melhorar a classificação automática de modo que tenha diretamente em consideração a qualidade de imagem, por forma a aumentar o desempenho destes métodos na deteção de RD.

Isto implica que todo o processo de aquisição e classificação da imagem pelos sistemas de classificação automática se torne bastante mais vantajoso, aumentando a eficácia destes sistemas e otimizando a automatização de todo o processo de classificação de imagem.



Figura 1.1.1 – Exemplo de imagem com qualidade adequada

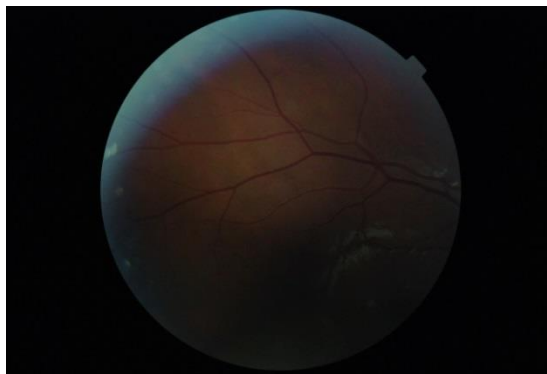


Figura 1.1.2 – Exemplo de imagem com qualidade inadequada

## 1.2 Estado da arte

Em 2004, Usher et al. [3] desenvolveram e testaram um sistema capaz de reconhecer características de RD usando imagens do fundo do olho centradas na mácula, adquiridas com um campo de 45°. Imagens de 1273 pacientes foram submetidas a um pré-processamento para otimização da cor e contraste destas. Recorreram à segmentação para deteção de lesões, sendo estas posteriormente classificadas de forma automática através de uma rede neuronal artificial. Foi conseguida uma sensibilidade de 95,1% para uma especificidade de 46,3%, sendo que ao otimizaram a especificidade para 78,9% a sensibilidade baixou para 70,8%.

Fleming et al. [4], em 2006, testaram vários métodos de normalização de contraste, mostrando que este processo potencia um melhor desempenho dos sistemas de deteção automática de microaneurismas (MA's). Conseguindo, para o melhor método, uma sensibilidade de deteção de imagens com MA's de 85,4% para um especificidade de 83,1%.

Mais tarde, em 2007, Philip et al. [5] treinaram classificadores automáticos de deteção de RD a partir de uma base de dados de 1067 imagens obtidas a partir de um programa de rastreio de RD regional. Foram testados usando uma diferente base de dados com 14 406 imagens de 6722 pacientes. Tanto a classificação automática, como a manual, foram feitas de acordo com a escala "Disease"/"No disease", sendo que a classificação "No disease" foi utilizada não só para marcar casos sem RD, bem como para a marcação de imagens de qualidade inadequada. A deteção dos casos marcados como "No Disease", para a classificação manual, obteve uma sensibilidade de 86,5% e especificidade de 95,3%, contra 90,5% de sensibilidade e 67,4% de especificidade obtidas pelo sistema automático.

Em 2009, Niemeijer et al. [6] estudaram diferentes métodos de fusão de classificadores para combinar algoritmos de deteção de diferentes lesões, sendo avaliados numa base de dados de 60 000 imagens. O melhor método produziu uma AUC de 0.881.

Abràmoff et al. [7] conduziram em 2010 um estudo no qual foi feita a comparação entre o algoritmo de deteção automática de RD proposto em 2008, por Quellec et al.[8], o qual ganhou a *Retinopathy Online Challenge Competition* em 2009 (denominado por "Challenge2009") e o algoritmo então usando na ferramenta de rastreio *EyeCheck*, publicado por Niemeijer et al. [6] em 2009 (denominado por "*EyeCheck*"). Os resultados obtidos foram similares para os dois algoritmos. Imagens contendo mais do que RD leve, foram detetadas com uma AUC de 0,839 para o algoritmo denominado "*EyeCheck*" e com uma AUC – *Area under*

*ROC (Receiver Operating Characteristic) curve* (forma de representar a relação entre a sensibilidade e especificidade – reflete a performance de um sistema classificador binário), de 0,821 para o algoritmo “*Challenge2009*”. Para uma sensibilidade de 90% foi conseguida uma especificidade de 47,7% e 43,6% para cada algoritmo respetivamente.

Mais tarde, em 2011, Niemeijer et al. [9] aplicaram também métodos de fusão de classificadores combinando diferentes detetores automáticos de microaneurismas. Contudo os resultados mostraram que a combinação de diferentes detetores para a mesma lesão não reflete necessariamente uma melhoria no desempenho.

Em 2011, Oliveira et al. [2] realizou um estudo no qual imagens de 5386 pacientes, adquiridas em 2007, foram avaliadas automaticamente para a existência de RD pelo sistema RetmarkerSR, gerando um *output* binário, “*Disease*” ou “*No Disease*”. Os resultados obtidos foram de 96,1% de sensibilidade para uma especificidade de 51,7%. Efetuou-se uma segunda análise na qual se avaliou, a par das imagens previamente adquiridas, as imagens provenientes de uma segunda aquisição realizada em 2008, para 289 pacientes. Esta resultou numa diminuição em 26,8% dos FP’s aumentando a especificidade para 63,2%, com uma sensibilidade de 95,8%.

No ano a seguir, Jelinek et al. [10] avaliaram diferentes técnicas de fusão para combinar diferentes classificadores automáticos de diferentes lesões de RD. Mostrando vantagens na combinação de detetores de lesões complementares.

No que concerne à correta avaliação da qualidade de imagem da retina, existem alguns parâmetros considerados de extrema importância. Segundo o estudo “*Atherosclerotic Risk in Communities (ARIC)*” levado a cabo pela *University of Wisconsin-Madison* [11], estes **parâmetros de qualidade de imagem** são a focagem e clareza de imagem, a definição do campo (relacionada com a porção da retina mostrada na imagem), visibilidade da mácula e do disco ótico, e inexistência artefactos. Os artefactos são perturbações na imagem que, segundo este mesmo estudo, são causados maioritariamente por existência de névoas, presença de sujidade, oclusão parcial provocada pelas pestanas, oclusão total devido ao piscar de olho, iluminação não homogénea, e limpeza insuficiente da lente da câmara.

Estes parâmetros podem ser divididos em duas categorias: os **parâmetros de qualidade de imagem genéricos**, que englobam a focagem e clareza de imagem e a ausência de artefactos; e os **parâmetros de qualidade de imagem estruturais**, como é o caso do campo da imagem e da visibilidade da mácula e do disco ótico.

Fazendo uso apenas de parâmetros genéricos, a classificação da qualidade de imagem é feita a partir de uma estimativa que tem em conta as medidas obtidas, não se recorrendo a métodos de segmentação estrutural, que são *per si*, mais complexos e demorados. A grande vantagem da utilização de métodos de avaliação de qualidade de imagem baseados em parâmetros genéricos é portanto a sua simplicidade face aos outros métodos. Usando apenas os critérios genéricos vários estudos foram publicados, nomeadamente o de Lee e Wang [12], de 1999, no qual foi proposto um método baseado numa abordagem por análise de histogramas. Neste foi calculado um índice por convolução entre o histograma de intensidades da imagem de entrada e o histograma de intensidades modelo obtido a partir de imagens da retina com uma boa qualidade de imagem. Em 2001, Lalonde et al. [13] avaliaram o gradiente de magnitude do histograma de uma imagem de entrada e a informação local do histograma dos seus valores de cinzento, comparativamente com histogramas de referência, obtidos a partir de imagens com boa qualidade. Bartling et al. [14] em 2009, desenvolveram um trabalho focado na avaliação da nitidez e iluminação da imagem. Repartiram a imagem em regiões quadradas não sobrepostas e analisaram-nas individualmente, obtendo indicadores individuais de qualidade para cada uma delas, posteriormente agrupados para formar um indicador geral da qualidade de imagem. Neste trabalho, para quantificar a iluminação, recorreu-se à avaliação do contraste e brilho da imagem da retina. Os indicadores de qualidade foram testados num conjunto de 1000 imagens, e os resultados foram concordantes com aqueles obtidos pelos especialistas em classificação manual.

Davis et al. [15], também em 2009, desenvolveu um método de avaliação de qualidade de imagem da retina, baseado no cálculo de 17 características diferentes para cada canal do espaço *CIE Lab*, um espaço de cor uniforme que na prática se trata de um modelo matemático para descrever cada cor. Este modelo colorimétrico é definido por L – luminosidade e a, b – gamas de cor. Os autores, fazendo uso destas características, puderam avaliar as gamas de cor da imagem, a sua luminosidade e o seu contraste. Os resultados obtidos sobre 5 conjuntos de 200 imagens cada foram uma sensibilidade de 100% e uma especificidade de 96% no que respeita à identificação de imagens de qualidade não adequada.

O primeiro trabalho proposto no qual o método de avaliação da qualidade de imagem se baseou nos parâmetros estruturais foi publicado por Usher et al. [16] em 2003. Neste estudo os autores usaram como índice de clareza da imagem a área dos vasos detectados sobre toda a imagem. Se a área medida estivesse acima de um determinado *threshold* por eles definido, a imagem era classificada como tendo uma qualidade de imagem adequada, pois quando uma imagem está turva ou desfocada, alguns dos vasos presentes deixam de ser detetáveis. Como tal,

será de esperar que a área dos vasos detetados seja reduzida em imagens não adequadas. Este método atingiu uma sensibilidade de 84,3% e uma especificidade de 95%. Tem que se ter em conta que os grandes vasos continuam a ser detetáveis, ou seja, que a imagem perde apenas uma percentagem dos seus vasos detetáveis quando sujeita a uma desfocagem mínima mas significativa. Tem ainda que se ter em consideração a variação anatómica na área vascular entre diferentes pacientes.

Em 2005 Lowell et al. [17] propuseram um algoritmo baseado em critérios estruturais, analisando também a vasculatura presente na imagem, mas considerando apenas uma área circular em torno da mácula. Em 2006 foi proposto o *Image Structure Clustering*, ISC, por Niemeijer et al. [18]. Este método é de forma geral para conseguir compreender as estruturas presentes em conjuntos de imagens. A estrutura local em cada pixel pode ser descrita através das saídas obtidas a partir de um conjunto de filtros. Este conjunto de filtros gera vetores de resposta que caracterizam estruturas particulares da imagem. O algoritmo proposto agrupa estes vetores de resposta em grupos que representam estruturas da imagem similares, os chamados *clusters*. Portanto, primeiro, cada imagem é filtrada usando um conjunto de filtros gerando vetores de resposta para todos os pixels da imagem, de seguida a distância entre o vetor de resposta e cada *cluster* é medida e o pixel é destacado para o *cluster* com a média mais próxima. O número de pixels destacados para cada *cluster* e os seus rácios relativos oferecem assim uma descrição compacta das estruturas presentes na imagem. Baseado nos *clusters* obtidos é criado um vector de característica que é classificado resultando numa medida de qualidade de imagem. Este método foi testado num total de 2000 e resultou numa AUC de 0,9968 o que é muito perto da área ótima, 1.

Também em 2006, Fleming et al. [19] analisou a vasculatura detetável sobre uma região centrada na fóvea para avaliar a clareza da imagem, usando a visibilidade dos vasos como indicador de clareza, distanciando-se do método mais geral que avalia a clareza da imagem a partir do contraste dos vasos relativamente ao fundo, comparando posteriormente com um *threshold*, podendo ser a escolha deste um pouco difícil. Apesar das semelhanças com o trabalho publicado por Usher et al.[16], para avaliação da clareza da imagem foram apenas considerados os vasos detetados mais estreitos, aumentando assim a capacidade de discriminação. Adicionalmente, avaliou-se a definição do campo, visto que até então não existia nenhum estudo no qual esta característica fosse automaticamente avaliada. Segundo os autores, cerca de 50% das imagens com qualidade não adequada falha por ter uma definição de campo inadequada apesar de possuir uma clareza suficiente. Assim, eles avaliaram esta característica recorrendo à verificação de um conjunto de condições relativas a medidas como o diâmetro do

disco ótico, a distância entre o disco ótico e a fóvea, o comprimento das arcadas e o alinhamento vertical da imagem. Foi obtida uma sensibilidade de 99,1% e especificidade de 89,4% para a detecção de imagens com qualidade inadequada.

Em 2008, Giancardo et al. [20] propôs um método que apesar de inspirado no trabalho de Niemeijer et al. [18] apenas tem em conta a vasculatura do olho e não toda a estrutura ocular. Foi neste estudo utilizada a densidade dos vasos medida em diferentes regiões e um histograma de 5 intervalos de cada canal de cor como características para avaliação.

Hunter et al. [21] propôs em 2011 um novo método de avaliação da qualidade da imagem da retina que obteve uma sensibilidade de 100% e especificidade de 93%, para um conjunto de 200 imagens. Este método foi baseado no contraste e quantificação de vasos sanguíneos dentro de um diâmetro de disco ótico da fóvea e no contraste entre a fóvea e o fundo.

Os dois tipos de parâmetros utilizados na avaliação da qualidade de imagens da retina, os parâmetros genéricos e os parâmetros estruturais foram pela primeira vez combinados por Paulus et al. [22] em 2011. Os critérios usados foram baseados na identificabilidade e diferenciação de certas estruturas no fundo do olho, bem como a homogeneidade da iluminação e a nitidez. O método apresentado baseou-se nestes critérios usando uma técnica de *clustering* similar ao ISC proposto por Niemeijer et al. [18], usando uma medida de nitidez e as características de Haralick. A técnica de *clustering* utilizada serviu para detetar certas estruturas e para avaliar como elas podem ser reconhecidas e diferenciadas. A medida de nitidez avalia o quão claramente os componentes estão separados entre si e as características de Haralick funcionam como indicadores dos seguintes parâmetros genéricos de avaliação: nitidez, contraste e homogeneidade. Este método atingiu uma sensibilidade de 96,9% e uma especificidade de 80% no reconhecimento de imagens de qualidade inadequada. Para comparação Paulus et al. [22] testou o método ISC no mesmo conjunto de imagens e obteve uma sensibilidade de 50,8% e 96,6% de especificidade, tendo o método proposto por Paulus et al. [22] obtido um melhor desempenho.

Já em 2012, Fleming et al. [23], descreveu um novo método para avaliar automaticamente a clareza de imagens da retina. Ao contrário do método acima referido publicado em 2006, este novo método não depende da presença da mácula na imagem. Neste trabalho foi efetuada uma comparação entre medidas estruturais e estatísticas, demonstrando-se que a avaliação da clareza pode ser restrita a certas regiões selecionadas da imagem e que o desempenho é mantido independentemente do FOV (*field of view*). No entanto, o desempenho deste novo método foi similar ao do método apresentado pela equipa em 2006. Concluiu-se

então que a avaliação da clareza baseada em medidas estatísticas possui um desempenho igualmente bom ao da avaliação da clareza baseada na análise da vasculatura.

Também em 2012, Dias et al. [24] propôs um algoritmo de avaliação da qualidade da imagem baseado em medidas de parâmetros genéricos, como a cor, a focagem, o contraste e a iluminação, cada um analisado por um algoritmo próprio. Até então, não tinha sido feita nenhuma abordagem como esta, em que os quatro parâmetros são avaliados distintamente, combinando-se posteriormente para se conseguir uma classificação geral da imagem. Segundo os autores, esta classificação da qualidade de imagem obtida a partir de indicadores individuais é bastante importante pois produz indicadores de qualidade de imagem parciais, o que pode ser útil para uma avaliação de qualidade em tempo real ou para apontar correções durante o processo de captura. Os algoritmos propostos para avaliar os quatro parâmetros e os respectivos quatro classificadores mostraram um bom desempenho, perto do ideal, com as seguintes áreas por baixo da curva ROC:  $AUC_{\text{cor}}=0,9993$ ,  $AUC_{\text{focagem}}=0,9867$ ,  $AUC_{\text{contraste}}=0,9783$ ,  $AUC_{\text{iluminação}}=0,9984$ . Baseando-se nas avaliações e classificações obtidas por cada um deles, um classificador final permitiu a avaliação da qualidade geral da imagem, atingindo uma sensibilidade de 99,76%, especificidade de 99,49% e área sob a curva ROC de 0,9987, para um conjunto de 2032 imagens. Comparativamente com os trabalhos desenvolvidos por Davis et al. [15] e Niemeijer et al. [18], este método possui um melhor desempenho no sentido em que possui uma maior uniformidade de resultados e uma maior área sob a curva ROC sendo também consideravelmente mais rápido. Assim, este trabalho veio mostrar que características genéricas são boas bases para o desenvolvimento de algoritmos de avaliação de qualidade de imagens da retina, na medida em que permitem obter classificações confiáveis.

Apesar de alguns dos estudos de desempenho dos sistemas de deteção automática de RD terem em consideração indiretamente a qualidade da imagem, recorrendo a algoritmos de pré-processamento para otimização de algumas das características desta, nenhum faz uso direto da informação proveniente da mesma como parâmetro de classificação. De tal forma que, com o objetivo de melhorar desempenho dos algoritmos de classificação automática na deteção de RD, desenvolvidos pela Retmarker S.A. ir-se-á proceder à combinação destes classificadores com os algoritmos de classificação dos parâmetros genéricos, propostos por Dias et al. [24], conjuntamente (Cor, Focagem, Contraste e Iluminação) e com a Qualidade Geral. A principal técnica de combinação de classificadores é a fusão, na qual é considerado que todos os classificadores contribuem para a decisão final [10]. Pretende-se aqui perceber qual das combinações produz resultados mais promissores no âmbito do rastreio automático de RD.

Numa segunda abordagem, estudar-se-á a performance dos classificadores automáticos de RD, desconsiderando as imagens classificadas com 0 (má qualidade) pelo algoritmo de classificação de Qualidade Geral. Reproduzindo a situação em que uma imagem é adquirida, é avaliada a sua qualidade e, caso não tenha uma qualidade adequada, é solicitada a reaquisição da imagem, não servindo essa imagem de *input* aos algoritmos de classificação automática de RD.

De forma mais geral, apresenta-se em seguida o diagrama de blocos do projeto (Figura 1.2.1).

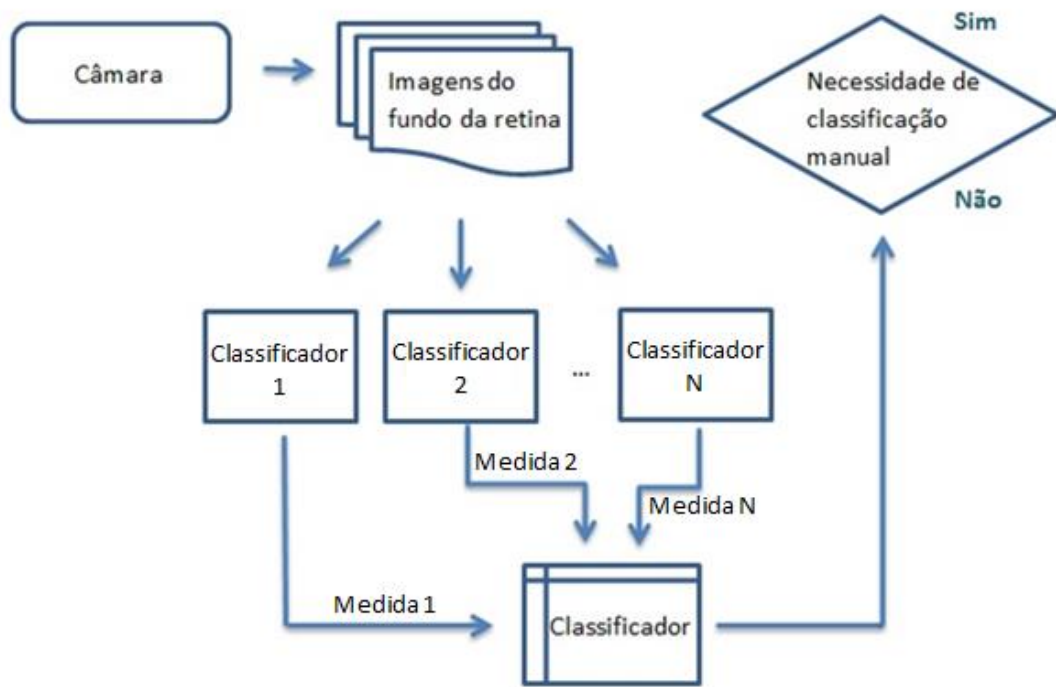


Figura 1.2.1 – Diagrama de blocos do projeto



## 2 Conceitos teóricos

### 2.1 Classificador automático

Um classificador automático [25] é um sistema capaz de avaliar as características de um determinado objeto ou situação e classificá-los dentro de um conjunto limitado de classes. Este é construído a partir de uma aprendizagem supervisionada utilizando um conjunto pré-classificado de dados, chamado conjunto de treino. Uma vez construído, é testado num outro conjunto de dados, o conjunto de validação, do qual não deve fazer parte dados que tenham sido utilizados no treino do classificador, de maneira a que os resultados sejam íntegros.

Um classificador pode ser um conjunto de regras de classificação, uma árvore de decisão ou uma rede neuronal.

Entre as diferentes técnicas de prospeção de dados usadas para classificação, as mais utilizadas são, as árvores de decisão e as regras de classificação.

