

VALIDAÇÃO DO SOFTWARE RetmarkerAMD

João Pires¹, Inês Láins², Dalila Alves³, Sandrina Nunes³,

Rufino Silva^{1, 3, 4}, João Pedro Marques^{1, 3, 4}

1. Faculdade de Medicina da Universidade de Coimbra, 3004-504 Coimbra, Portugal
2. Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA
3. Association for Innovation and Biomedical Research on Light and Image (AIBILI),
3000-548 Coimbra, Portugal
4. Centro Hospitalar e Universitário de Coimbra (CHUC), 3000-075 Coimbra, Portugal

Correspondência deve ser enviada a João Pires: jpires010@hotmail.com

Abstract

Purpose

The aim of this study is to validate RetmarkerAMD (Retmarker SA) as a semiautomatic grading software to detect early and late fundoscopic features of age-related macular degeneration (AMD).

Materials and Methods

Multicenter, cross-sectional study, including both control and AMD patients from the medical retina units of Centro Hospitalar e Universitário de Coimbra (CHUC) and Massachusetts Eye and Ear Infirmary (Boston, MA, USA). Two hundred and two color fundus photographs (CFP) previously graded by a senior retina specialist (RS) using a validated software (Topcon IMAGENet 2000) were used in this study. Two graders (JP and JPM), both certified by the Coimbra Ophthalmology Reading Centre (CORC), independently classified the 202 CFP images using RetmarkerAMD and Topcon IMAGENet 2000. To improve the detection of the AMD features, a certified automated software was used to standardize brightness, contrast and color balance. Both intra and intergrader agreement was evaluated. For the Topcon IMAGENet 2000, the graders' classification was also compared to a senior retina specialist grading, defined as goldstandard. Agreement was assessed using the percentage of agreement (i.e. number of images with the same classification) and the weighted Kappa coefficient considering linear weights.

Results

The inter-grader analysis (JPM vs JP) for all the features analyzed with RetmarkerAMD revealed an average agreement of 86,2%. For the AMD stage, a 93% agreement was observed (Kappa=0.95, $p<0.001$). Using Topcon IMAGEnet 2000, the inter-grader analysis showed an almost perfect agreement in grading AMD stage (90.1%; Kappa=0.87, $p<0.001$). Both graders also showed an almost perfect agreement with the goldstandard (90.1%; Kappa=0.88, $p<0.001$ for JPM and 87.1%; Kappa=0.86, $p<0.001$ for JP). Regarding the intra-grader analysis between RetmarkerAMD and Topcon IMAGEnet 2000, a 76,8% agreement (Kappa = 0.73, $p<0.001$) and a 70,8% agreement (Kappa = 0.67, $p<0.001$, was reached for JPM and JP, respectively.

Conclusion

The use of RetmakerAMD was shown to be reliable, reproducible and accurate for the grading of AMD features. This semiautomated software may thus be useful for the longitudinal monitoring of AMD features and contribute to identify early imaging biomarkers of the advanced forms of the disease.

Key words: Age-related macular degeneration; grading; color fundus photography, teleophthalmology

Resumo

Objetivo

Este trabalho tem como objetivo a validação do *RetmarkerAMD* como um *software* de *grading* semi-automático para deteção de alterações precoces e tardias da degenerescência macular da Idade (DMI).

Materiais e Métodos

Desenhou-se um estudo multicêntrico e transversal, incluindo indivíduos com e sem DMI (controlos), oriundos dos serviços de Retina Médica do Centro Hospitalar e Universitário de Coimbra (CHUC) e do *Massachusetts Eye and Ear Infirmary (Boston, MA, USA)*. Foram usadas 202 retinografias anteriormente classificadas por um especialista sénior da retina (RS), utilizando um *software* validado (*Topcon IMAGEnet 2000*). Dois *graders* (JP e JPM), certificados pelo *Coimbra Ophthalmology Reading Centre (CORC)*, classificaram as 202 imagens usando o *RetmarkerAMD* e o *Topcon IMAGEnet 2000*. Para maior facilidade na deteção das lesões, foi utilizado um *software* certificado que automaticamente otimiza a cor, brilho e contraste. Procedeu-se a uma análise de concordância intra e *intergrader*. No *Topcon IMAGEnet 2000*, a análise de ambos os *graders* foi comparada com a análise de um especialista sénior de retina, definida como *goldstandard*. A concordância foi avaliada utilizando a percentagem de concordância (i.e. número de imagens com a mesma classificação) e o coeficiente *Kappa* tendo em conta ponderações lineares.