As principais vantagens das árvores de decisão, prendem-se com o facto de estas possuírem estruturas simples e perceptíveis. São facilmente aplicáveis a grandes conjuntos de dados, e a qualquer tipo de dados, são capazes de trabalhar com atributos contínuos e discretos. Porém quando utilizadas grandes bases de dados, deixam de ser tão facilmente legíveis pelo usuário. As regras de classificação não possuem esta desvantagem, na medida em que podem ser compreendidas per si, isto é, sem necessidade de referenciar outras regras. Contudo, não são aconselhadas para trabalhar com atributos numéricos.

As regras de classificação podem ser extraídas a partir de árvores de decisão, sendo a eficiência da classificação mantida quando se dá esta conversão.

### 2.2 Qualidade de um Classificador

A taxa de erro de um classificador é um indicador da qualidade deste. Obtida através do teste do mesmo, pode ser estimada de diversas maneiras [25], sendo que umas produzem resultados mais realistas que outras.

O erro de um classificador é dado pelo quociente entre o número de erros e o número de casos testados:

$$\text{Percentagem de erro (\%)} = \frac{N^{\circ} \text{ de erros}}{N^{\circ} \text{ de casos testados}}$$

- **Estimação por Re-substituição ou Erro Aparente**
  - O classificador é testado pelo conjunto de treino;
  - Sendo o conjunto de treino utilizado para testar a validade do classificador, claramente é uma estimativa demasiado otimista, não é uma boa estimativa;
- **Estimação por utilização de um conjunto independente**
  - O conjunto de dados disponíveis é dividido no conjunto de treino e no conjunto teste;
  - Ao reduzir o número de dados do conjunto de treino, a estimativa torna-se um pouco pessimista;
- **Estimação por custos**
  - O custo de um erro é a penalização imposta ao sistema caso ele cometa algum erro de classificação;
  - Com o objetivo de minimizar o custo de erros de classificação, é utilizada uma **matriz de custos** (Tabela 2.1) que contém o custo de cada erro;

Tabela 2.1 – Exemplo de uma matriz de custos para um problema com 3 classes

Classe atribuída	Classe real		
	I	II	III
I	0	2	12
II	10	0	8
III	5	6	0

- É também usada uma **matriz confusão** (Tabela 2.2), que representa o número de exemplos por classe obtida como resultado do teste do classificador;

Tabela 2.2 – Exemplo de uma matriz confusão para um problema com 3 classes

Classe atribuída	Classe real		
	I	II	III
I	92	6	2
II	5	89	6
III	10	15	75

- O custo do classificador é dado por:

$$Custo = \sum_{i=1}^N \sum_{j=1}^N C_{i,j} M_{i,j}$$

Onde:

- $C_{i,j}$  é o valor da linha  $i$  e coluna  $j$  da matriz de custos
- $M_{i,j}$  é o valor da linha  $i$  e coluna  $j$  da matriz confusão
- $N$  é o número de classes

Assim, para o exemplo das Tabelas 2.1 e 2.2, temos:

$$Custo = 0*92+2*6+12*2+10*5+0*89+8*6+5*10+6*15+0*75=274$$

Desta forma o custo médio será:

$$Custo\ médio = \frac{Custo}{Número\ de\ Classificações}$$

Para o exemplo anterior:

$$Custo\ médio = \frac{274}{300} = 0.91$$

- A desvantagem deste método é que tem em conta apenas os custos que provêm dos erros, não considerando os ganho através das classificações corretas;
- **Estimação por camadas ou cruzada**
  - O conjunto de dados é dividido em  $N$  subconjuntos com aproximadamente igual número de exemplos;
  - São construídos  $N$  classificadores, cada um deles é treinado com  $N-1$  conjuntos e testado com o conjunto que fica de fora;
  - O resultado final é a média dos resultados;

- Tem que se ter em atenção que ao fazer-se a divisão dos conjuntos, eles devem possuir aproximadamente a mesma proporção de casos;
- Estimção complexa e ainda com alguma tendncia pessimista;
- Se o nmero de dados igual ao nmero de subconjuntos, ento estamos perante o mto *leave-one-out* – estimao mais realista;
- **Estimao por *bootstrapping e0***
  - Para formar o conjunto de treino feita uma amostragem aleatria de N exemplos, com reposio, do conjunto original de tamanho N, e so retirados os exemplos repetidos, ficando um conjunto, em mdia, com cerca de 62.3% dos exemplos do conjunto original;
  - Depois de construdo o classificador com base neste conjunto de treino, testado com o conjunto de teste, formado pelos exemplos que ficam de fora;
  - Tendncia pessimista:
- **Estimao por *bootstrapping .623***
  - Na estimao por *bootstrapping .623* em cada iterao calculado o erro, fazendo a combinao linear do erro de Re-substituio ponderado de um facto de 0.623 com o erro *e0* ponderado de um facto de 0.377 – Tendncia menos pessimista;

A escolha do melhor mto de validao depende do tamanho do conjunto de treino [26]. Para conjuntos de treino grandes, so recomendados os mtodos de validao por camadas, ou de diversas camadas ou o *leave-one-out*; para conjuntos de treino pequenos recomendado o *leave-one-out*; para conjuntos de treino muito pequenos dever ser calculado o *leave-one-out*, *bootstrapping .623* e validao por camadas utilizando duas camadas estratificadas, seguidamente a heurstica abaixo dever ser aplicada:

- $leave-one-out < bootstrapping .623 \Rightarrow$  usa-se *bootstrapping .623*
- $leave-one-out > estimao por camadas \Rightarrow$  usa-se estimao por camadas
- Caso contrrio usa-se *leave-one-out*

## 2.3 Indicadores de Performance

Com intuito de encontrar o método que melhor se ajusta ao problema e produz assim os melhores resultados podem ser avaliados alguns indicadores de performance, entre os quais:

- **CA (*Classification Accuracy*)**

A *Classification Accuracy*, CA, é uma medida de precisão da classificação, calculada a partir do número total de classificações corretas sobre o número total de classificações:

$$CA = \frac{\text{Número de classificações correctas}}{\text{Número total de classificações}} \times 100 (\%)$$

- **Sensibilidade**

A sensibilidade mede a capacidade do sistema identificar corretamente os casos que possuem doença:

$$\text{Sensibilidade} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}} \times 100 (\%)$$

- **Especificidade**

A especificidade mede a capacidade do sistema identificar corretamente os casos que não possuem doença:

$$\text{Especificidade} = \frac{\text{Verdadeiros Negativos}}{\text{Verdadeiros Negativos} + \text{Falsos Positivos}} \times 100 (\%)$$

- **AUC (*Area under ROC curve*)**

A **curva ROC** (*Receiver Operating Characteristic*) ilustra a performance de um sistema de classificação binário, e traduz a relação, normalmente antagónica, entre a sensibilidade e especificidade do sistema. A curva ROC é obtida traçando um diagrama que representa a sensibilidade em função dos Falsos Positivos, isto é,  $(1 - \text{Especificidade})$  para determinados **valores de corte**.

## 2.4 Árvores de decisão

### 2.4.1 Introdução

Uma árvore de decisão é um modelo, cuja estrutura se assemelha ao formato de uma árvore. É constituída por um número finito de nós e folhas, onde cada nó representa uma decisão entre um número finito de alternativas, isto é, um teste sobre um atributo, e cada folha, ou *nó terminal*, representa uma classificação, obtendo-se assim uma regra por cada folha sendo possível extrair conteúdo informativo de cada atributo [27]. Cada árvore possui um nó localizado no topo, denominado *nó raiz*, que é o pai de todas as sub-árvores cujos os nós denominam-se nós filhos do nó raiz. Os nós que não possuem nós filhos são então, os nós terminais [28].

O *nível* de um nó diz respeito à distância, isto é, número de ligações, entre este e o nó raiz. A definição de *profundidade* de uma árvore de decisão é a maior distância entre um nó terminal, folha, e o nó raiz. A árvore pode ser *binária*, *ternária* ou *mista*, dependendo se possui duas, três ou número variável de ligações em cada nodo, respetivamente. Em suma, os **nós** de uma árvore de decisão correspondem aos **atributos**, as **ramificações** de cada nó representam os **valores deste atributo**, e as **folhas** representam as **classes** [28]. A classificação de determinado objeto é feita testando os valores dos seus atributos nos nós ao longo da árvore, formando um percurso desde o nó raiz, até um dos nós terminais, o qual representará a classe predita para aquele objeto [27]. Significa que, após a construção de uma árvore de decisão, a mesma pode ser utilizada na classificação de novos objetos, efetuando o percurso de forma *top-down*, através dos nós de decisão, desde o nó raiz até ao nó terminal, classe.

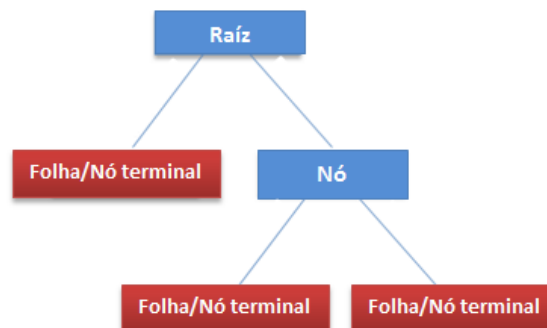


Figura 2.4.1 – Exemplo de árvore de decisão

Algumas das características das árvores de decisão, referidas por J. Gama [29]:

- *Flexibilidade*: não assumem uma distribuição única dos dados;
- *Robustez*: tendência na construção de árvores robustas, ainda que possuindo variáveis não relevantes, devido a seleção interna dos atributos/testes;
- *Interpretabilidade*: produzem resultados inteligíveis, usando valores conhecidos dos atributos para interpretar o problema;
- *Velocidade*: a maioria dos algoritmos de indução de árvores de decisão, constrói a árvore de forma *top-down*, dividindo-se pelos diversos percursos, sem necessidade de retornar ao início para percorrer outro percurso;

Das desvantagens mais comuns, têm-se [29]:

- *Fragmentação*: induz a partição dos dados em subconjuntos, o que não é ideal quando se está a fazer uso de vários atributos com relevância;
- *Replicação*: pode ocorrer a duplicação de uma sequência de testes em diferentes ligações, diminuindo a precisão da classificação;
- *Valores em falta*: dado que uma árvore de decisão trata-se de uma hierarquia de testes, se o valor de um atributo é desconhecido, poderá resultar numa indecisão no percurso a seguir;
- *Valores contínuos*: dificuldade no tratamento destes valores, o que implica um maior tempo de processamento;
- *Instabilidade*: pequenas variações no conjunto de treino, podem produzir grandes variações na árvore de decisão final;

### 2.4.2 Indução de árvores de decisão

O processo de construção de uma árvore de decisão é denominado, **Indução**. A maioria dos algoritmos de indução de árvores de decisão efetua a construção destas de forma *top-down*, isto é, de cima para baixo, ou seja, do nó a raiz até aos nós terminais. A partir do conjunto de treino, cada iteração, os algoritmos de indução procuram o melhor atributo/teste, que é aquele que permite uma melhor separação das classes, para realizarem a ramificação da árvore.

A estratégia *top-down* é dividida em duas abordagens [28]:

- *Por profundidade (depth-first)*: dado um determinado nó inicial é aplicada uma ligação para atingir um novo nó; se este não for um nó terminal, então é adicionada uma nova ligação e por conseguinte um novo nó, e assim sucessivamente. Quando se chegar a um nó terminal, então dá-se um recuo ao nó pai deste, e adiciona-se uma nova ligação para um novo nó. Se não for possível adicionar um novo nó, dá-se um novo recuo até ao nó que permite incluir uma nova ligação. O processo termina quando não se consegue encontrar mais nenhum nó. Exemplo descrito na Figura 2.4.2:

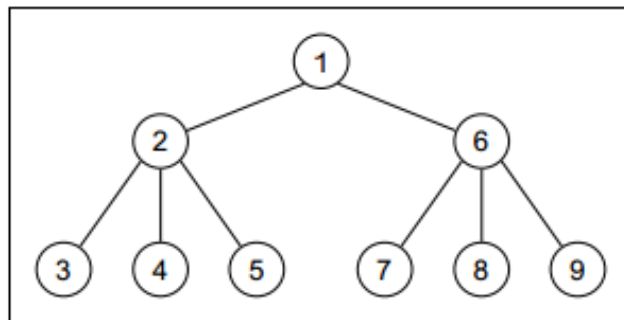


Figura 2.4.2 – Indução por profundidade [28]

- *Por largura (breadth-first)*: são aplicadas todas as ligações de um nó, atingindo novos nós. Para cada nó resultante é testado o critério de paragem, resultando na adição de um nó terminal ou de uma nova ligação e por conseguinte, um novo nó. Esta técnica não utiliza *backtracking*, isto é, percorre um caminho sem necessidade de retornar ao início para percorrer um novo caminho. Exemplo descrito na Figura 2.4.3:

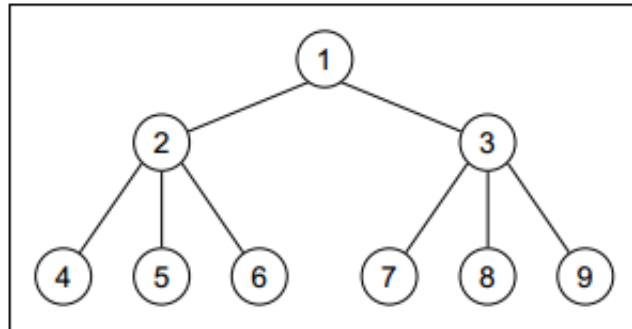


Figura 2.4.3 – Indução por largura [28]

A indução por profundidade nem sempre produz uma solução ótima, porém não necessita de muitos requisitos de memória. A indução por largura produz uma solução mais completa, ótima, no entanto é uma estratégia mais pesada computacionalmente.

### Escolha do nó

Para eleger o melhor atributo/teste pode ser usado um dos seguintes critérios [28] [30]:

- **Critério de Gini**

A impuridade de um nó é máxima quando todas as classes possuem igual distribuição, e mínima quando existe apenas uma classe. Este critério, tem assim como objetivo, minimizar a impuridade de cada nó.

$$i(t) = 1 - \sum_{classes} (p(classe|t))^2 = 1 - \sum_{i=1}^N p_i^2$$

Onde:

- $i(t)$  é a impuridade
- $p_i$  é a probabilidade da classe  $i$
- $N$  é o número de classes

- **Cr terio da Entropia**

Este crit rio d -nos uma medida da quantidade de informa  o necess ria para codificar a classe de um determinado n . O atributo que possuir o maior ganho de informa  o,   o selecionado para o n  em quest o.

$$i(t) = - \sum_{i=1}^N p_i \times \log_2 p_i$$

- **Cr terio da Paridade**

D -nos diretamente o ganho da impuridade,

$$\Delta i(t) = \frac{p_L p_R}{4} \left[ \sum_j |p(j|t_L) - p(j|t_R)| \right]^2$$

Onde:

- $p_L$    a probabilidade do n  descendente esquerdo
- $p_R$    a probabilidade do n  descendente direito
- $p(j|t_L)$    a probabilidade da classe  $j$  no n  descendente esquerdo
- $p(j|t_R)$    a probabilidade da classe  $j$  no n  descendente direito

Este crit rio j  tem inerente o equil brio entre os ramos e divis o das classes.

Fazendo uso deste m todo, o atributo escolhido ser  aquele que minimiza a impuridade, que implica o menor ganho de impuridade.

### Defini o da parti o

Para efetuar a parti o em cada n , quando estamos a lidar com **caracter sticas discretas**, pode utilizar-se uma das seguintes propostas [30]:

- **Um ramo por cada valor do atributo**
  - A cada valor do atributo   atribu do um ramo;
  - Extrai todo o conte do informativo do atributo;
  - Grande complexidade;
- **Solu o de Hunt**
  -   atribu do um ramo a cada valor do atributo, e um outro ramo aos restantes valores;
  - N o explora os atributos na sua totalidade;

- Tendência a criar árvores de decisão desequilibradas;
- **Características ordenadas**
  - Se for possível estabelecer uma relação de ordem entre os valores de um atributo, então este diz-se ordenado;
  - Os teste efetuados em cada nó são do tipo  $x < C \Rightarrow$  nós binários;
  - Simplicidade de perceção;
  - Não maximiza a informação extraída dos atributos;
- **Agrupamentos de características em dois conjuntos**
  - São criados e associados dois subconjuntos a cada nó binário, sendo que para um atributo com cardinalidade N existem  $2^{N-1} - 1$  partições possíveis;
  - Grande complexidade;
- **Agrupamentos de características em vários conjuntos**
  - Os valores de cada atributo são agrupados inicialmente em grupos de dois, testando-se as partições, seguidamente são agrupados em grupos de três, voltando a ser testados, e assim sucessivamente, até que não se observe nenhuma melhoria nos resultados;
  - Grande complexidade;

Para **características contínuas** podem ser usadas também várias abordagens para efetuar a partição em cada nó, sendo a mais simples a do **teste simples ou pesquisa exaustiva**, na qual são ordenados os valores do atributo e é escolhido um dos valores médios entre eles.

### Definição da folha/classe

A decisão de quando um nó é terminal, pode ser feita quando [31]:

- i. Todos os exemplos que atingem um nó pertencem à mesma classe;
- ii. Todos os exemplos que atingem o nó possuem os mesmos valores dos atributos, mas não pertencem necessariamente à mesma classe;
- iii. O número de exemplos de um nó é menor que um certo valor percentual;

- iv. Para todos os possíveis testes que dividem o conjunto de exemplos, o resultado do melhor atributo é muito baixo;

A definição de qual a classe a atribuir a um determinado nó terminal pode ser feita de duas maneiras [30]:

- i. *Atribuição da classe com maior probabilidade*: é atribuída a classe mais frequente dentro dos exemplos que se encontram no nó terminal:

$$\max(p_j) = \max_{j=1}^k \left( \frac{n_j}{n} \right)$$

Onde:

- $n$  é o número total de exemplos no nó terminal
- $n_j$  é o número da classe  $C_j$  no nó terminal
- $k$  é o número de classes

- ii. *Atribuição baseada nos custos*: é atribuída uma classe que minimiza os custos provenientes desta classificação:

$$\text{custo}(j) = \sum_{i=1}^k p_i C_{i,j}$$

Onde:

- $k$  é o número de classes
- $C_{i,j}$  é o valor da linha  $i$ , coluna  $j$  da matriz de custo
- $p_i$  é a probabilidade da classe  $C_i$

## Limites

Durante a construção de uma árvore de decisão ocorre o crescimento desta enquanto são encontradas novas partições, que melhoram o seu desempenho em separar os objetos do conjunto de treino nas diversas classes. Isto pode levar a um *overfitting*, que ocorre quando a árvore em vez de generalizar, aprende detalhadamente os padrões existentes. Isto é, a árvore gerada é muito específica para o conjunto de treino, mas falha na generalização do conhecimento do conjunto de teste.

Com vista a contornar este problema pode-se proceder a uma substituição dos nós profundos por nós terminais, removendo-se assim as ligações que não contribuem para a

precisão da classificação. Esta técnica é denominada por *poda* e realiza-se fazendo uma análise comparativa entre a taxa de erro do nó e a taxa de erro que ocorre quando se poda o mesmo [28]. A taxa de erro é dada por:

$$E(T) = \frac{ce}{ce + cc}$$

Onde:

- *ce* é o número de casos com classificação errada
- *cc* é o número de casos com a classificação correcta

A estratégia de poda pode induzir alguns erros na classificação do conjunto de treino, no entanto quando se faz a classificação de novos exemplos, não presentes no conjunto de treino, o classificador irá produzir um erro de generalização menor.

Os métodos de poda podem dividir-se em duas abordagens principais [28]:

- *Pré-poda*: quando os critérios de paragem aquando a indução da árvore são satisfeitos, parando assim a construção da árvore;
- *Pós-poda*: a poda da árvores ocorre depois da construção completa desta;

A pós-poda é a abordagem mais comum e confiável, mas trata-se de um processo mais demorado. A pré-poda tem a vantagem de não gastar tempo desnecessário na construção de uma árvore completa, que no final não irá ser toda utilizada.

Mesmo que nem sempre a poda de uma determinada árvore melhore a precisão desta em relação à sua original, sempre tem a vantagem em tornar a árvore numa estrutura menos complexa.

### 2.4.3 Algoritmo CART (*Classification and Regression Trees*)

Este algoritmo foi apresentado em 1984, por L. Breiman, J. Friedman, R. Olsen e C. Stone [32]. Realiza a indução das árvores de decisão de forma automática, a partir da estratégia *top-down*. Baseia-se num conjunto de treino, rotulado com o respetivo *Ground Truth*, e constrói a árvore de decisão realizando duas ligações por cada nó, em função de um atributo apenas. O atributo é escolhido como aquele que minimiza a diversidade dos grupos gerados, isto é, predominando uma única classe. [28]

Como resultado, este algoritmo gera um árvore de decisão binária univariada, bastante inteligível.