Resultados

A análise *inter-grader* (JPM vs JP), para todas as alterações estudadas, utilizando o *RetmarkerAMD*, revelou uma concordância média de 86,2%. Para o estadió de DMI, a concordância foi de 93% ($Kappa=0.95$, $p<0.001$). Utilizando o *Topcon IMAGEnet 2000*, a análise *inter-grader* mostrou uma concordância *almost perfect* para o estadió de DMI (90.1%; $Kappa=0.87$, $p<0.001$). Ambos os *graders* mostraram uma concordância *almost perfect* quando comparadas as suas análises com o *goldstandard* (90.1%; $Kappa=0.88$, $p<0.001$ para JPM e 87.1%; $Kappa=0.86$, $p<0.001$ para JP). Em relação à análise *intra-grader* entre *RetmarkerAMD* e *Topcon IMAGEnet 2000*, foi obtida uma concordância de 76,8% ($Kappa = 0.73$, $p<0.001$) e 70,8% ($Kappa = 0.67$, $p<0.001$) para JPM e JP, respetivamente.

Conclusão

A utilização do *RetmarkerAMD* mostrou ser confiável, reprodutível e precisa no que respeita ao *grading* das lesões típicas de DMI. Este *software* semi-automático pode assim ser útil na monitorização longitudinal destes doentes e contribuir para identificar biomarcadores imagiológicos precoces de formas avançadas da doença.

Palavras-chave: Degenerescência macular da idade; grading; retinografia; teleoftalmologia

Introdução

A Degenerescência Macular da Idade (DMI) é uma doença crónica e progressiva (1) que afeta a região central da retina - a mácula. Na sua forma avançada, a doença tem um carácter extremamente lesivo em termos funcionais, sendo a principal causa de cegueira irreversível em doentes ≥ 65 anos, em países desenvolvidos (2-5). A prevalência da DMI aumenta exponencialmente com o avançar da idade (3,5), passando de 10% entre os 65 e 75 anos a cerca de 30% a partir dos 75 anos (6). As previsões indicam que em 2040, 288 milhões de doentes em todo o mundo sofrerão desta patologia, tornando-a numa verdadeira epidemia à escala global (3).

O *Beaver Dam Eye Study*, um estudo de larga escala na população desta cidade norte-americana, estimou uma prevalência de 14,1% nas formas precoces e 1,6% nas formas tardias [1,2% de lesões neovasculares e 0,6% de lesões não-neovasculares - atrofia Geográfica (AG)] (4). Este estudo comprovou que a incidência e a progressão da DMI estão intimamente relacionadas com a idade: num período de 5 anos, indivíduos > 75 anos têm uma probabilidade ~ 10 vezes superior de desenvolver formas precoces da doença, do que pessoas com idades entre os 43 e 54 anos (5).

Em Portugal, o *Coimbra Eye Study* foi o primeiro estudo epidemiológico da DMI na população portuguesa (2,7). Verificou-se uma prevalência de 11,22% para formas precoces e 1,16% (0,57% para formas neovasculares e 0,61% para a AG) para formas tardias da doença (2).

Sabe-se hoje que a DMI é uma doença multifatorial com uma complexa interação entre fatores genéticos e ambientais e que esta interação é importante tanto no desenvolvimento como na progressão da doença (8,9). A nível ocular, a presença de alterações pigmentares, o número,

a área e a localização dos *drusens*, constituem fatores de risco conhecidos para a progressão para formas tardias (10). O estudo *AREDS (Age-Related Eye Disease Study)* foi um estudo prospetivo, multicêntrico, onde participaram 4757 doentes com DMI. Os resultados deste estudo permitiram o desenvolvimento de uma classificação em 4 estadios, tendo por base as alterações fundoscópicas encontradas: sem DMI; DMI precoce; DMI intermédia; DMI avançada. (*vide* Material e Métodos, Tabela 2) (11). Além disso, levou à criação de um modelo de predição do risco de desenvolvimento de formas tardias da doença, através de um *point scoring system* (*vide* Tabela 1) (12) este modelo permite predizer o risco de progressão a 5 anos para estadios avançados de DMI (*vide* Tabela 1) (12).