Os critérios utilizados por este algoritmo são [30]:

- i. A eleição do melhor atributo é efetuada principalmente através do critério de Gini, sendo possível efetuar-se a partir do critério de entropia ou de paridade;
- ii. Para atributos discretos, através da criação de nós binários, o algoritmo forma dois subconjuntos possíveis;
- iii. Para atributos contínuos, faz uso da combinação linear para repartir os nós
- iv. Para tratamento de dados desconhecidos, este algoritmo em vez de atribuir a cada nó apenas o atributo que minimiza a impureza, atribui-lhe um conjunto de atributos, em *ranking*, que produzem uma partição semelhante. Assim, quando está a ser classificado um novo objeto cujo valor do melhor atributo é desconhecido, é observado o valor do atributo de valor conhecido seguindo a ordem do *ranking*;
- v. A classe de cada nó terminal pode ser atribuída através da maior probabilidade ou da minimização de custos
- vi. Utiliza um método de pós-poda, por redução do fator custo-complexidade;

#### 2.4.4 Extração de Regras de Classificação

Como mencionado anteriormente, é possível representar o conteúdo informativo extraído a partir das árvores de decisão sob a forma de regras de classificação. O processo de indução é repetido para cada percurso possível a partir do *nó raiz*, até a um *nó terminal* da árvore de decisão, formando para cada um destes uma regra de classificação.

As regras de classificação são apresentadas sob a forma: SE <condição> ENTÃO <classe>, em que as condições são dadas através dos pares <atributo, valor do atributo> ao longo do percurso, e a classe é dada pelo *nó terminal*.

Após a inferência de uma regra, esta pode ser *podada*, removendo-se condições que não contribuem para a precisão desta. A regra é avaliada considerando a exclusão de determinada condição, caso a taxa de erro da regra diminua, esta condição será então removida. Avalia-se também a existência de regras similares, sendo as dualidades eliminadas. Sendo este

procedimento de remoção de condições, geralmente, computacionalmente caro. Para eliminar esta desvantagem, foram criados métodos diretos de indução de regras de classificação, sem ser necessária a construção de árvores de decisão [27].



## 3 Métodos

### 3.1 Base de dados

Foram disponibilizadas duas bases de dados pela Retmarker S.A., a partir do quais foram obtidos os conjuntos de treino para a aprendizagem supervisionada dos classificadores. A primeira base de dados foi cedida pelo TECSA - The Eye Field (East London – South Africa). Trata-se de uma base de dados obtida em contexto clínico, constituído por 382 imagens do fundo do olho, todas elas classificadas pelo mesmo oftalmologista, Dr. Steve Cook, para a existência de RD. A escala utilizada neste contexto foi a seguinte:

- R0 – No visible retinopathy
- R1 – Mild
- R2 – Observable background
- R3 – Referable background
- R4 – Proliferative
- R6 – Inadequate
- M0 – No visible maculopathy
- M1 – Observable
- M2 – Referable

Para análise de performance dos algoritmos de deteção automática de RD, foram tomadas as seguintes considerações:

- Todas as imagens classificadas como R0M0, R1M0 e R1M1 foram consideradas como não tendo doença – *NO DISEASE*. Sendo as restantes combinações consideradas como tendo doença – *DISEASE*, de maneira a obtermos assim, um output binário;
- As imagens onde foram observadas marcas de tratamento prévio (cicatrizes de laser) foram retiradas da análise. Isto porque, do ponto de vista de rastreio estas não deverão ser consideradas, dado que estas marcas indicam que o olho em questão já obteve tratamento para a patologia;

Desta forma, a base de dados final possui então 356 imagens estando todas elas classificadas binariamente como *DISEASE* ou *NO DISEASE*.

Para realizar a segunda abordagem, na qual não se pretende considerar os casos de má qualidade de imagem apontados por um algoritmo automático de validação de qualidade [24], foram removidos todos esses casos, resultando numa base de dados de 336 imagens.

A segunda base de dados utilizada no estudo foi cedida pelo Sankara Nethralaya Eye Hospital (Chennai, Índia), obtida em contexto de rastreio em 2004 . É constituída por 5850 imagens, pertencentes a 1519 pacientes. A classificação manual foi executada por técnicos do referido *Eye Hospital* e efetuada por paciente. Pacientes considerados com imagens com qualidade inadequada, não foram classificados para a existência de RD. Dos 1519, apenas 1469 foram classificados manualmente de acordo com a seguinte escala:

- No DR
- Mild NPDR
- Moderate NPDR
- Severe NPDR
- PDR

Dado que os algoritmos de deteção automática de RD fazem a avaliação de cada imagem individualmente, o facto da classificação manual ter sido por paciente e não por imagem, conduziu a uma necessidade de uma reavaliação desta, a fim de conseguir fazer corresponder a cada classificação a imagem mais ajustada. Considerou-se que a imagem mais ajustada à classificação manual do paciente é aquela que, em caso de existência de patologia, demonstre mais lesões. Desta forma passou-se a ter para avaliação uma imagem por paciente, tornando-se possível avaliar o desempenho dos algoritmos de classificação automática por si e efetuar a aprendizagem supervisionada dos algoritmos resultantes da posterior fusão.

Ao efetuar-se a análise detalhada da base de dados, de modo a realizar a correta correspondência entre as classificações e as imagens, observou-se que em alguns casos a classificação atribuída pelos técnicos não corresponde a nenhuma da imagem do respetivo paciente. Assim, e para tentar salvaguardar possíveis lapsos efetuados aquando a classificação manual feita pelos técnicos do Sankara Nethralaya Eye Hospital, foi executada em simultâneo, uma nova classificação.

Reparou-se também que para 790 pacientes todas as respetivas imagens foram adquiridas com o ângulo errado de acordo com o protocolo de rastreio, e possuem artefactos, possivelmente devido à sujidade da lente. Assim, as imagens destes pacientes tiveram que ser

desconsideradas da base de dados para realizar o estudo em questão. A base de dados a ser assim utilizada, contém um total de 729 imagens correspondendo a 729 pacientes.

Na classificação feita a par da original foi considerada a existência de mais dois pacientes com imagens de qualidade inadequada, sendo desta nova análise desconsiderados, resultando numa base de dados de 727 imagens, correspondentes a 727 pacientes.

Nesta base de dados não consta a informação sobre a existência de marcas de laser, como tal, nenhuma consideração foi tomada em relação à presença ou ausência destas marcas. Sendo que esta base de dados foi obtida num contexto de rastreio, após esta observação cuidada de todas as imagens, achou-se ajustado considerar todas aquelas com a classificação “No DR” como não tendo doença – *NO DISEASE* e as restantes como tendo doença – *DISEASE*, aproximando um pouco mais o rácio Doença/Não Doença ao da base de dados cedida pelo TECSA.

O resultado final são assim duas bases de dados possuindo então 729 imagens para a classificação efectuada pelo Sankar Nethralaya Eye Hospital e 727 imagens para a nova classificação, estando todas elas classificadas binariamente como *DISEASE* ou *NO DISEASE*.

Para efetuar a segunda abordagem, foram retiradas todas as imagens consideradas pelo algoritmo automático de classificação da qualidade como sendo inadequadas, resultando numa base de dados de 659 imagens para o classificação original, e 657 imagens para a nova classificação efetuada.

Em suma, as bases de dados:

Tabela 3.1 – Descrição sumária das bases de dados

Base de dados	Total de imagens	Sem má qualidade	Fonte	Classificação manual
1	356	336	TECSA	Dr. Steve Cook
2	729	659	Sankara Nethralaya - EH	Técnico especializado
3	727	657	Sankara Nethralaya - EH	Rita Figueiredo

## 3.2 Processamento

Todas as imagens foram submetidas a três algoritmos de processamento diferentes, dois de detecção/classificação automática de RD, os quais denominar-se-ão em diante por Retmarker e DMUC, e um de avaliação da qualidade da imagem

O algoritmo Retmarker diz respeito ao algoritmo de classificação automática de RD desenvolvido pela Retmarker S.A. para detecção de microaneurismas, e do qual podemos reproduzir um *output* binário: *DISEASE* ou *NO DISEASE*.

O algoritmo DMUC trata-se de um algoritmo de classificação automática de RD desenvolvido pelo Departamento de Matemática da Universidade de Coimbra em parceria com a Retmarker S.A., capaz de detetar microaneurismas, hemorragias e outras evidências de lesões, como exsudatos de proteínas, referidas como BL's (*Bright Lesions*). Este algoritmo permite também a extração de um *output* binário: *DISEASE* ou *NÃO DISEASE*.

O algoritmo de processamento da qualidade da imagem é constituído por 5 algoritmos de classificação: Cor, Focagem, Contraste, Iluminação e a Qualidade Geral da imagem, sendo este último uma fusão entre os outros 4 classificadores. Os 4 algoritmos de classificação das características da imagem são avaliadas numa escala de 0 a 3, sendo que 0 é a pior classificação e 3 a melhor, e a Qualidade Geral desta é reproduzida num *output* binário: 1 – qualidade adequada, 0 – qualidade inadequada.

## 3.3 Construção dos classificadores

### 3.3.1 Conjunto de treino

O conjunto de treino usado para a aprendizagem automática dos classificadores foi então obtido a partir do processamento da base de dados mencionada acima. Ir-se-á proceder a combinações entre os 7 classificadores disponíveis: Retmarker, DMUC e os algoritmos de classificação da qualidade (Qualidade Geral, Cor, Focagem, Contraste e Iluminação). Portanto, os classificadores construídos a partir destas bases de dados irão avaliar estes atributos, e atribuir a cada imagem a respetiva classe, baseada na decisão do classificador humano, como sugere a Figura 3.3.1.



Figura 3.3.1 – Diagrama da combinação de classificadores para a primeira abordagem

Seguidamente, proceder-se-á ao estudo do desempenho dos classificadores automáticos de RD, Retmarker e DMUC, desconsiderando as imagens classificadas como 0 (má qualidade) pelo algoritmo de classificação de Qualidade Geral. Desta forma teremos, no máximo, 6 atributos no conjunto de treino, dado que o resultado da Qualidade Geral será sempre 1. Figura 3.3.2.



Figura 3.3.2 – Diagrama da combinação de classificadores para segunda abordagem

### 3.3.2 Material

Inicialmente para a construção dos classificadores tentou usar-se o programa *Orange Canvas*<sup>®</sup> versão 2.7, treinando 5 classificadores distintos: *Naive Bayes*, SVM, k-NN, algoritmo CN2 e *Random Forests*, e validados através dos seguintes métodos: validação por camadas ou cruzada (*Cross-validation*), *leave-one-out* e *random sampling*. No entanto, este programa produz resultados de complexa interpretação, dado que constrói os classificadores em *black boxes*, não se conseguindo observar e modificar certos parâmetros, como o custo dos erros produzidos, e não se conseguindo reproduzir resultados. Assim, e como no caso em estudo existe uma necessidade de impor um custo maior à ocorrência de Falsos Negativos, concluiu-se que este *software* de aprendizagem automática não é o adequado para a análise em questão.

Desta forma, recorreu-se ao uso do programa de aprendizagem automática, CART<sup>®</sup> [33], versão 4.0, para o sistema operativo *Windows*<sup>®</sup>. Este *software* é baseado no algoritmo proposto por L. Breiman et al. em 1984 [32], vocacionado para a construção de árvores de decisão, sendo possível manipular parâmetros, como os custos dos erros, probabilidade de ocorrência das classes, entre outros, produzindo resultados inteligíveis e passíveis de serem ajustados aos problemas em estudo.

## 4 Resultados

Pretende-se, através de combinações entre os classificadores descritos na secção 4.2, obter uma melhor performance do sistema de classificação automática de RD.

Inicialmente estudou-se a performance de cada um dos classificadores automáticos de RD *per si*, Retmarker e DMUC, para a base de dados em questão.

Seguidamente procedeu-se a diferentes combinações entre os algoritmos de classificação automática distintos, para perceber se alguma dessas combinações produz resultados mais promissores no âmbito de rastreio de RD.

Numa segunda abordagem procedeu-se ao estudo dos classificadores automáticos de RD, Retmarker e DMUC, e respetivas fusões, desconsiderando as imagens classificadas como 0 (má qualidade) pelo algoritmo de classificação de Qualidade Geral.

Desta forma, espera-se obter uma análise comparativa entre a fusão dos vários classificadores automáticos e perceber de que maneira a qualidade da imagem deverá ser usada para potenciar o desempenho dos algoritmos de classificação automática de RD, se antes da utilização destes ou em paralelo.

### 4.1 Apresentação de Resultados

Repare-se que, no contexto de rastreio, a existência de um Falso Negativo significa a classificação de um paciente com patologia como sendo saudável. Desta forma o paciente não será encaminhado para o respetivo técnico de saúde a fim de receber tratamento ou controlo da doença, o que pode implicar consequências gravíssimas. Assim, é de máxima importância minimizar estes acontecimentos. Deste forma, há uma necessidade de impor ao classificador um maior custo para estes erros por forma a diminuir a probabilidade da sua ocorrência. Vai ser portanto estudada a performance dos classificadores resultantes para diferentes custos.

O método usado para validação dos classificadores resultantes das fusões foi a estimação cruzada ou por camadas (*Cross-Validation*), com 10 camadas. A probabilidade de cada classe é dada pela ocorrência destas no conjunto de treino.

O critério utilizado para seleção das melhores árvores de decisão foi a relação custo-complexidade.

Para as 3 bases de dados foram obtidos os seguintes resultados para o desempenho dos algoritmos de deteção de RD individual, e para as diferentes combinações entre estes e os algoritmos de Qualidade, respetivamente:

### Base de dados nº 1

Tabela 4.1 – Resultados obtidos utilizando a base de dados nº 1

Nº	Classificador	VP	FP	VN	FN	Sensibilidade	Especificidade	Custos (FN/FP)
1	RET	63	112	162	19	76,83%	59,12%	(1/1)
2	DMUC	57	47	227	25	69,51%	82,85%	(1/1)
3	RET+DMUC	50	25	249	32	60,98%	90,88%	(1/1)
4	RET+DMUC	52	45	229	30	63,41%	83,58%	(4/1)
5	RET+DMUC	67	129	145	15	81,71%	52,92%	(8/1)
6	RET+QUAL	65	118	156	17	79,27%	56,93%	(4/1)
7	RET+QUAL	67	119	155	15	81,71%	56,57%	(8/1)
8	RET+QUALs	63	112	162	19	76,83%	59,12%	(4/1)
9	RET+QUALs	71	174	100	11	86,59%	36,50%	(8/1)
10	DMUC+QUAL	56	46	228	26	68,29%	83,21%	(1/1)
11	DMUC+QUAL	57	47	225	25	69,51%	82,72%	(4/1)
12	DMUC+QUAL	61	59	215	21	74,39%	78,47%	(8/1)
13	DMUC+QUALs	42	27	247	40	51,22%	90,15%	(1/1)
14	DMUC+QUALs	64	60	214	18	78,05%	78,10%	(4/1)
15	DMUC+QUALs	73	99	175	9	89,02%	63,87%	(8/1)
16	RET+DMUC+QUAL	50	25	249	32	60,98%	90,88%	(1/1)
17	RET+DMUC+QUAL	37	25	249	45	45,12%	90,88%	(4/1)
18	RET+DMUC+QUAL	69	136	138	13	84,15%	50,36%	(8/1)
19	RET+DMUC+QUALs	46	20	254	36	56,10%	92,70%	(1/1)
20	RET+DMUC+QUALs	64	50	224	18	78,05%	81,75%	(4/1)
21	RET+DMUC+QUALs	73	99	175	9	89,02%	63,87%	(8/1)
22	RET s/ QUAL=0	61	105	155	15	80,26%	59,62%	(4/1)
23	DMUC s/ QUAL=0	56	45	215	20	73,68%	82,69%	(4/1)
24	RET+DMUC s/ QUAL=0	49	23	237	27	64,47%	91,15%	(1/1)
25	RET+DMUC s/ QUAL=0	56	45	215	20	73,68%	82,69%	(4/1)
26	RET+DMUC s/ QUAL=0	56	45	215	20	73,68%	82,69%	(8/1)
27	RET+QUALs s/ QUAL=0	61	105	155	15	80,26%	59,62%	(4/1)
28	RET+QUALs s/ QUAL=0	61	105	155	15	80,26%	59,62%	(8/1)
29	DMUC+QUALs s/ QUAL=0	45	29	231	31	59,21%	88,85%	(1/1)
30	DMUC+QUALs s/ QUAL=0	58	53	207	18	76,32%	79,62%	(4/1)
31	DMUC+QUALs s/ QUAL=0	67	92	168	9	88,16%	64,62%	(8/1)
32	RET+DMUC+QUALs s/ QUAL=0	47	18	242	29	61,84%	93,08%	(1/1)
33	RET+DMUC+QUALs s/ QUAL=0	57	46	214	19	75,00%	82,31%	(4/1)
34	RET+DMUC+QUALs s/ QUAL=0	65	89	171	11	85,53%	65,77%	(8/1)

## Base de dados nº 2

Tabela 4.2 – Resultados obtidos utilizando a base de dados nº 2

Nº	Classificador	VP	FP	VN	FN	Sensibilidade	Especificidade	Custos (FN/FP)	
1	RET		54	47	571	57	48,65%	92,39%	(1/1)
2	DMUC		60	48	570	51	54,05%	92,23%	(1/1)
3	RET+DMUC		46	9	609	65	41,44%	98,54%	(1/1)
4	RET+DMUC		60	48	570	51	54,05%	92,23%	(4/1)
5	RET+DMUC		68	86	532	43	61,26%	86,08%	(8/1)
6	RET+QUAL		53	47	571	58	47,75%	92,39%	(1/1)
7	RET+QUAL		53	47	571	58	47,75%	92,39%	(4/1)
8	RET+QUAL		63	105	513	48	56,76%	83,01%	(8/1)
9	RET+QUALs		42	26	592	69	37,84%	95,79%	(1/1)
10	RET+QUALs		41	70	568	50	45,05%	89,03%	(4/1)
11	RET+QUALs		67	121	490	44	60,36%	80,20%	(8/1)
12	DMUC+QUAL		57	42	576	54	51,35%	93,20%	(1/1)
13	DMUC+QUAL		58	48	570	53	52,25%	92,23%	(4/1)
14	DMUC+QUAL		60	48	570	51	54,05%	92,23%	(8/1)
15	DMUC+QUALs		54	38	580	57	48,65%	93,85%	(1/1)
16	DMUC+QUALs		60	48	570	51	54,05%	92,23%	(4/1)
17	DMUC+QUALs		70	109	509	41	63,06%	82,36%	(8/1)
18	RET+DMUC+QUAL		46	9	609	65	41,44%	98,54%	(1/1)
19	RET+DMUC+QUAL		58	48	570	53	52,25%	92,23%	(4/1)
20	RET+DMUC+QUAL		73	130	488	38	65,77%	78,96%	(8/1)
21	RET+DMUC+QUALs		46	9	609	65	41,44%	98,54%	(1/1)
22	RET+DMUC+QUALs		60	48	570	51	54,05%	92,23%	(4/1)
23	RET+DMUC+QUALs		76	117	501	35	68,47%	81,07%	(8/1)
24	RET s/ QUAL=0		53	46	513	47	53,00%	91,77%	(4/1)
25	DMUC s/ QUAL=0		57	42	517	43	57,00%	92,49%	(4/1)
26	RET+DMUC s/ QUAL=0		45	9	550	55	45,00%	98,39%	(1/1)
27	RET+DMUC s/ QUAL=0		57	42	517	43	57,00%	92,49%	(4/1)
28	RET+DMUC s/ QUAL=0		65	79	480	35	65,00%	85,87%	(8/1)
29	RET+QUALs s/ QUAL=0		42	26	533	58	42,00%	95,35%	(1/1)
30	RET+QUALs s/ QUAL=0		53	43	516	47	53,00%	92,31%	(4/1)
31	RET+QUALs s/ QUAL=0		53	46	513	47	53,00%	91,77%	(8/1)
32	DMUC+QUALs s/ QUAL=0		55	36	523	45	55,00%	93,56%	(1/1)
33	DMUC+QUALs s/ QUAL=0		57	36	523	43	57,00%	93,56%	(4/1)
34	DMUC+QUALs s/ QUAL=0		57	42	517	43	57,00%	92,49%	(8/1)
35	RET+DMUC+QUALs s/ QUAL=0		45	9	550	55	45,00%	98,39%	(1/1)
36	RET+DMUC+QUALs s/ QUAL=0		57	42	517	43	57,00%	92,49%	(4/1)
37	RET+DMUC+QUALs s/ QUAL=0		63	82	477	37	63,00%	85,33%	(8/1)

## Base de dados nº 3

Tabela 4.3 – Resultados obtidos utilizando a base de dados nº 3

Nº	Classificador	VP	FP	VN	FN	Sensibilidade	Especificidade	Custos (FN/FP)
1	RET	69	38	547	73	48,59%	93,50%	(1/1)
2	DMUC	77	42	543	65	54,23%	92,82%	(1/1)
3	RET+DMUC	55	6	579	87	38,73%	98,97%	(1/1)
4	RET+DMUC	91	74	511	51	64,08%	87,35%	(4/1)
5	RET+DMUC	91	74	511	51	64,08%	87,35%	(8/1)
6	RET+QUAL	69	38	547	73	48,59%	93,50%	(1/1)
7	RET+QUAL	69	38	547	73	48,59%	93,50%	(4/1)
8	RET+QUALs	69	38	547	73	48,59%	93,50%	(1/1)
9	RET+QUALs	80	61	524	62	56,34%	89,57%	(4/1)
10	RET+QUALs	112	343	242	30	78,87%	41,37%	(8/1)
11	DMUC+QUAL	77	42	543	65	54,23%	92,82%	(1/1)
12	DMUC+QUAL	77	50	535	65	54,23%	91,45%	(4/1)
13	DMUC+QUAL	88	92	493	54	61,97%	84,27%	(8/1)
14	DMUC+QUALs	75	34	551	67	52,82%	94,19%	(1/1)
15	DMUC+QUALs	77	42	543	65	54,23%	92,82%	(4/1)
16	DMUC+QUALs	120	328	257	22	84,51%	43,93%	(8/1)
17	RET+DMUC+QUAL	55	7	578	87	38,73%	98,80%	(1/1)
18	RET+DMUC+QUAL	88	74	511	54	61,97%	87,35%	(4/1)
19	RET+DMUC+QUAL	100	123	462	42	70,42%	78,97%	(8/1)
20	RET+DMUC+QUALs	55	6	579	87	38,73%	98,97%	(1/1)
21	RET+DMUC+QUALs	99	104	481	43	69,72%	82,22%	(4/1)
22	RET+DMUC+QUALs	100	132	453	42	70,42%	77,44%	(8/1)
23	RET s/ QUAL=0	66	37	493	61	51,97%	93,02%	(4/1)
24	DMUC s/ QUAL=0	73	37	493	54	57,48%	93,02%	(4/1)
25	RET+DMUC s/ QUAL=0	53	6	524	74	41,73%	98,87%	(1/1)
26	RET+DMUC s/ QUAL=0	86	68	462	41	67,72%	87,17%	(4/1)
27	RET+DMUC s/ QUAL=0	86	68	462	41	67,72%	87,17%	(8/1)
28	RET+QUALs s/ QUAL=0	66	34	496	61	51,97%	93,58%	(1/1)
29	RET+QUALs s/ QUAL=0	66	34	496	61	51,97%	93,58%	(4/1)
30	RET+QUALs s/ QUAL=0	101	271	259	26	79,53%	48,87%	(8/1)
31	DMUC+QUALs s/ QUAL=0	73	31	499	54	57,48%	94,15%	(1/1)
32	DMUC+QUALs s/ QUAL=0	73	31	499	54	57,48%	94,15%	(4/1)
33	DMUC+QUALs s/ QUAL=0	109	280	250	18	85,83%	47,17%	(8/1)
34	RET+DMUC+QUALs s/ QUAL=0	53	6	524	74	41,73%	98,87%	(1/1)
35	RET+DMUC+QUALs s/ QUAL=0	86	68	462	41	67,72%	87,17%	(4/1)
36	RET+DMUC+QUALs s/ QUAL=0	86	68	462	41	67,72%	87,17%	(8/1)

## 4.2 Análise de resultados

Como referido na subsecção 4.1, no caso em estudo é de extrema importância minimizar o número de casos de Doença erroneamente classificados, referidos como FN's, que idealmente deveria ser zero. Deste modo, o indicador de performance que melhor se ajusta ao caso em estudo, dando uma medida de segurança do sistema de classificação, será aquele que tem, diretamente, em consideração o total de ocorrências destes erros. O indicador que melhor representa esta relação e que será portanto usado como principal critério de performance dos

métodos de classificação é a **Sensibilidade**. A **Especificidade** é um bom critério a ser analisado em paralelo, pois reflete a vantagem do sistema automático, isto é, traduz a redução do esforço humano (*human burden*), estando diretamente relacionado com o Verdadeiros Negativos.

Repare-se que, para a fusão de classificadores, os melhores resultados foram obtidos para um custo de erro FN 8x maior do que o erro FP. Serão esses resultados que serão agora submetidos a uma análise comparativa mais detalhada.

Analisando os resultados obtidos a partir da base de dados nº1, comparando os dois algoritmos de classificação automática de RD, observa-se que o Retmarker (Tabela 4.1 – linha 1) possui uma maior sensibilidade (76,83%) que o DMUC (69,51%) à custa de um aumento de número de Falsos Positivos, tendo uma maior taxa de erro (131 erros / 356 casos = 36,79%). Assim constata-se que, apesar do algoritmo de classificação DMUC (Tabela 4.1 – linha 2) ser o melhor classificador no sentido em que produz menos erros (72 erros / 356 casos = 20,22%), o algoritmo Retmarker satisfaz melhor a condição do caso em estudo, isto é, produz maior sensibilidade.

Quando se efetuou a fusão de ambos os algoritmos de classificação automática, Retmarker e DMUC (Tabela 4.1 – linha 5), conseguiu-se obter um classificador com uma maior sensibilidade do que cada uma dos algoritmos por si, 81,71%. No entanto este classificador, produz uma maior taxa de erro (144 erros / 356 casos = 40,45%), dado que reflete um aumento do número de FP's.

A fusão do algoritmo Retmarker com o algoritmo de Qualidade (Tabela 4.1 – linha 7), *Overall Quality*, produziu a mesma sensibilidade da que o classificador resultante da fusão do algoritmo Retmarker com o algoritmo DMUC, 81,71%. No entanto, a especificidade aumentou, de 52,92% para 56,57%, refletindo uma diminuição da taxa de erro, (134 erros / 356 casos = 37,64%). Desta forma podemos imediatamente constatar que a fusão do algoritmo Retmarker com o algoritmo de Qualidade produz resultados mais promissores que a fusão do algoritmo Retmarker, com o algoritmo DMUC. Fazendo a combinação do algoritmo de Qualidade Geral, com o DMUC (Tabela 4.1 – linha 12), observa-se um aumento da sensibilidade (74,39%), com diminuição de especificidade (78,47%), em relação ao algoritmo DMUC por si, mas não supera a sensibilidade do algoritmo Retmarker, nem de nenhum dos classificadores obtidos a partir das restantes fusões.

Se em vez de se usar o classificador da Qualidade Geral, se usar os algoritmos de classificação das características da qualidade: Cor, Focagem, Iluminação e Contraste e proceder à fusão destes com o algoritmo Retmarker (Tabela 4.1 – linha 9) observa-se um aumento da sensibilidade do classificador resultante (86,59%) à custa de uma grande diminuição da especificidade deste (36,50%), o que traduz um aumento acentuado da taxa de erro (184 erros / 356 casos = 51,59%), sendo, de todas as fusões, a que resulta no classificador que apresenta a maior taxa de erro. Efetuado a fusão desta com o algoritmo DMUC (Tabela 4.1 – linha 15), obtém-se uma melhoria significativa no valor de sensibilidade em relação aos classificadores resultantes das restantes combinações, (89,02%), sendo o maior otimização da sensibilidade obtida para a base de dados em questão. No entanto, ocorre um aumento da taxa de erro (108 erros / 356 casos = 30,34%), que se traduz num decréscimo da especificidade (63,87%) em relação ao algoritmo DMUC (82,85%), e DMUC + Qualidade (78,47%). Ainda assim, este valor da especificidade supera os valores obtidos para as restantes combinações.

Considerando agora a fusão dos dois algoritmos de classificação automática de RD, Retmarker e DMUC com o algoritmo da Qualidade Geral (Tabela 4.1 – linha 18), obteve-se uma sensibilidade superior aos métodos de classificação individuais (84,15%), mas inferior à obtida pela fusão dos algoritmos Retmarker e DMUC com o algoritmos das características da qualidade (86,59% e 89,02% respetivamente). No entanto, produz menor taxa de erro (149 erros / 356 casos = 41,85%) que o classificador resultante Retmarker combinado com os algoritmos das características da qualidade (51,69%), mas maior de que estes combinados com o DMUC (30,34%).

Combinando os dois algoritmos de deteção automática de RD, Retmarker e DMUC com os algoritmos das características da Qualidade (Tabela 4.1 – linha 21) obteve-se exatamente a mesma árvore de decisão que para a junção do algoritmo DMUC com as características da Qualidade (sensibilidade = 89,02% e especificidade = 63,87%), o que indica que é atribuída importância zero ao algoritmo Retmarker. Repare-se que as combinações que não incluem o Retmarker produzem menor taxa de erro, o que indica que este atributo tem menos semelhanças com o *Ground Truth*, isto é, a classificação manual de RD atribuída pelo técnico a cada imagem. Isto constitui uma explicação para que reduzida importância que é dada à decisão do Retmarker, quando presente o conhecimento da decisão do DMUC.

Para a segunda abordagem, far-se-á uma análise dos dados obtidos para a performance dos algoritmos Retmarker e DMUC, individualmente e combinados entre si e com as quatro características da Qualidade, desconsiderando as imagens assinaladas pelo algoritmo de Qualidade Geral como tendo qualidade inadequada.

Para o algoritmo Retmarker (Tabela 4.1 – linha 22), foi obtida uma sensibilidade superior (80,26%) à sensibilidade obtida (73,68%) pelo DMUC (Tabela 4.1 – linha 23), no entanto esta diferença de +6,58% na sensibilidade do método é conseguida através de uma diferença de cerca de +0.846x no número de Falsos Negativos, resultando numa taxa de erro 1.846x superior (120 erros / 336 casos = 35,71% *versus* 65 erros / 336 casos = 19,35%). Ainda assim, estes resultados mostram seguir a mesma tendência que os obtidos para a totalidade das imagens (Tabela 4.1 – Linha 1 e 2) e observam-se melhores que estes.

Juntando o algoritmo Retmarker com o DMUC (Tabela 4.1 – linha 26), obtém-se exatamente o mesmo resultado que para o DMUC por si só (Tabela 4.1 – linha 23), o que mostra que mais uma vez é atribuída importância zero à decisão resultante pelo classificador Retmarker.

Ao efetuar a combinação das características da Qualidade com cada um dos classificadores de RD, obteve-se uma melhoria no valor da sensibilidade em relação a cada um deles individualmente. A combinação DMUC + 4 características da Qualidade (Tabela 4.1 – linha 31) provou ser melhor que a combinação destas com o Retmarker (Tabela 4.1 – linha 28), possuindo não só maior sensibilidade (88,16% contra 80,26%), bem como a menor taxa de erro (101 erros / 336 casos = 30,06% *versus* 120 erros / 336 casos = 35,71%), provando ser o método que possui melhor desempenho e produzindo os melhores resultados no âmbito em estudo, para esta segunda abordagem. Podemos observar que para a primeira abordagem a combinação que reflete os melhores resultados é também a fusão do DMUC com as características da Qualidade (Tabela 4.1 – linha 15).

De maneira a finalizar a análise da base de dados nº 1, combinou-se os algoritmos, Retmarker e DMUC com as características da Qualidade (Tabela 4.1 – linha 34), observando-se um bom valor de sensibilidade (85,53%) para uma taxa de erro de 29,76% (100 erros / 336 casos) . Podemos ver que a introdução das características da Qualidade à fusão Retmarker +

DMUC (Tabela 4.1 – linha 26) aumentou bastante a sensibilidade do método, no entanto vemos que a junção do Retmarker à fusão DMUC + 4 características da Qualidade (Tabela 4.1 – linha 31) introduz um aumento dos FN's e conseqüente diminuição da sensibilidade.

Procedendo à análise dos dados obtidos para a 2ª base de dados, é possível observar que o classificador automático de RD que produz melhores resultados para o estudo em causa é o DMUC, apresentando uma sensibilidade de 54,05% (Tabela 4.2 – linha 2) em contraste com 48,65% obtidos pelo Retmarker (Tabela 4.2 – linha 1). O DMUC representa também o melhor classificador entre os dois, possuindo uma taxa de erro menor (99 erros / 729 casos = 13,58%) que o Retmarker (104 erros / 729 casos = 14,27%).

Ao efetuar-se a fusão entre os dois algoritmos de classificação automática de RD, Retmarker e DMUC (Tabela 4.2 – linha 5) observou-se uma melhoria significativa na sensibilidade (61,26%) comparativamente ao valor obtido por cada um deles isoladamente (Tabela 4.2 – linha 1 e 2), à custa de um aumento de Falsos Positivos, resultando numa diminuição da especificidade (86,08%) e aumento da taxa de erro (129 erros / 729 casos = 17,70%).

Fundindo cada um dos classificadores, Retmarker e DMUC, individualmente com o algoritmo de classificação automática de Qualidade Geral, obteve-se um melhor valor de sensibilidade (56,76%) para fusão Retmarker + Qualidade (Tabela 4.2 – linha 8), do que o valor obtido (54,05%) para fusão DMUC + Qualidade (Tabela 4.2 – linha 14). Contudo esta pequena diferença de 2,71 pontos percentuais é obtida à custa de um enorme aumento da taxa de erro (99 erros / 729 casos = 13,58% *versus* 153 erros / 729 casos = 20,99%) e conseqüente perda de especificidade (92,23% para 83,01% respetivamente). Repare-se ainda que os resultados para a fusão DMUC+ Qualidade (Tabela 4.2 – linha 14) é exatamente igual ao resultados produzidos pelo DMUC por si (Tabela 4.2 – linha 2), constatando-se que a Qualidade tem importância nula na combinação desta com o algoritmo DMUC.

Ao utilizarem-se os algoritmos de classificação das 4 características da qualidade de imagem, em vez da Qualidade Geral, e combinando-os com cada um dos algoritmos de classificação automática de RD obteve-se uma melhor sensibilidade (63,06%) para a combinação DMUC + 4 características da Qualidade (Tabela 4.2 – linha 17) do que a obtida (60,36%) para a combinação destas com o Retmarker (Tabela 4.2 – linha 11). Além do

classificador resultante da fusão DMUC + 4 características da Qualidade (Tabela 4.2 – linha 17) ser, dos dois, o que melhor se ajusta ao estudo, é também aquele que produz menor taxa de erro ( $150 \text{ erros} / 729 \text{ casos} = 20,58\%$  *versus*  $165 \text{ erros} / 729 \text{ casos} = 22,63\%$ ) e portanto maior especificidade (82,36% contra 80,20% respetivamente), tratando-se assim do melhor classificador entre os dois.

Quando combinados os Retmarker e DMUC como o algoritmo da Qualidade Geral (Tabela 4.2 – linha 20) obteve-se uma melhoria no valor da sensibilidade em relação às restantes combinações (65,77%) mas à custa de um aumento da taxa de erro ( $168 \text{ erros} / 729 \text{ casos} = 23,05\%$ ), sendo o classificador que produz o maior número de erros para esta abordagem na base de dados em questão.

Ao proceder-se à combinação dos dois algoritmos de deteção automática de RD, Retmarker e DMUC com os algoritmos das características da Qualidade (Tabela 4.2 – linha 23), não só há uma melhoria na sensibilidade do método de classificação, em relação à combinação anterior (68,47%), como existe ainda um decréscimo da taxa de erro em relação à mesma ( $152 \text{ erros} / 729 \text{ casos} = 20,85\%$ ). Trata-se portanto do classificador que produz os resultados mais ajustados ao estudo em curso, para a base de dados em questão.

Realizada a segunda abordagem, na qual todas as imagens de que o algoritmo automático de classificação da Qualidade Geral, considerou não adequadas foram removidas da base de dados, observou-se que o desempenho do classificador automático de RD, DMUC por si (Tabela 4.2 – linha 25), é melhor que o desempenho do Retmarker (Tabela 4.2 – linha 24), possuindo uma sensibilidade maior (57,00% contra 53,00% respetivamente) e uma taxa de erro menor ( $85 \text{ erros} / 659 \text{ casos} = 12,90\%$  *versus*  $93 \text{ erros} / 659 \text{ casos} = 14,11\%$ ). De notar ainda, que estes valores são melhores que os obtidos para estes classificadores na primeira abordagem (Tabela 4.2 – linha 1 e 2).

Juntando os dois classificadores, Retmarker e DMUC (Tabela 4.2 – linha 28), obteve-se uma melhoria significativa na sensibilidade (65,00%) em relação a cada um destes individualmente, tratando-se do maior valor de sensibilidade obtido com esta abordagem, para esta base de dados. Contudo, esta melhoria é conseguida implicando um aumento dos Falsos Positivos e consequentemente da taxa de erro ( $114 \text{ erros} / 659 \text{ casos} = 17,30\%$ ).

Analisando as fusões Retmarker + 4 características da Qualidade (Tabela 4.2 – linha 31), e DMUC + 4 características da Qualidade (Tabela 4.2 – linha 34), observam-se exatamente os mesmos valores obtidos para cada um dos classificadores de RD respetivos, individualmente. Assim conclui-se que os algoritmos de classificação das características da Qualidade têm nestes dois casos um peso nulo na construção das respetivas árvores de decisão.

Para finalizar a análise desta base de dados, observa-se os resultados da fusão dos três classificadores em caso, Retmarker + DMUC + 4 características da Qualidade (Tabela 4.2 – linha 37). A sensibilidade obtida (63,00%) é melhor que as restantes à exceção da obtida pela fusão Retmarker + DMUC (Tabela 4.2 – linha 28), no entanto possui a maior taxa de erro de todos classificadores resultantes desta abordagem (119 erros / 659 casos = 18,06%).

Procedendo à análise dos resultados da última base de dados em estudo, consegue observar-se que o classificador DMUC (Tabela 4.3 – linha 2) traduz por si, melhor desempenho que o classificador Retmarker (Tabela 4.3 – linha 1), tendo uma maior sensibilidade (54,23% contra 48,59%) e menor taxa de erro (107 erros / 727 casos = 14,72% *versus* 111 erros / 727 casos = 15,27%), apesar de possuir uma especificidade ligeiramente mais baixa (92,82% contra 93,50%, respectivamente). Fazendo a fusão destes dois classificadores (Tabela 4.3 – linha 5), observa-se uma melhoria significativa na sensibilidade (64,08%) implicando uma descida na especificidade (87,35%) e aumento da taxa de erro (125 erros / 727 casos = 17,19%) como consequência do aumento dos Falsos Positivos.

Combinou-se o algoritmo de classificação da Qualidade Geral, com cada um dos classificadores de deteção automática de RD. A combinação da Qualidade Geral com o algoritmo DMUC (Tabela 4.3 – linha 13) produziu não só maior sensibilidade (61,97%) mas também maior taxa de erro (146 erros / 727 casos = 20,08%) que a combinação da mesma com o algoritmo Retmarker (Tabela 4.3 – linha 7), que obteve os mesmos valores que o mesmo por si, indicando que a Qualidade teve importância nula na construção da árvore de decisão resultante desta combinação.

Ao proceder-se à fusão de ambos os algoritmos, Retmarker e DMUC com os algoritmos das quatro características da qualidade, os resultados obtidos melhoram bastante relativamente à sensibilidade mas à custa de um enorme aumento da taxa de erros, devido a subida acentuada do

número de falsos negativos. A fusão que obteve melhores resultados, tanto a nível de sensibilidade (84,51%) como de taxa de erro (350 erros / 727 casos = 48,14%) foi aquela entre o algoritmo DMUC e os algoritmos das características (Tabela 4.3 – linha 16), sendo o melhor resultado de sensibilidade obtido para esta abordagem com esta base de dados. A combinação das características da Qualidade com o Retmarker (Tabela 4.3 – linha 10) deu origem a uma sensibilidade de 78,87% e taxa de erro de 51,31% (373 erros / 727 casos).

As duas últimas combinações feitas no âmbito desta abordagem foram entre os dois classificadores automáticos de RD e a Qualidade Geral (Tabela 4.3 – linha 19), e posteriormente na vez deste, as características da Qualidade (Tabela 4.3 – linha 22). Estas duas fusões, resultaram na mesma sensibilidade (70,42%), contudo a junção dos algoritmos DMUC e Retmarker com as quatro características da Qualidade produziu uma maior taxa de erro (174 erros / 727 casos = 23,93%) que a combinação entre o DMUC, Retmarker e Qualidade Geral, (165 erros / 727 casos = 22,70%).

Finalizando a análise, observe-se agora a segunda abordagem para esta base de dados. Novamente o algoritmo DMUC (Tabela 4.3 – linha 24), por si, resultou num melhor valor de sensibilidade (57,48%) que o valor obtido (51,97%) com algoritmo Retmarker (Tabela 4.3 – linha 23). Esta diferença entre a sensibilidade dos métodos foi conseguida sem um decréscimo na especificidade, sendo a mesma entre eles (93,02%), o que significa que o classificador DMUC produz assim uma menor taxa de erro (91 erros / 657 casos = 13,85%) que o Retmarker (98 erros / 657 casos = 14,92%). Os valores da sensibilidade para cada um dos métodos de classificação automática de RD, mostraram ser melhores que os obtidos para os mesmos utilizando a primeira abordagem (Tabela 4.3 – linha 1 e 2).

Procedendo-se à combinação dos algoritmos de classificação Retmarker e DMUC (Tabela 4.3 – linha 27) conseguiu-se uma melhoria significativa na sensibilidade (67,72%) em relação a cada um deles por si, a custo de um aumento dos Falsos Positivos traduzido no decréscimo da sensibilidade (87,17%) e aumento da taxa de erro do classificador (109 erros / 657 casos = 16,59%).

Introduzindo agora os classificadores das características da Qualidade, e juntando-os ao Retmarker e DMUC individualmente, a combinação DMUC + 4 características da Qualidade

(Tabela 4.3 – linha 33) produziu melhores resultados que a combinação Retmarker + 4 características da Qualidade (Tabela 4.3 – linha 30) tendo conseguido uma sensibilidade significativamente maior (85,83% contra 79,53%) sem comprometer demasiado a especificidade deste (47,17% contra 48,87%) traduzindo assim uma pequena diferença entre eles na taxa de erro (298 erros / 657 casos = 45,36% *versus* 297 erros / 657 casos = 45,21%). Esta combinação provou assim ser, desta última abordagem, a que produz melhores resultados no âmbito do estudo desenvolvido. Repara-se que na primeira abordagem a fusão DMUC + 4 características da Qualidade (Tabela 4.3 – linha 16) foi também aquela que produziu resultados mais promissores em relação à sensibilidade do método de classificação resultante, tendo obtido valores muito similares.

Por fim, efetuou-se a fusão das características da qualidade com os dois classificadores automáticos de RD, Retmarker e DMUC (Tabela 4.3 – linha 36). Esta combinação, resultou numa árvore de decisão igual à obtida pela fusão Retmarker + DMUC (Tabela 4.3 – linha 27), produzindo os mesmos resultados, sensibilidade de 67,72% e taxa de erro de 16,59% (109 erros / 657 casos). Desta maneira observa-se que as características da Qualidade não foram tomadas em consideração aquando da construção da árvore de decisão.

## 5 Conclusão e trabalho futuro

A base de dados nº 2 produz resultados em alguns aspetos discrepantes em relação às outras duas bases de dados, de forma que não será tomada em consideração para aferição de conclusões. Essa discrepância prende-se com o objetivo da classificação manual inicial ser distinta do objetivo no caso dos trabalhos descritos.

Considerando a base de dados nº 1, na primeira análise, comparando os dois algoritmos de classificação de automática de RD, Retmarker e DMUC, é possível concluir que o Retmarker é o método mais sensível, produzindo uma sensibilidade = 76,83% enquanto que o DMUC possui uma sensibilidade de 69,51%. No entanto, esta melhoria é conseguida à custa de um aumento dos falsos positivos, produzindo uma maior taxa de erro (36,79%) que o DMUC (20,22%). Assim, apesar de no caso em estudo o melhor método ser aquele que possui maior sensibilidade, ou seja, o Retmarker, para avaliação do classificador em si, o melhor é aquele que produz menor taxa de erro e portanto é mais similar ao *Ground Truth*, neste caso o DMUC. O mesmo não acontece para a base de dados nº 3, na qual o classificador que obtém um melhor desempenho é o DMUC com uma sensibilidade de 54,23% em contraste com 48,59% obtida pelo Retmarker. Nesta análise, verifica-se também que este classificador, produz não só a melhor sensibilidade como a menor taxa de erro (14,74% contra 15,27%), o que é indicativo de que não só este algoritmo de classificação automática de RD é o mais adequado ao âmbito do estudo realizado como é, de uma forma geral, o melhor classificador de entre os dois.

A diferença observada entre os resultados das duas bases de dados em questão, pode refletir o tipo de lesões presentes nas imagens do fundo do olho presentes em cada uma destas, pois recorde-se que o algoritmo DMUC é um algoritmo de classificação de RD mais completo, estando desenhado para encontrar diferentes tipo de lesões, enquanto que o Retmarker foi construído de maneira a encontrar apenas microaneurismas. Por outro lado um detetor que tente detetar mais lesões pode gerar mais Falsos Positivos.

Na segunda abordagem, na qual todas as imagens consideradas como qualidade não adequada pelo Qualidade Geral foram removidas, para a primeira base de dados observou-se que a relação entre os classificadores Retmarker e DMUC manteve-se, tendo o Retmarker produzido novamente a melhor sensibilidade e maior taxa de erro em relação ao DMUC. Na

base de dados nº 3 também a relação entre eles também foi mantida tendo o DMUC produzido a melhor sensibilidade e menor taxa de erro. Contudo, em ambas as bases de dados ocorreu uma melhoria significativa na sensibilidade de ambos os classificadores automáticos de RD, o que indica que a Qualidade é de facto um fator importante a ter em conta na avaliação automática desta patologia, e que as imagens de qualidade inadequada estão diretamente relacionadas com falhas por parte destes sistemas de classificação automática, levando ao aumento do número de Falsos Negativos.

Avaliando agora os resultados de todos os testes efetuados, de todas as combinações estudadas, para a base de dados nº 1 os classificadores que obtiveram a maior sensibilidade e portanto aqueles que mais se ajustam ao caso em estudo foram as seguintes combinações (representadas a amarelo na Tabela 4.1): DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste (Tabela 4.3 – linha 15) e o Retmarker + DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação (Tabela 4.3 – linha 21). Estas duas fusões produziram exatamente os mesmos resultados, sensibilidade = 89,02% e especificidade = 63,87%. Assim concluiu-se que o Retmarker tem peso nulo na construção da respetiva árvore de decisão, o que indica que o DMUC apesar de ter produzido menor sensibilidade por si que o Retmarker, aparenta ser o classificador que produz, de forma geral, mais resultados em acordo com a classificação manual e mais influenciados pelos parâmetros da qualidade da imagem. Pelo que, o melhor método é resultante da fusão do classificador DMUC com os classificadores das características da Qualidade de imagem. O erro obtido para este método foi de 30,34%, o que comparativamente com os restantes, representa um bom compromisso entre sensibilidade e taxa de erro.

Na análise da base de dados nº3 foi possível constatar que tal como na base de dados nº1 a combinação que produziu os resultados mais favoráveis para o estudo desenvolvido foi também a combinação DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste, marcada a amarelo na Tabela 4.3, tendo conseguido uma sensibilidade de 84,51% e especificidade de 43,93% (Tabela 4.3 – linha 16). Tal reforça a observação de que o algoritmo de classificação automática DMUC, não só possui melhor desempenho no âmbito do estudo desenvolvido, como este pode ser potenciado pela informação proveniente dos algoritmos de classificação das características da Qualidade, sendo esta mais completa e com maior influência positiva sobre os classificadores automáticos de RD, do que a informação proveniente do algoritmo da Qualidade Geral.

Os resultados obtidos a partir da segunda abordagem mostraram estar de acordo com a primeira abordagem, tendo sido a combinação DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste aquela que revelou os melhores resultados para ambas as bases de dados (marcadas a cor-de-rosa nas Tabelas 4.1 e 4.3 respetivamente). Na base de dados nº 1 foi conseguido valores muito similares, tendo sido a sensibilidade ligeiramente menor relativamente à conseguida na primeira abordagem, contudo na base de dados nº3 apesar de os valores serem muito aproximados dos produzidos na primeira abordagem, a sensibilidade obtida foi ligeiramente maior. Dado que, esta diferença na sensibilidade entre as duas abordagens para a referida combinação é bastante pequena, sendo de apenas 0,86% para a base de dados nº1 e 1,32% para a base de dados nº3, pode concluir-se que quando presente a informação das características da qualidade, a Qualidade Geral da imagem deixa de influenciar tanto desempenho do classificador DMUC, tornando-se a presença ou ausência de imagens consideradas inadequadas pelo algoritmo Qualidade Geral irrelevante.

Apesar das diferenças quantitativas entre os resultados obtidos pelas duas bases de dados (nº1 e nº3), que pode ser explicada pelo facto destas bases de dados terem sido obtidas em contextos diferentes, sendo ambas fora do contexto mais frequente de rastreio. A primeira foi, como já referido, adquirida em contexto clínico, havendo uma prevalência maior de casos de patologia. A outra, apesar de ter sido adquirida em contexto de rastreio, a qualidade desta em termos gerais aparenta mais falhas do que seria desejável. Observa-se que, e de forma geral, o mérito dos classificadores obtidos pelas diferentes combinações e para as diferentes abordagens, mantém-se. Ou seja, que se torna possível a utilização de algoritmos automáticos em condições mais adversas, ao fazer a *fusion* com algoritmos de qualidade, conseguindo-se assim obter um sistema ainda mais robusto. Desta forma, sumariamente, podemos tirar as seguintes conclusões:

- O melhor método de classificação automática de RD, por si, é o algoritmo DMUC;
- A combinação que otimiza o deteção automática de RD é a fusão entre os seguintes classificadores: DMUC + 4 características da Qualidade: Cor, Focagem, Contraste e Iluminação.
- A utilização da informação proveniente dos algoritmos de classificação das características da Qualidade: Cor, Focagem, Contraste e Iluminação, melhora bastante o

desempenho dos classificadores de deteção automática de RD, tendo uma influência mais positiva do que a utilização da informação proveniente da Qualidade Geral;

- Desconsiderando as imagens marcadas pelo algoritmo de Qualidade Geral como inadequadas, simulando o impedimento do avanço destas imagens no programa de rastreio, consegue-se um grande melhoria no desempenho dos classificadores automáticos de deteção de RD, evidenciando que as imagens de má qualidade estão na origem de erros de classificação por parte destes, reforçando portanto a importância de ter em consideração a qualidade das imagens para fins de rastreios automáticos;
- Desconsiderando as imagens marcadas pelo algoritmo de Qualidade Geral como inadequadas, ao ter-se em conta a informação proveniente dos algoritmos de classificação das características da Qualidade: Cor, Focagem, Contraste e Iluminação a par com os algoritmos de deteção automática de RD chega-se à conclusão que a presença ou ausência de imagens de má qualidade deixa de influenciar positiva ou negativamente os algoritmos de deteção;
- Na base de dados nº3, a qual apresenta de forma geral pior qualidade, observou-se que a melhoria na sensibilidade dos classificadores resultantes das fusões é ainda mais acentuada. Sugerindo que a melhoria potenciada pelas combinações dos algoritmos de classificação automática de RD com os algoritmos de avaliação da Qualidade, depende das características da qualidade da base de dados.

As observações produzidas pelo estudo desenvolvido vieram mostrar que a fusão entre algoritmos de deteção automática de RD com algoritmos de avaliação de Qualidade, nunca feita até então, acrescenta uma melhoria significativa à sensibilidade dos algoritmos de deteção de lesões isoladamente. O acréscimo de sensibilidade revelou-se entre os 10p.p. e os 30p.p., dependendo do algoritmo de deteção e da base de dados, o que sugere uma influência das características das bases de dados no desempenho destes classificadores. Estes resultados vêm assim, acrescentar novas possibilidades ao processo de automatização do rastreio de Retinopatia Diabética, otimizando os seus resultados.

Repare-se que para as bases de dados utilizadas foram obtidos resultados de sensibilidade para o algoritmo Retmarker, por si, inferiores ao obtido pelo estudo realizado por Oliveira et al. [2], em 2011. Desta forma seria de grande interesse efetuar esta mesma análise

para a base de dados usada por Oliveira et al. [2] em 2011, e outras similares, de grandes dimensões, e mais perto do contexto normal de rastreio, de maneira a obter uma validação mais precisa e apurada destes resultados. Verificando deste modo se a melhoria esperada elevaria os resultados de sensibilidade para perto do ideal.



## Referências

- [1] J. Cunha-Vaz, R. Bernardes, C. Lobo, I. Pires, L. Ribeiro, and T. Santos, “Computer aided detection of diabetic retinopathy progression”, *Digital Teleretinal Screening*, pp. 59-66, 2012.
- [2] C. M. Oliveira, L. M. Cristóvão, M. L. Ribeiro, and J. R. Faria Abreu, “Improved Automated Screening of Diabetic Retinopathy”. *Ophthalmologica*, vol. 4, no. 226, pp. 191–197, Aug. 2011.
- [3] D. Usher, M. Dumskyj, M. Himaga, T.H. Williamson, S. Nussey, J. Boyce, “Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening”, *Diabet Med*, vol. 21, pp.84-90
- [4] A. D. Fleming, S. Philip, K. a Goatman, J. a Olson, and P. F. Sharp, “Automated microaneurysm detection using local contrast normalization and local vessel detection”, *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1223–32, Sep. 2006.
- [5] S. Philip, a D. Fleming, K. a Goatman, S. Fonseca, P. McNamee, G. S. Scotland, G. J. Prescott, P. F. Sharp, and J. a Olson, “The efficacy of automated ‘disease/no disease’ grading for diabetic retinopathy in a systematic screening programme”, *Br. J. Ophthalmol.*, vol. 91, no. 11, pp. 1512–7, Nov. 2007.
- [6] M. Niemeijer, M. D. Abràmoff, B. van Ginneken, “Information fusion for diabetic retinopathy CAD in digital color fundus photographs”, *IEEE Trans Med Imaging*, vol. 28, pp. 775–85, 2009.
- [7] M. D. Abràmoff, J. Reinhardt, S. R. Russell, J. Folk, V. Mahajan, M. Niemeijer, and G. Quellec, “Automated Early detection of Diabetic Retinopathy”, *Ophthalmology*, vol.117, no.6, pp. 1147-1154, 2010.
- [8] G. Quellec, M. Lamard, P.M. Josselin, G. Czuvgel, B. Cochener, C. Roux, “Optimal wavelet transform for the detection of microaneurysmsin retina photographs”, *IEEE Trans Med Imaging*, vol.27, pp.1230-1241, 2008.

- [9] M. Niemeijer, M. Loog, M. D. Abràmoff, B. van Ginneken, “On combining computer-aided detection systems”, *IEEE Trans Med Imaging*, pp. 215-233, 2011.
- [10] H. F. Jelinek, R. Pires, R. Padilha, S. Goldenstein, J. Wainer, T. Bossomaier, and A. Rocha, “Data fusion for multi-lesion Diabetic Retinopathy detection,” *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium*, pp. 1–4, 2012.
- [11] O. Dept. of, F.P.R.C. Visual Sciences of the University of Wisconsin –Madison. ARIC Grading Protocol, 1995.
- [12] S.C. Lee, Y. Wang, “Automatic Retinal image quality and assessment and enhancement”, *Proceedings of SPIE Medical Imaging: Image Processing*, pp. 1581-1590, 1999.
- [13] M. Lalonde, L. Gagnon, M.C. Boucher, “Automatic visual quality assessment in optical fundus images”, *Proceedings of Vision Interface*, pp. 259-264, 2001.
- [14] H. Bartling, P. Wanger, L. Martin, “Automated quality evaluation of digital fundus photographs”, *Acta Ophthalmologica*, vol. 87, pp. 643-647, 2009.
- [15] H. Davis, S. Russel, E. Barriga, M. Abramoff, P. Soliz, “Vision-based real-time retinal image quality assessment”, *22nd IEEE International Symposium on Computer-Based Medical Systems*, pp.1-6, 2009.
- [16] D.B. Usher, M.Himaga, M.J. Dumskyj, “Automated assessment of digital fundus image quality using detect vessel area”, *Proceedings of Medical Image Understanding and Analysis*, pp. 81-84, 2003.
- [17] J. Lowell, A. Hunter, M. Habib, D. Steel, “Automated quantification of fundus image quality”, *Proceedings of the 3<sup>rd</sup> European Medical and Biological, Engineering Conference*, 2005.
- [18] M. Niemeijer, M. D. Abràmoff, and B. van Ginneken, “Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening”, *Med. Image Anal.*, vol. 10, no. 6, pp. 888–98, Dec. 2006.

- [19] A. D. Fleming, S. Philip, K. a Goatman, J. a Olson, and P. F. Sharp, “Automated assessment of diabetic retinal image quality based on clarity and field definition”, *Invest. Ophthalmol. Vis. Sci.*, vol. 47, no. 3, pp. 1120–5, Mar. 2006.
- [20] L. Giancardo, M.D. Abramoff, E. Chaum, T.P. Karnowski, F. Meriaudeau, K.W. Tobin, “Elliptical local vessel density: a fast and robust quality metric for retinal images”, *Conference Proceedings of the international Conference of IEEE Engineering in Medicine and Biology Society*, pp.3534-3537, 2008.
- [21] A. Hunter, J.A. Lowell, M. Habib, B. Ryder, A. Basu, D. Steel, “An automated retinal image quality grading algorithm”, *Conference Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society*, pp.5955-5958, 2011.
- [22] J. Paulus, J. Meier, R. Bock, J. Hornegger, G. Michelson, “Automated quality assessment of retinal fundus photos”, *International Journal of Computer Assisted Radiology and Surgery* 5, pp. 557-564, 2010.
- [23] A.D. Fleming, S. Philip, K.A. Goatman, P.F. Sharp, J.A. Olson, “Automated clarity assessment of retinal images using regionally based structural and statistical measures”, *Medical Engineering & Physics*, vol.34, pp. 849-859, 2012.
- [24] J. M. Pires Dias, C. M. Oliveira, and L. A. da Silva Cruz, “Retinal image quality assessment using generic image quality indicators,” *Inf. Fusion*, Aug. 2012.
- [25] J. Fonseca, “Classificadores”, *Acetatos de aulas - Sistemas de Informação Médica*, Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa, 2012/2013
- [26] S.M. Weiss, C.A. Kullkowsky, *Computer Systems That Learn*, 1990.
- [27] L. Merschmann, “Classificação probabilística baseada em análise de padrões”, Tese (Pós-Graduação em Computação) – Universidade Federal Fluminense, 2007.
- [28] C. Halmenschlager, “Um algoritmo de indução de árvores e regras de decisão”, Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

- [29] J. Gama, “Combining classification algorithms”, Tese (Doutoramento) – Departamento de Ciência de Computadores, Universidade do Porto, Porto, 1999.
- [30] J. Fonseca, “Indução de árvores de decisão: HistClass – proposta de um algoritmo não paramétrico”, Dissertação (Mestrado em Engenharia Informática) – Departamento de Informática, Universidade Nova de Lisboa, Lisboa, 1994.
- [31] F. Esposito, D. Malerba, G. Semeraro, “A comparative analysis of methods for pruning decision trees”, *IEEE Transactions on Pattern Analysis And Machine Intelligence*, New York, vol.19, no.5, pp.476-491, May 1997.
- [32] L. Brieman, J. Friedman, R. Olshen, C. Stone, “Classification and Regression Trees”, *Pacific Grove: Wadsworth*, 1984.
- [33] D. Steinberg, P. Colla, “CART® - Classification and Regression Trees”, *CA: Salford Systems*, San Diego, 1997.

## A. Anexo 1: Árvores de decisão obtidas

### A.1 Base de Dados nº 1

- Retmarker + DMUC

i. Custos FN = 1; Custos FP=1

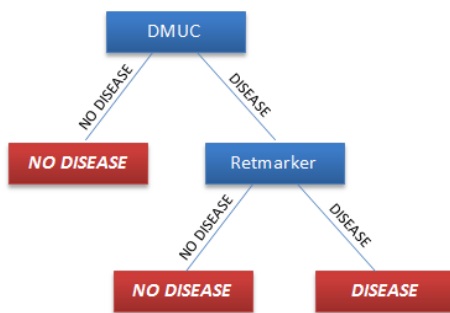


Figura A.1.1- árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=1 e FP =1

ii. Custos FN = 4; Custos FP=1

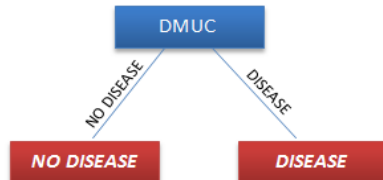


Figura A.1.2 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

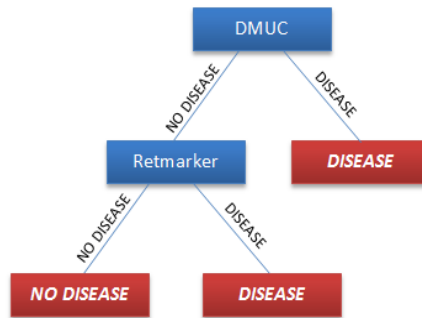


Figura A.1.3 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=8 e FP =1

- **Retmarker + Qualidade Geral**

- i. **Custos FN = 1; Custos FP=1**

Para custos FN=1 e FP=1, o CART não consegue construir nenhuma árvore de decisão.

- ii. **Custos FN = 4; Custos FP=1**

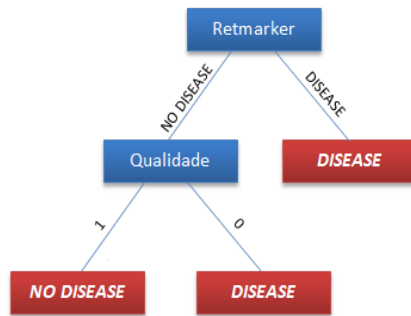


Figura A.1.4 -árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=4 e FP =1

- iii. **Custos FN = 8; Custos FP=1**

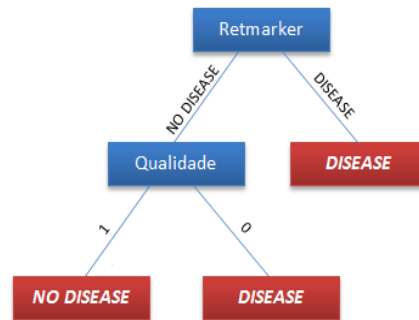


Figura A.1.5 -árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=8 e FP =1

- **Retmarker + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

**i. Custos FN = 1; Custos FP=1**

Para custos FN=1 e FP=1, o CART não consegue construir nenhuma árvore de decisão.

**ii. Custos FN = 4; Custos FP=1**

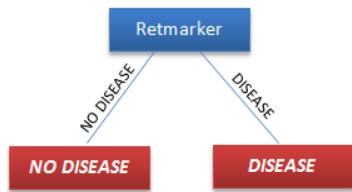


Figura A.1.6 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=4 e FP =1

**iii. Custos FN = 8; Custos FP=1**

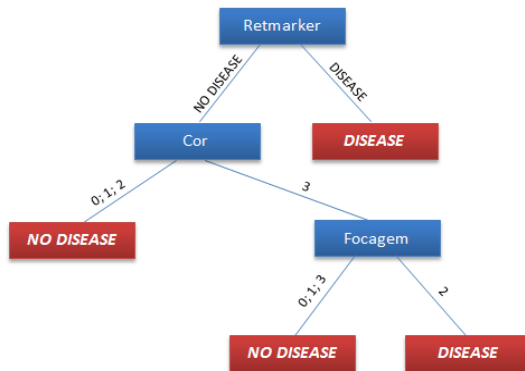


Figura A.1.7 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=8 e FP =1

• **DMUC + Qualidade Geral**

**i. Custos FN = 1; Custos FP=1**

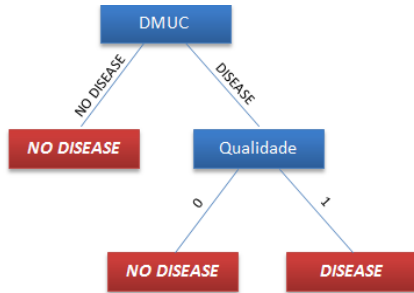


Figura A.1.8 -árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

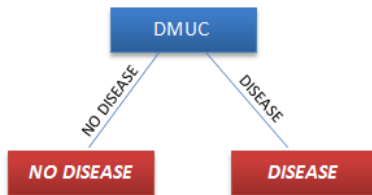


Figura A.1.9 -árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=4 e FP =1

**iii. Custos FN = 8; Custos FP=1**

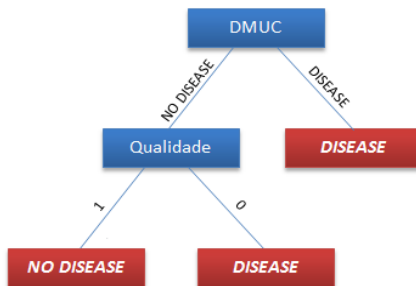


Figura A.1.10 -árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=8 e FP =1

- **DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

**i. Custos FN = 1; Custos FP=1**

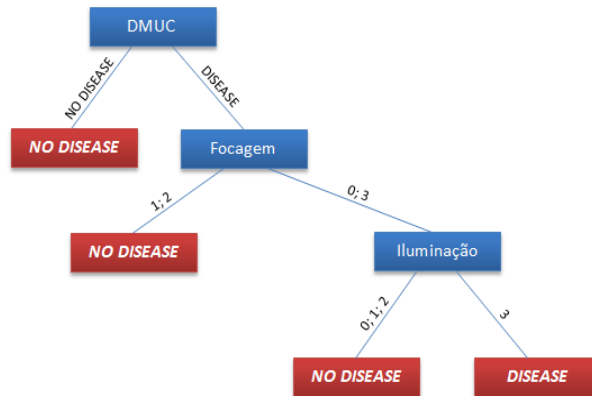


Figura A.1.11 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

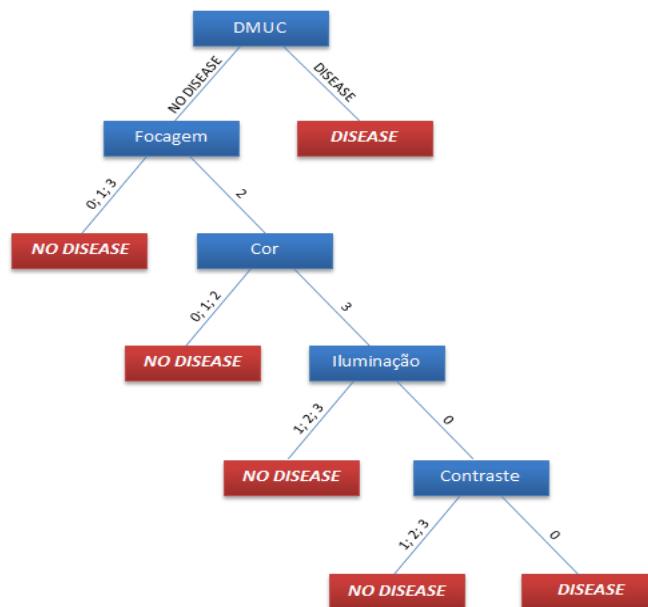


Figura A.1.12 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

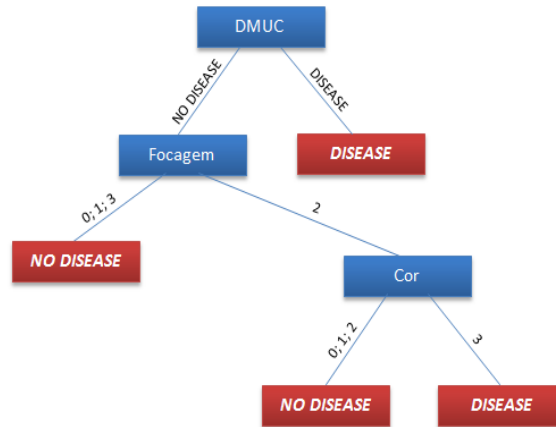


Figura A.1.13 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=8 e FP =1

- **Retmarker + DMUC + Qualidade Geral**

**i. Custos FN = 1; Custos FP=1**

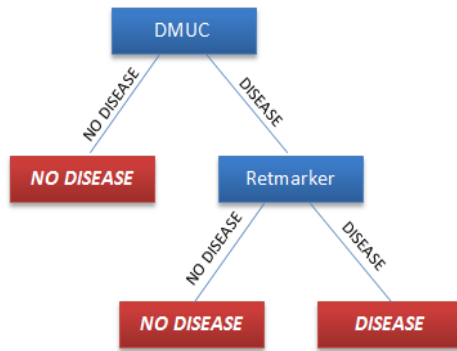


Figura A.1.14 -árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

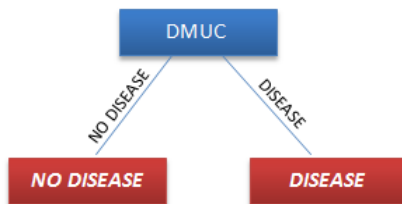


Figura A.1.15 -árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

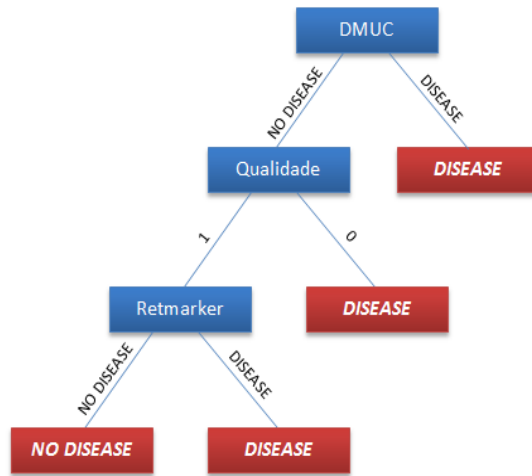


Figura A.1.16 -árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=8 e FP =1

- **Retmarker + DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

i. Custos FN = 1; Custos FP=1

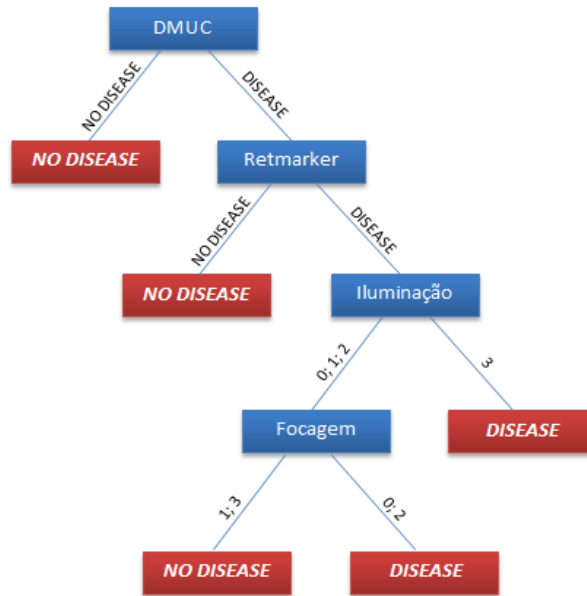


Figura A.1.17 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=1 e FP =1

ii. Custos FN = 4; Custos FP=1

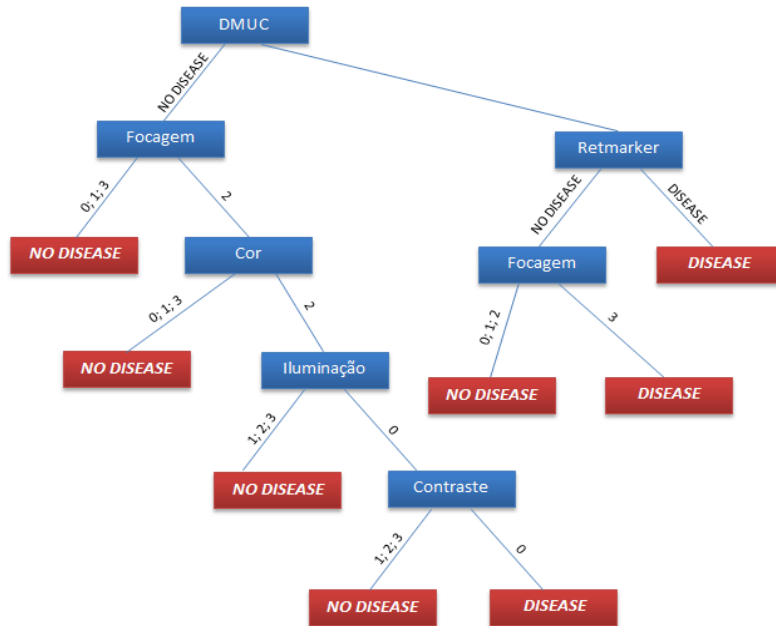


Figura A.1.18 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

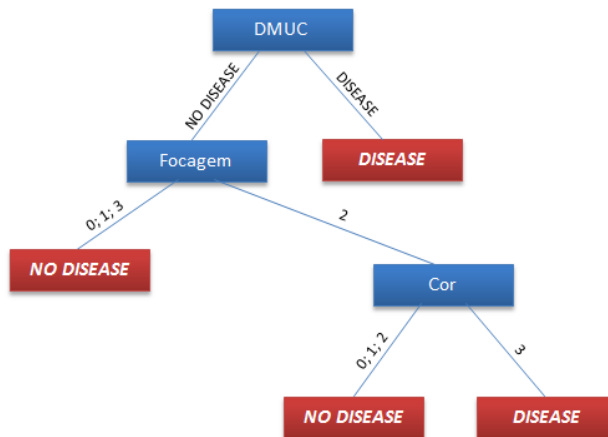


Figura A.1.19 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=8 e FP =1

- **Retmarker + DMUC; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

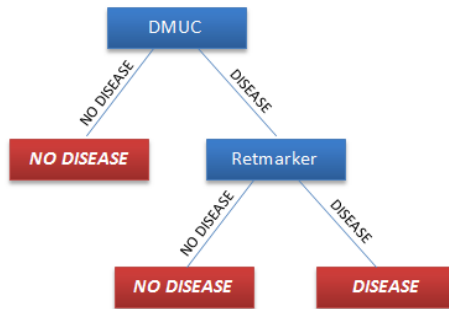


Figura A.1.20 -árvore de decisão resultante da fusão: Retmarker + DMUC desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

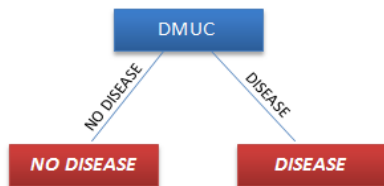


Figura A.1.21 -árvore de decisão resultante da fusão: Retmarker + DMUC desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1

**iii. Custos FN = 8; Custos FP=1**

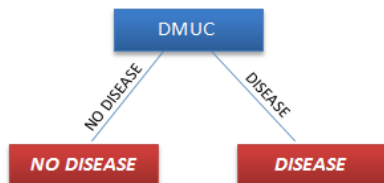


Figura A.1.22 -árvore de decisão resultante da fusão: Retmarker + DMUC desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1

- **Retmarker + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 4; Custos FP=1**

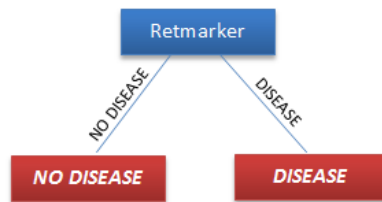


Figura A.1.23 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP=1

**ii. Custos FN = 8; Custos FP=1**

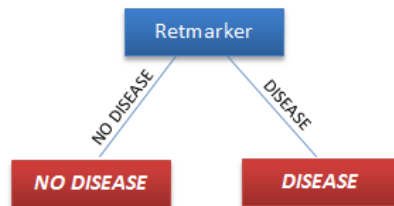


Figura A.1.24 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP=1

- **DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

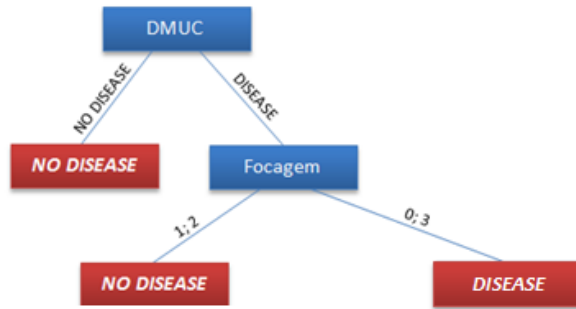


Figura A.1.25 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

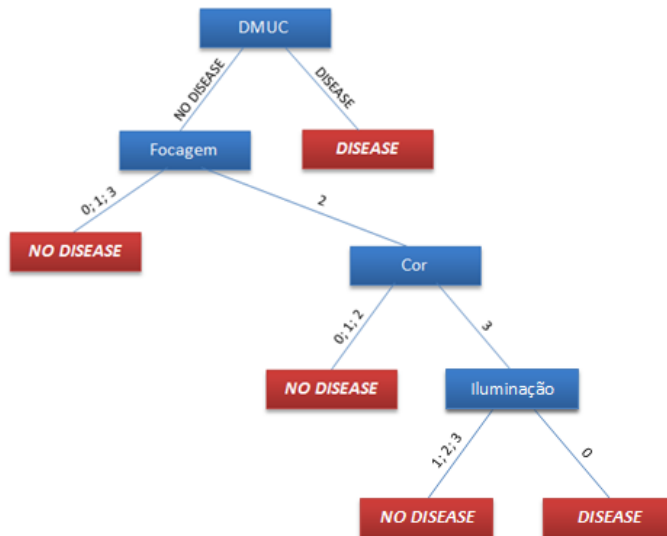


Figura A.1.26 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

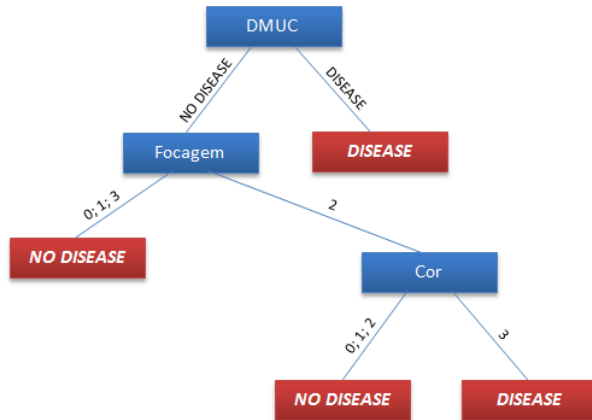


Figura A.1.27 -árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1

- **Retmarker + DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

Devido às grandes dimensões da árvore de decisão resultante desta fusão procedeu-se à sua partição, de maneira a ser aqui apresentada:

Se DMUC = *NO DISEASE*:

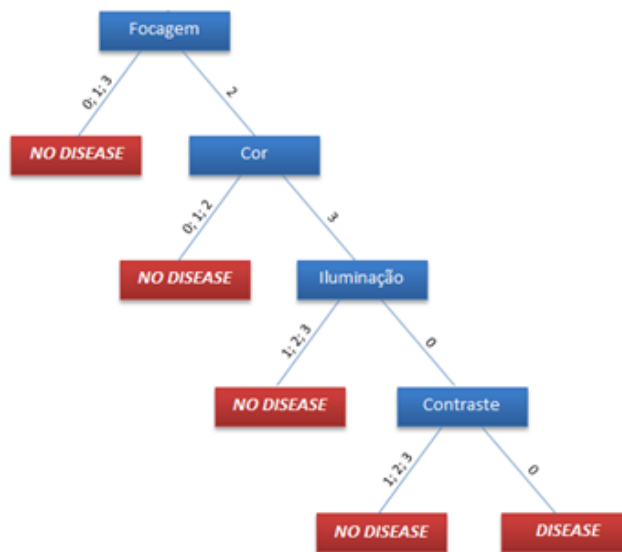


Figura A.1.28 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1; Parte 1

Se  $DMUC = DISEASE$  e  $Retmarker = NO DISEASE$ :

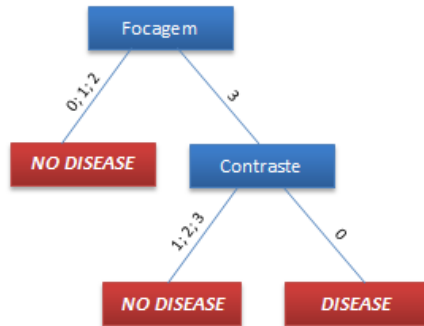


Figura A.1.29 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1; Parte 2

Se  $DMUC = DISEASE$  e  $Retmarker = DISEASE$ :

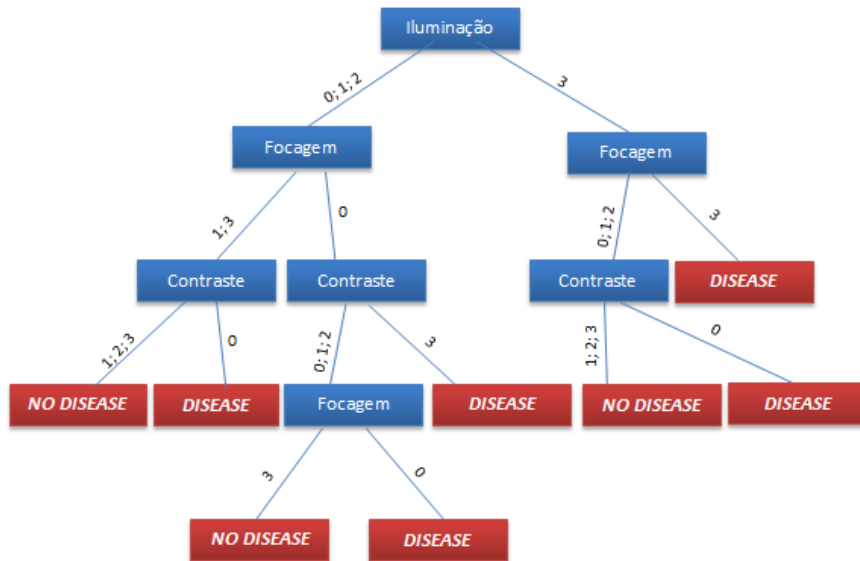


Figura A.1.30 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1; Parte 3

ii. Custos FN = 4; Custos FP=1

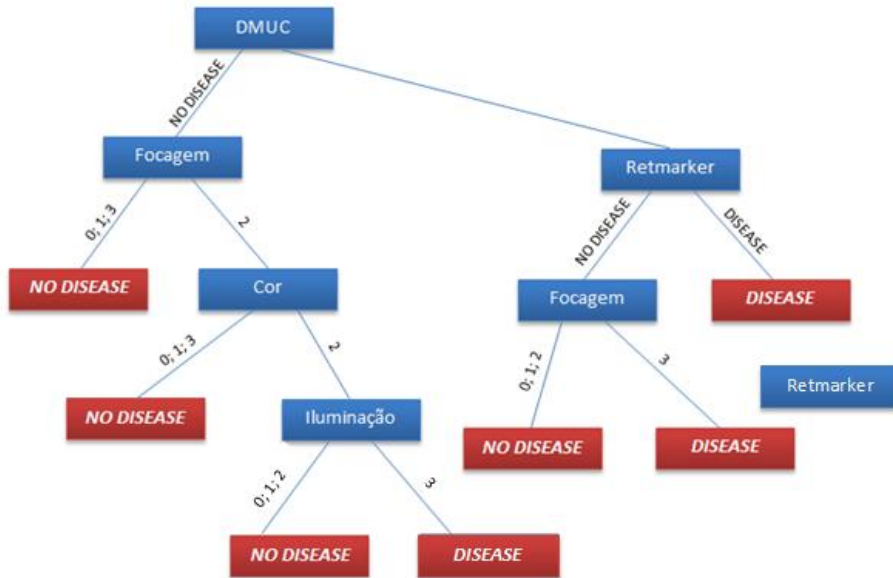


Figura A.1.31 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP=1

iii. Custos FN = 8; Custos FP=1

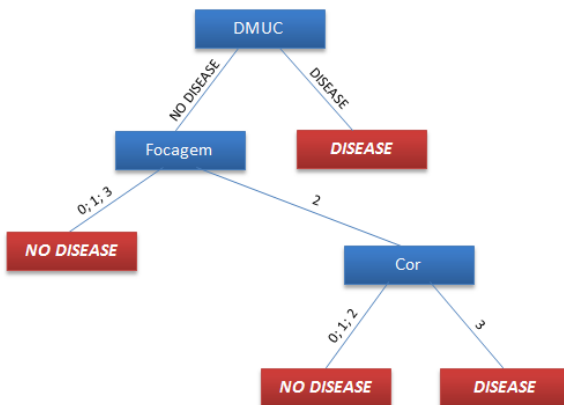


Figura A.1.32 -árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP=1

## A.2 Base de Dados nº 2

- **Retmarker + DMUC**

**i. Custos FN = 1; Custos FP=1**

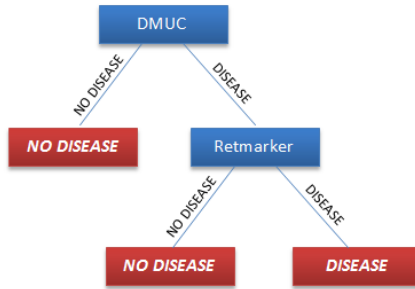


Figura A.2.1 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

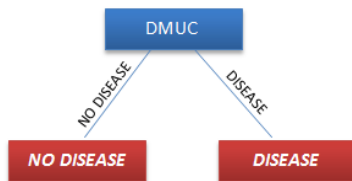


Figura A.2.2 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=4 e FP =1

**iii. Custos FN = 8; Custos FP=1**

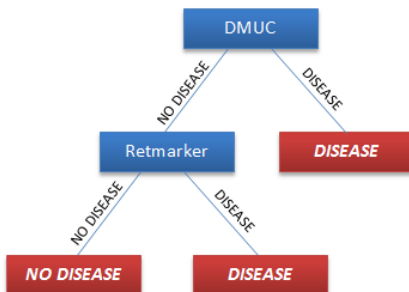


Figura A.2.3 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=8 e FP =1

• **Retmarker + Qualidade Geral**

i. **Custos FN = 1; Custos FP=1**

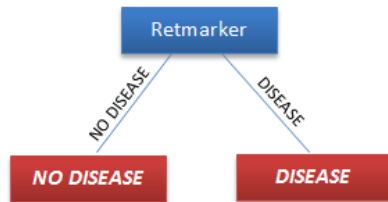


Figura A.2.4 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=1 e FP =1

ii. **Custos FN = 4; Custos FP=1**

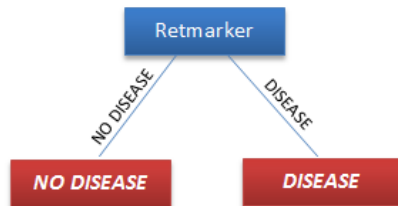


Figura A.2.5 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=4 e FP =1

iii. **Custos FN = 8; Custos FP=1**

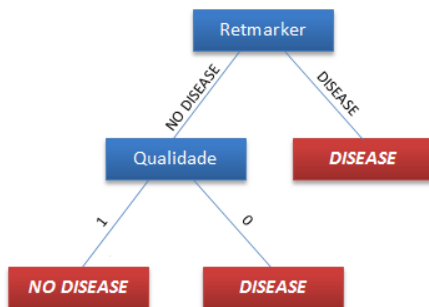


Figura A.2.6 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=8 e FP =1

- **Retmarker + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

i. **Custos FN = 1; Custos FP=1**

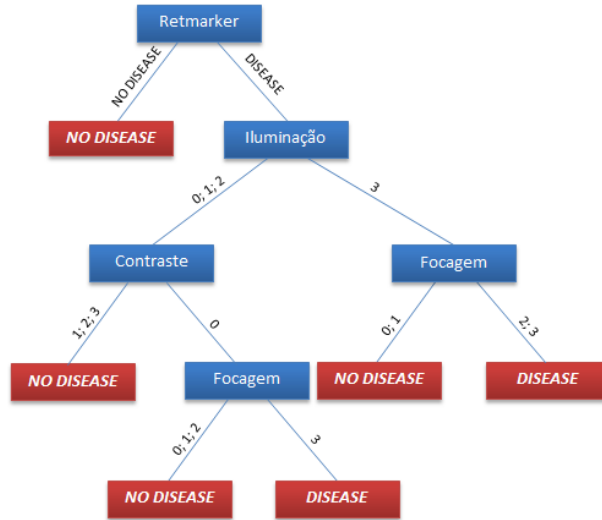


Figura A.2.7 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=1 e FP =1

ii. **Custos FN = 4; Custos FP=1**

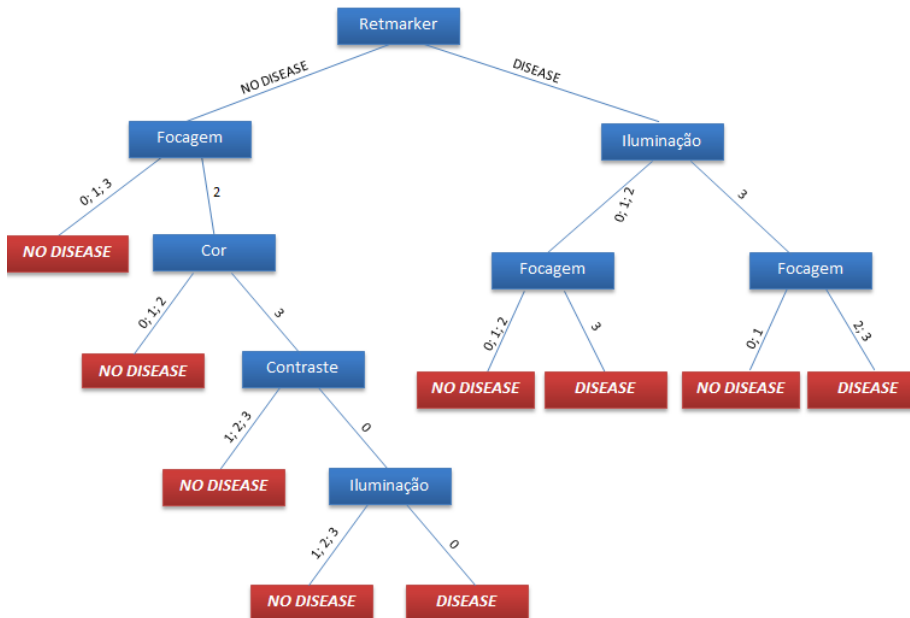


Figura A.2.8 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

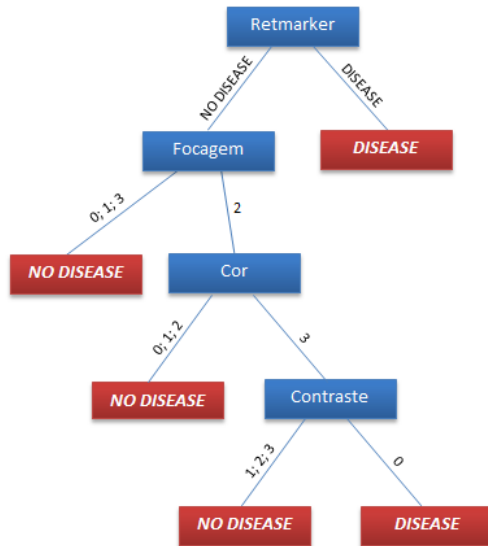


Figura A.2.9 -árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=8 e FP =1

- **DMUC + Qualidade Geral**

**i. Custos FN = 1; Custos FP=1**

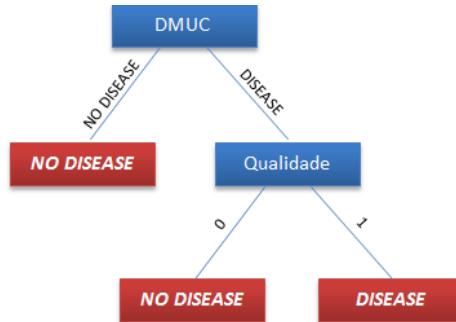


Figura A.2.10 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

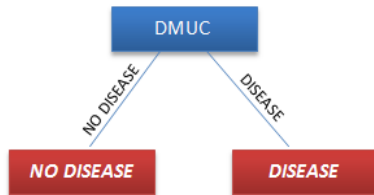


Figura A.2.11 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=4 e FP =1

**iii. Custos FN = 8; Custos FP=1**

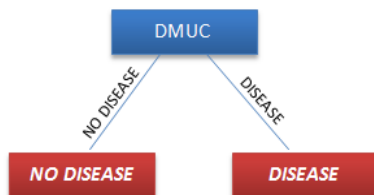


Figura A.2.12 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=8 e FP =1

- **DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

**i. Custos FN = 1; Custos FP=1**

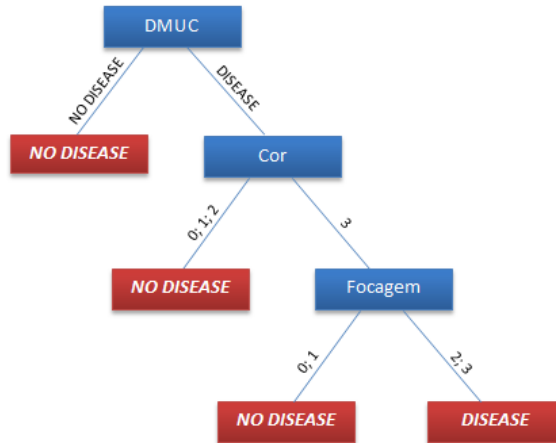


Figura A.2.13 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

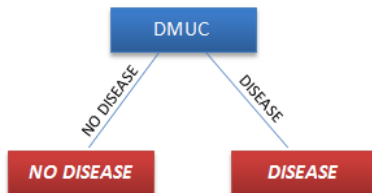


Figura A.2.14 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

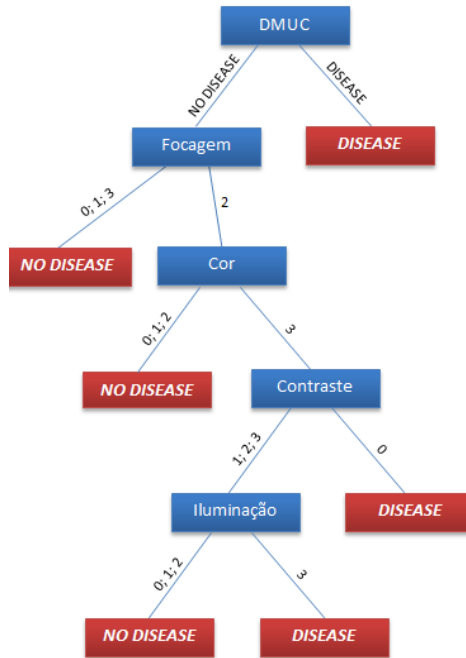


Figura A.2.15 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=8 e FP =1

- **Retmarker + DMUC + Qualidade Geral**

**i. Custos FN = 1; Custos FP=1**

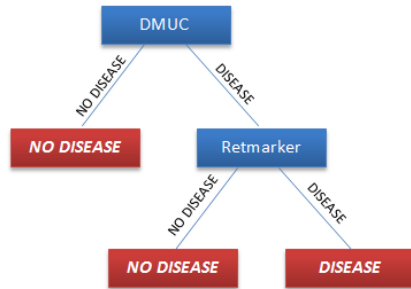


Figura A.2.16 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

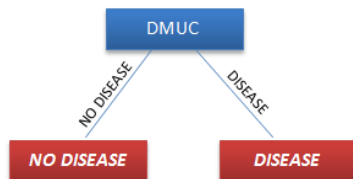


Figura A.2.17 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

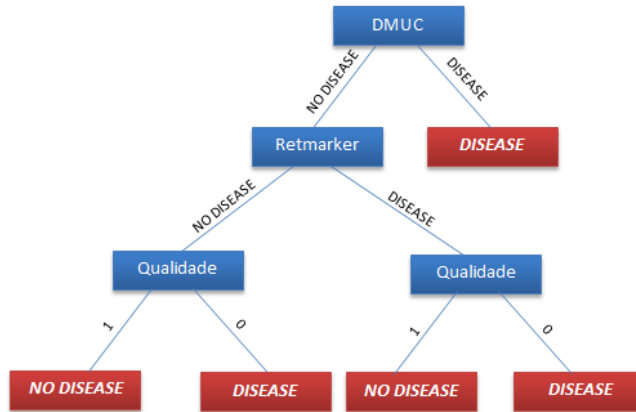


Figura A.2.18 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=8 e FP =1

- **Retmarker + DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

**i. Custos FN = 1; Custos FP=1**

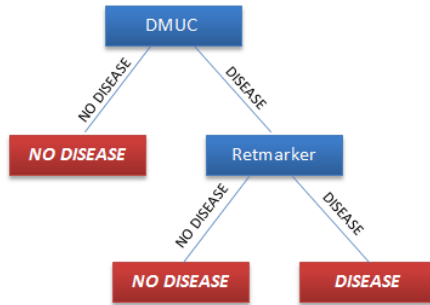


Figura A.2.19 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

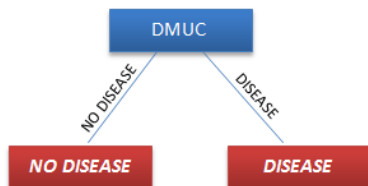


Figura A.2.20 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

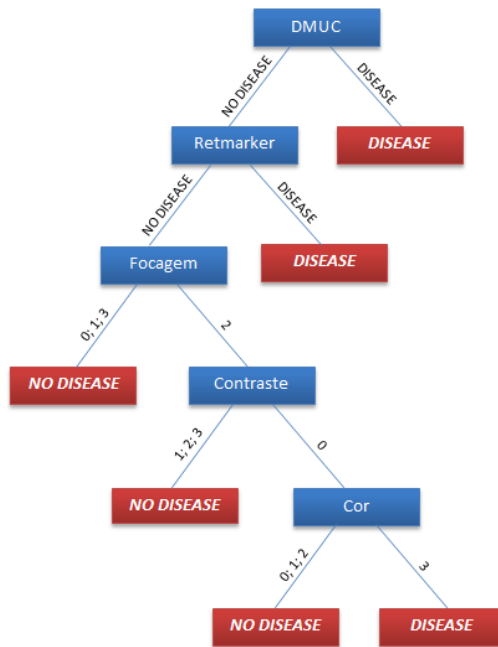


Figura A.2.21 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=8 e FP =1

- **Retmarker + DMUC; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

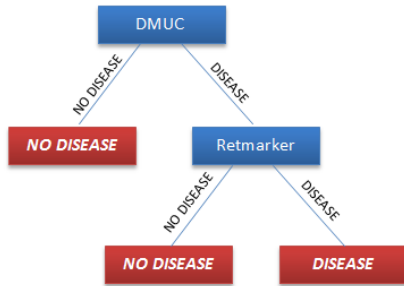


Figura A.2.22 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

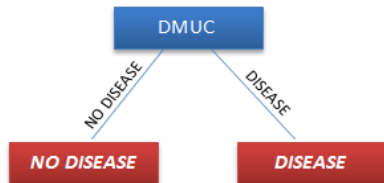


Figura A.2.23 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1

**iii. Custos FN = 8; Custos FP=1**

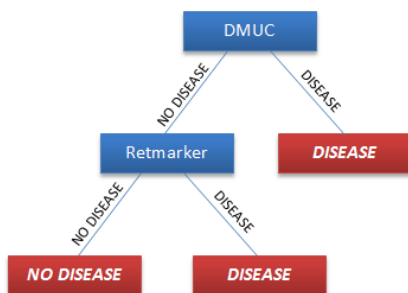


Figura A.2.24 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1

- **Retmarker + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

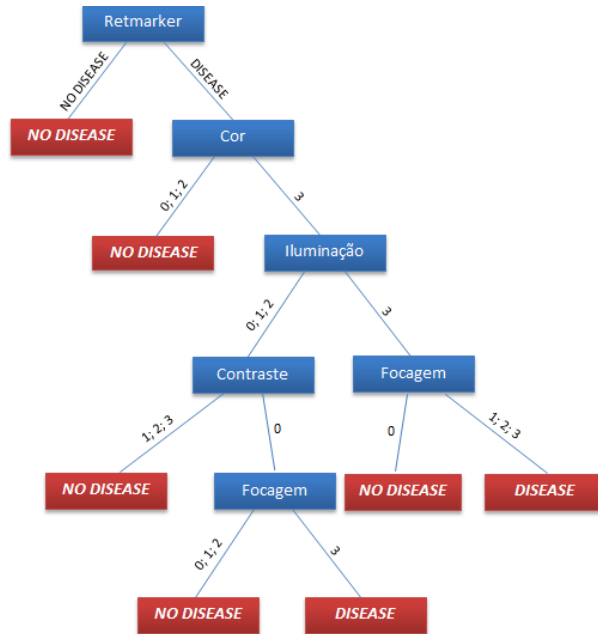


Figura A.2.25 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP=1

**ii. Custos FN = 4; Custos FP=**

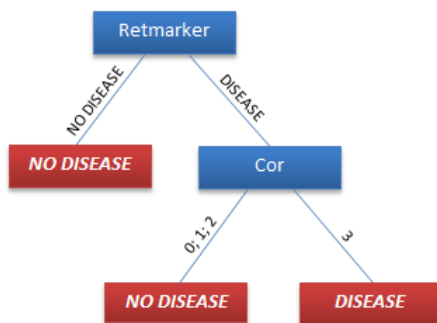


Figura A.2.26 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP=1

iii. Custos FN = 8; Custos FP=1

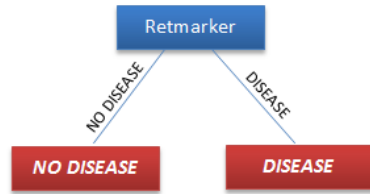


Figura A.2.27 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1

- **DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

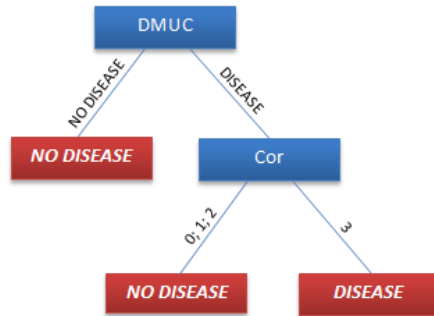


Figura A.2.28 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP=1

**ii. Custos FN = 4; Custos FP=1**

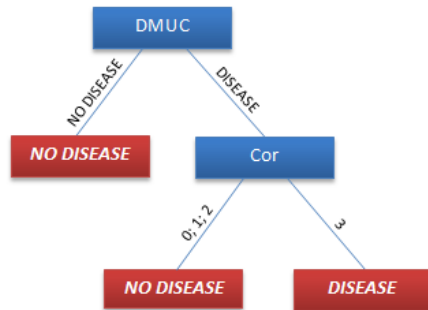


Figura A.2.29 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP=1

iii. Custos FN = 8; Custos FP=1

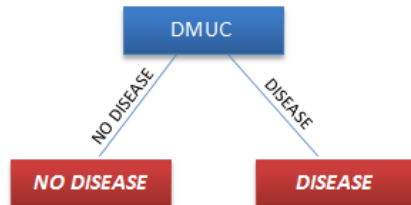


Figura A.2.30 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP=1

- **Retmarker + DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

i. **Custos FN = 1; Custos FP=1**

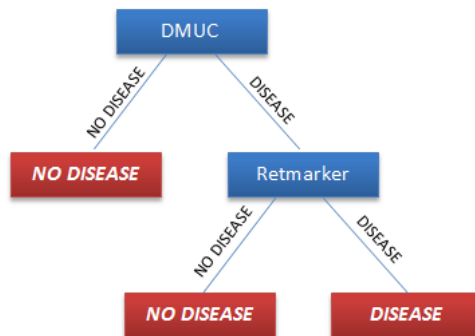


Figura A.2.31 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1

ii. **Custos FN = 4; Custos FP=1**

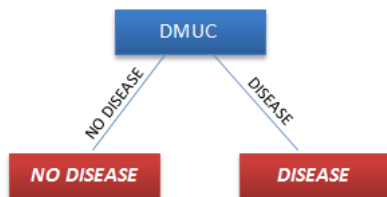


Figura A.2.32 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

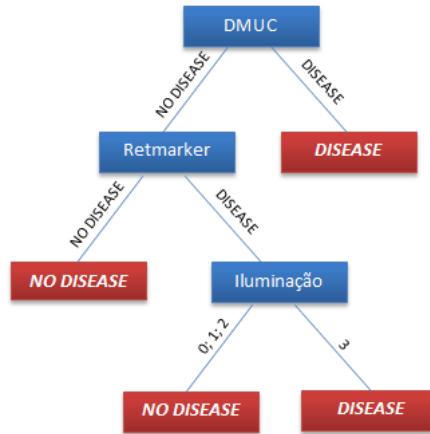


Figura A.2.33 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1

### A.3 Base de Dados nº 3

- Retmarker + DMUC

i. Custos FN = 1; Custos FP=1

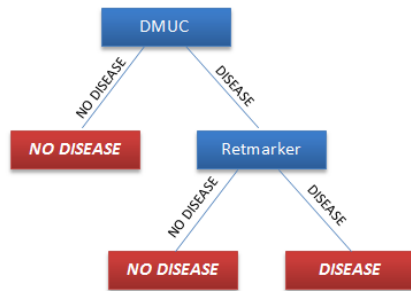


Figura A.3.1 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=1 e FP =1

ii. Custos FN = 4; Custos FP=1

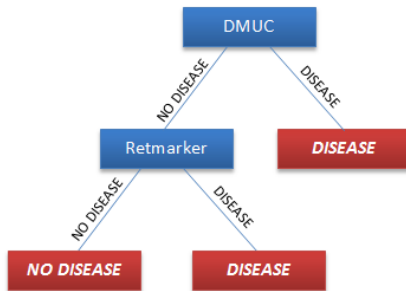


Figura A.3.2 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

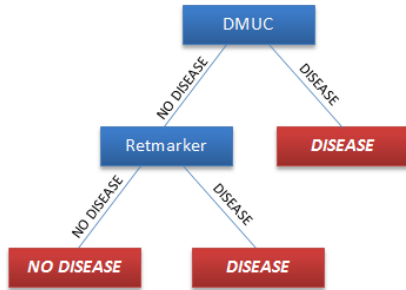


Figura A.3.3 - árvore de decisão resultante da fusão: Retmarker + DMUC, para custos FN=8 e FP =1

- **Retmarker + Qualidade Geral**

**i. Custos FN = 1; Custos FP=1**

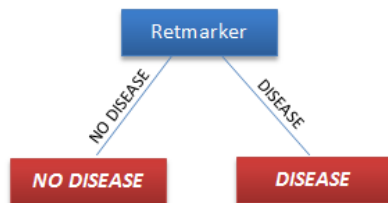


Figura A.3.4 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

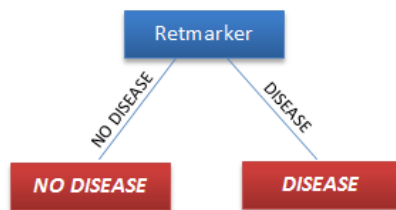


Figura A.3.5 - árvore de decisão resultante da fusão: Retmarker + Qualidade Geral, para custos FN=4 e FP =1

- **Retmarker + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

**i. Custos FN = 1; Custos FP=1**

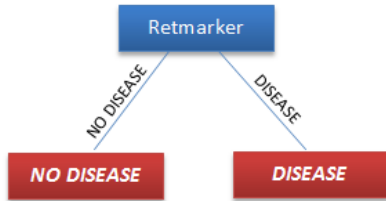


Figura A.3.6 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

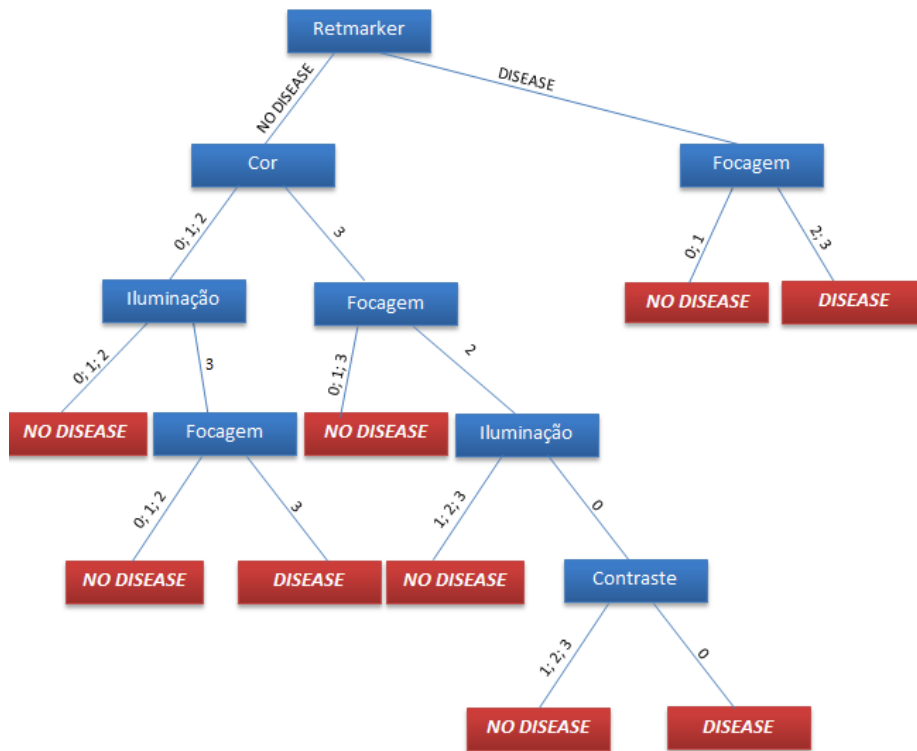


Figura A.3.7 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

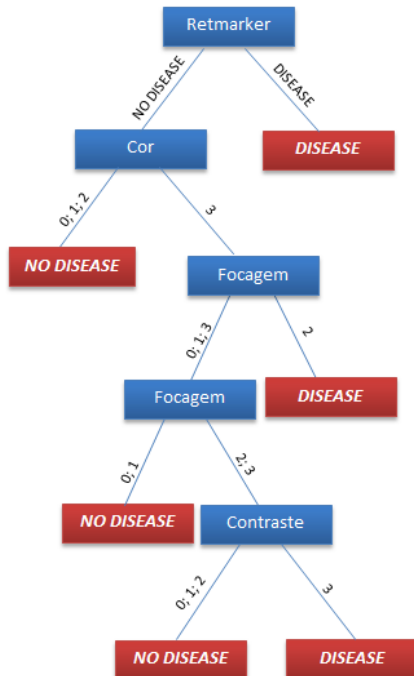


Figura A.3.8 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, para custos FN=8 e FP =1

- **DMUC + Qualidade Geral**

**i. Custos FN = 1; Custos FP=1**

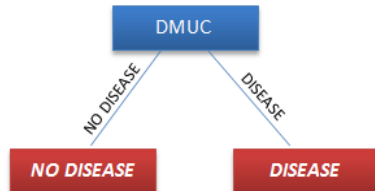


Figura A.3.9 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

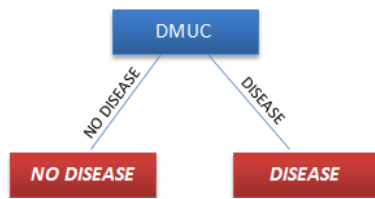


Figura A.3.10 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=4 e FP =1

**iii. Custos FN = 8; Custos FP=1**

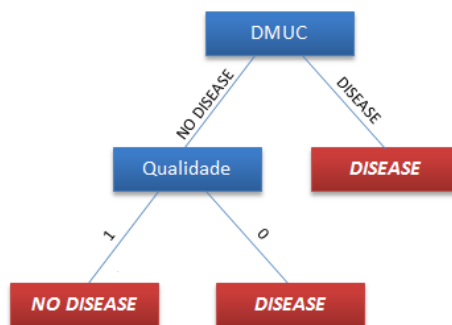


Figura A.3.11 - árvore de decisão resultante da fusão: DMUC + Qualidade Geral, para custos FN=8 e FP =1

- **DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

**i. Custos FN = 1; Custos FP=1**

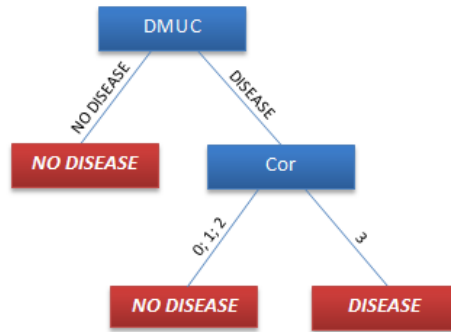


Figura A.3.12 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

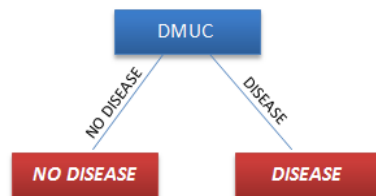


Figura A.3.13 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

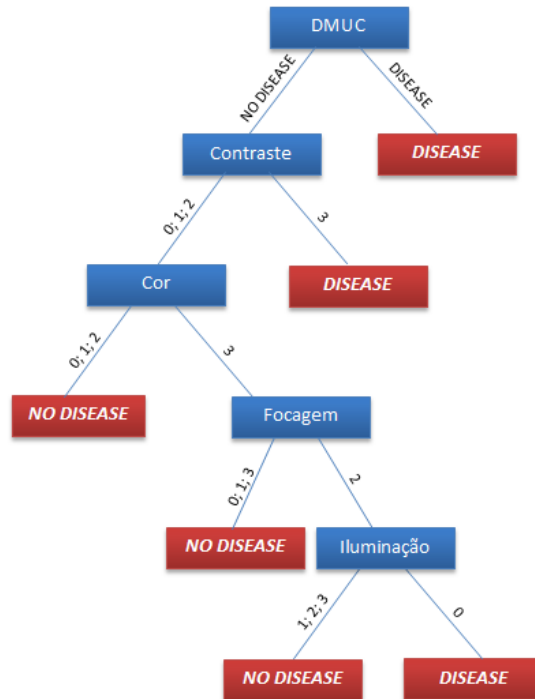


Figura A.3.14 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, para custos FN=8 e FP =1

- **Retmarker + DMUC + Qualidade Geral**

- i. **Custos FN = 1; Custos FP=1**

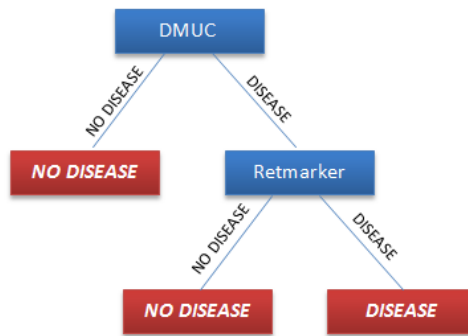


Figura A.3.15 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=1 e FP =1

- ii. **Custos FN = 4; Custos FP=1**

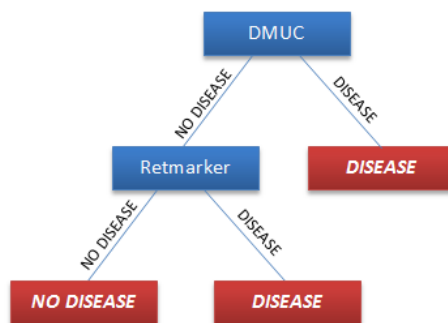


Figura A.3.16 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

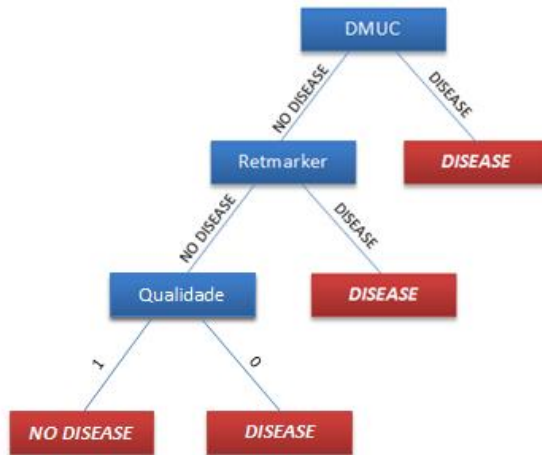


Figura A.3.17 - árvore de decisão resultante da fusão: Retmarker + DMUC + Qualidade Geral, para custos FN=8 e FP =1

- **Retmarker + DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste**

**i. Custos FN = 1; Custos FP=1**

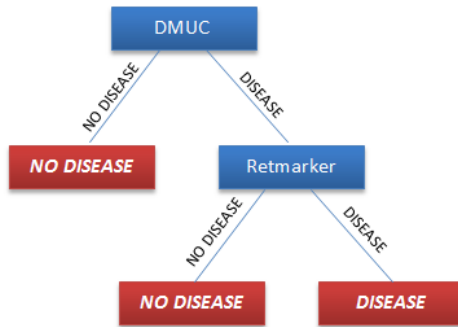


Figura A.3.18 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

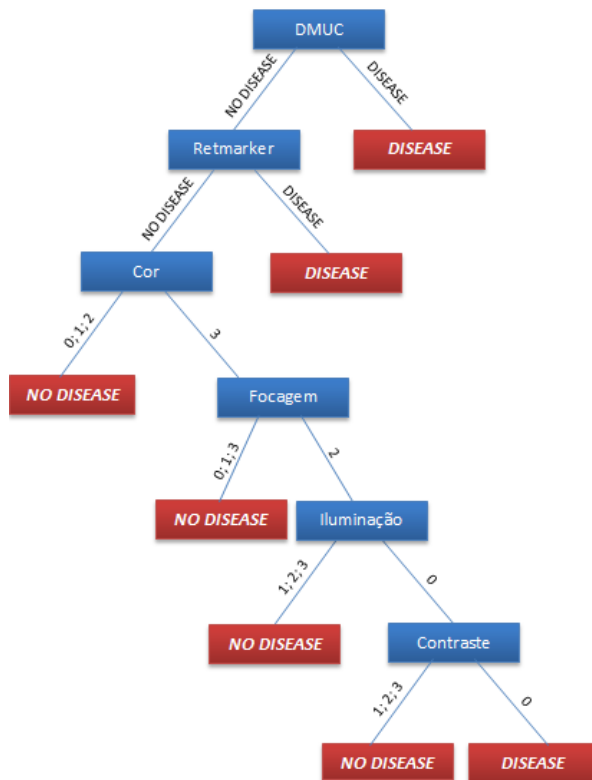


Figura A.3.19 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

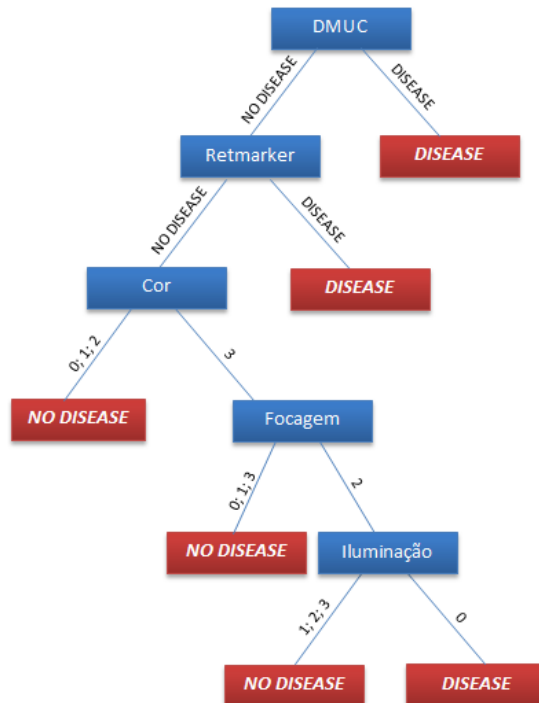


Figura A.3.20 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, para custos FN=8 e FP =1

- **Retmarker + DMUC; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

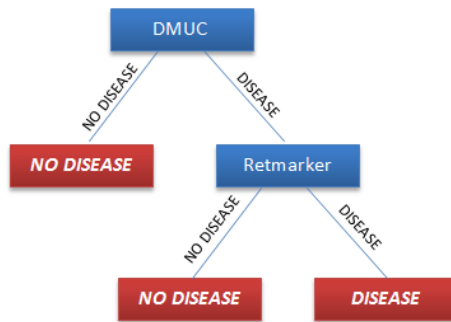


Figura A.3.21 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

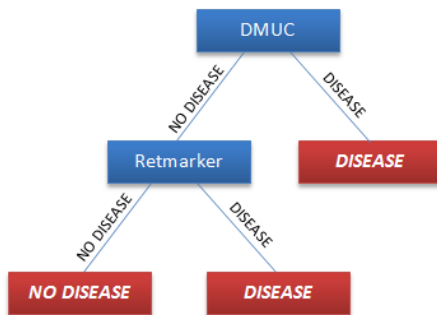


Figura A.3.22 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

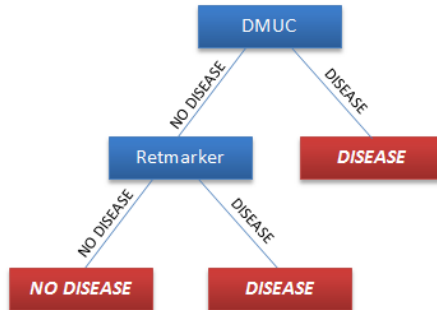


Figura A.3.23 - árvore de decisão resultante da fusão: Retmarker + DMUC, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1

- **Retmarker + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

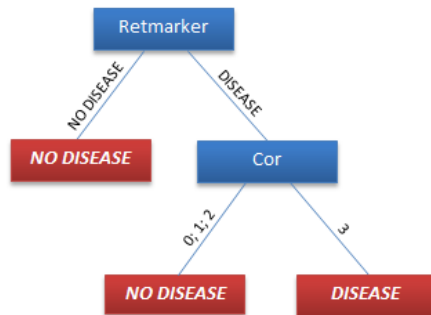


Figura A.3.24 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP=1

**ii. Custos FN = 4; Custos FP=1**

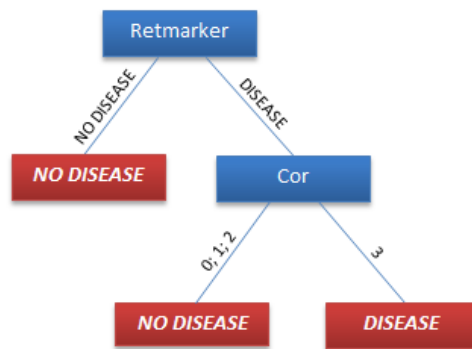


Figura A.3.25 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP=1

iii. Custos FN = 8; Custos FP=1

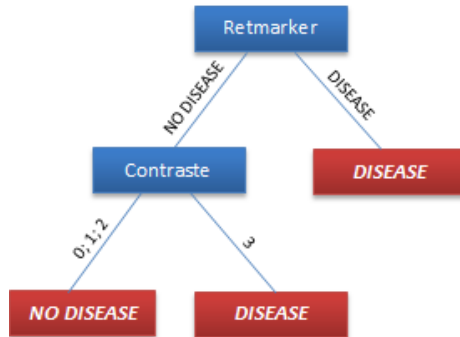


Figura A.3.26 - árvore de decisão resultante da fusão: Retmarker + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP=1

- **DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

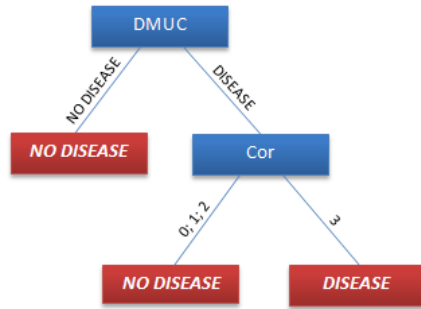


Figura A.3.27 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP=1

**ii. Custos FN = 4; Custos FP=1**

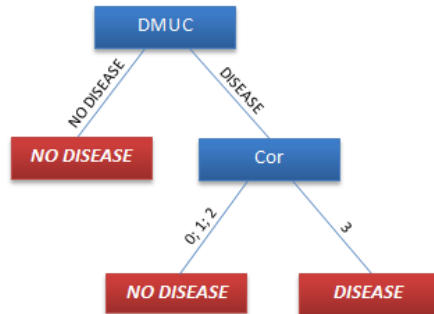


Figura A.3.28 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP=1

iii. Custos FN = 8; Custos FP=1

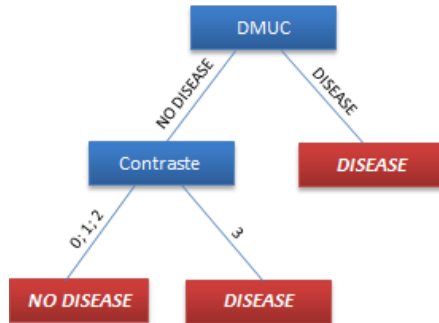


Figura A.3.29 - árvore de decisão resultante da fusão: DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP=1

- **Retmarker + DMUC + 4 características da Qualidade: Cor, Focagem, Iluminação e Contraste; desconsiderando imagens de qualidade inadequada**

**i. Custos FN = 1; Custos FP=1**

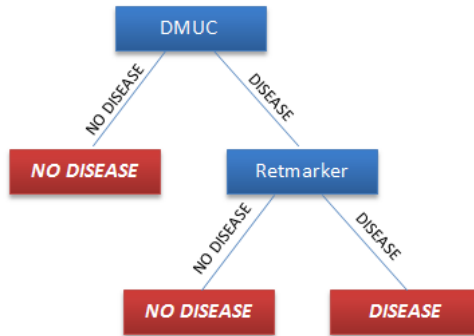


Figura A.3.30 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=1 e FP =1

**ii. Custos FN = 4; Custos FP=1**

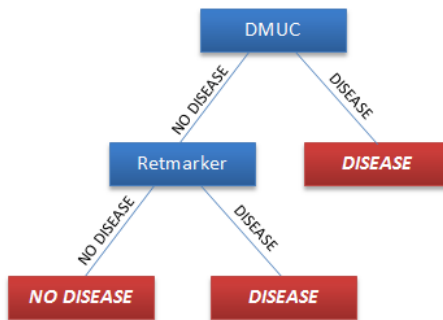


Figura A.3.31 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=4 e FP =1

iii. Custos FN = 8; Custos FP=1

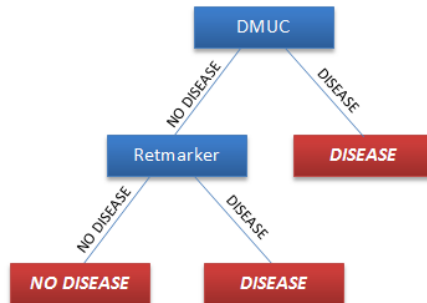


Figura A.3.32 - árvore de decisão resultante da fusão: Retmarker + DMUC + 4 características da Qualidade, desconsiderando imagens de qualidade inadequada, para custos FN=8 e FP =1