Tabela 1- *Point scoring system*, de acordo com o *AREDS*

Alterações fundoscópicas	Point score
<i>Drusens</i> intermédios (63 a 125 µm) em ambos os olhos	1 ponto
<i>Drusens</i> moles (≥ 125 µm) por cada olho	1 ponto
Alterações pigmentares por cada olho	1 ponto
DMI neovascular em 1 olho	2 pontos
Fatores de risco	Risco de progressão a 5 anos para estadios avançados de DMI (%)
0	0,5
1	3
2	12
3	25
4	50

Estabelecida a importância das alterações fundoscópicas na progressão para formas avançadas de DMI, vários *reading centres* iniciaram a classificação de retinografias de doentes com DMI. Este exame complementar de diagnóstico permite não só a documentação das lesões para efeitos de estadiamento, como a monitorização da sua evolução ao longo do tempo. Contudo, além de morosa, a classificação manual das imagens implica um grande investimento em termos humanos e financeiros, já que é 100% dependente do *grader*. A utilização e globalização de sistemas de classificação automáticos e semiautomáticos para identificação de doentes com DMI em risco de progressão é, por isso, altamente ambicionada. Quando comparados com métodos de análise manual, os sistemas de classificação automáticos e semi-automáticos permitem uma análise mais eficiente, com uma grande redução a nível de tempo despendido, recursos humanos e, em última análise, recursos financeiros (13,14).

O *RetmarkerAMD* é um *software* semiautomático desenvolvido para auxiliar no *grading* de imagens de doentes com DMI. Vários estudos piloto (15) atestaram uma boa sensibilidade e especificidade quando comparado com o *grading* manual, permitindo reduzir o custo económico, o tempo consumido (35% mais rápido), bem como uma maior identificação de lesões (mais 32%), com consequente redução do erro humano (15,16). O objetivo deste trabalho é a validação do *software RetmarkerAMD* como ferramenta de *grading* na DMI, comparando-o com um sistema já validado e amplamente utilizado – *Topcon IMAGEnet 2000*.

Materiais e Métodos

1. Tipo de estudo

Desenhou-se um estudo multicêntrico e transversal, para o qual foram selecionadas aleatoriamente 202 retinografias a partir de uma amostra de 372 doentes com ou sem DMI. As retinografias eram provenientes de doentes seguidos em Consulta de Retina Médica no Centro Hospitalar e Universitário de Coimbra (CHUC) ou no *Massachusetts Eye and Ear Infirmary (MEEI)*, em Boston, que consentiram a utilização das suas imagens para trabalhos de investigação. O cálculo do tamanho da amostra foi feito para um poder estatístico de 0,8 e uma significância de 0,05. Consideraram-se como critérios de inclusão: 1) doentes com >50 anos de idade sem doença oftalmológica ou 2) doentes com >50 anos de idade e diagnóstico de DMI precoce, intermédia ou tardia. Foram excluídos doentes com outras patologias oftalmológicas além de DMI ou doentes com opacidade significativa dos meios óticos, impossibilitando o *grading* das retinografias.

Para efeitos de validação do *software RetmarkerAMD*, procedeu-se a um *grading* independente efetuado por dois *graders* certificados (JPM e JP) pelo *Coimbra Ophthalmology Reading Center (CORC)*. Ambos os *graders* efetuaram a classificação das retinografias no *software RetmarkerAMD* e *Topcon ImageNet 2000*. Procedeu-se a uma análise de concordância *intra-* e *intergrader*. Salienta-se ainda que as medições efetuadas no *RetmakerAMD* e no *Topcon IMAGEnet 2000* tiveram uma diferença temporal de 2 meses. A classificação obtida no *software Topcon IMAGEnet 2000* foi comparada com uma classificação já realizada por 2 *graders* certificados pela mesma instituição e validada por um especialista sénior de retina (RS), tendo sido definida como *goldstandard*.

2. Certificação

Os *graders* foram previamente treinados, tendo obtido creditação e certificação por parte do *CORC*. Para isso, estabeleceu-se a necessidade de um valor de *Kappa* > 0.6, como sugerido por *Landis JR et al.* Sendo o valor de *Kappa* obtido a partir da análise estatística da concordância entre *graders* (17).

3. Aquisição e exportação das imagens

As retinografias utilizadas para *grading* foram obtidas com ângulos de 30°, centradas na fóvea (campo 2), após midríase farmacológica. Para este efeito utilizou-se uma máquina fotográfica *Topcon TRC- 50DX* (*Topcon Corporation, Tokyo, Japan*), ou uma *Zeiss FF-450 Plus* (*Carl Zeiss Meditec Inc., Dublin, CA*). Acopladas a estas máquinas existiam sensores, sendo no caso do *MEEI* um *Topcon cameras–Pike 11MP CCD* e no *CHUC* um *Nikon D2H*.

As imagens foram exportadas e guardadas em formato *jpeg* e *tif*. Posteriormente, foram carregadas para a rede segura do *CORC* e daí colocadas em dois computadores distintos, um para classificação pelo *software RetmarkerAMD* e outro para classificação com o *software Topcon IMAGEnet 2000*. Antes da exportação para as plataformas de *grading*, todas as imagens foram tratadas com um *software* automático de otimização de cor, brilho e contraste (Figura 1) (18).

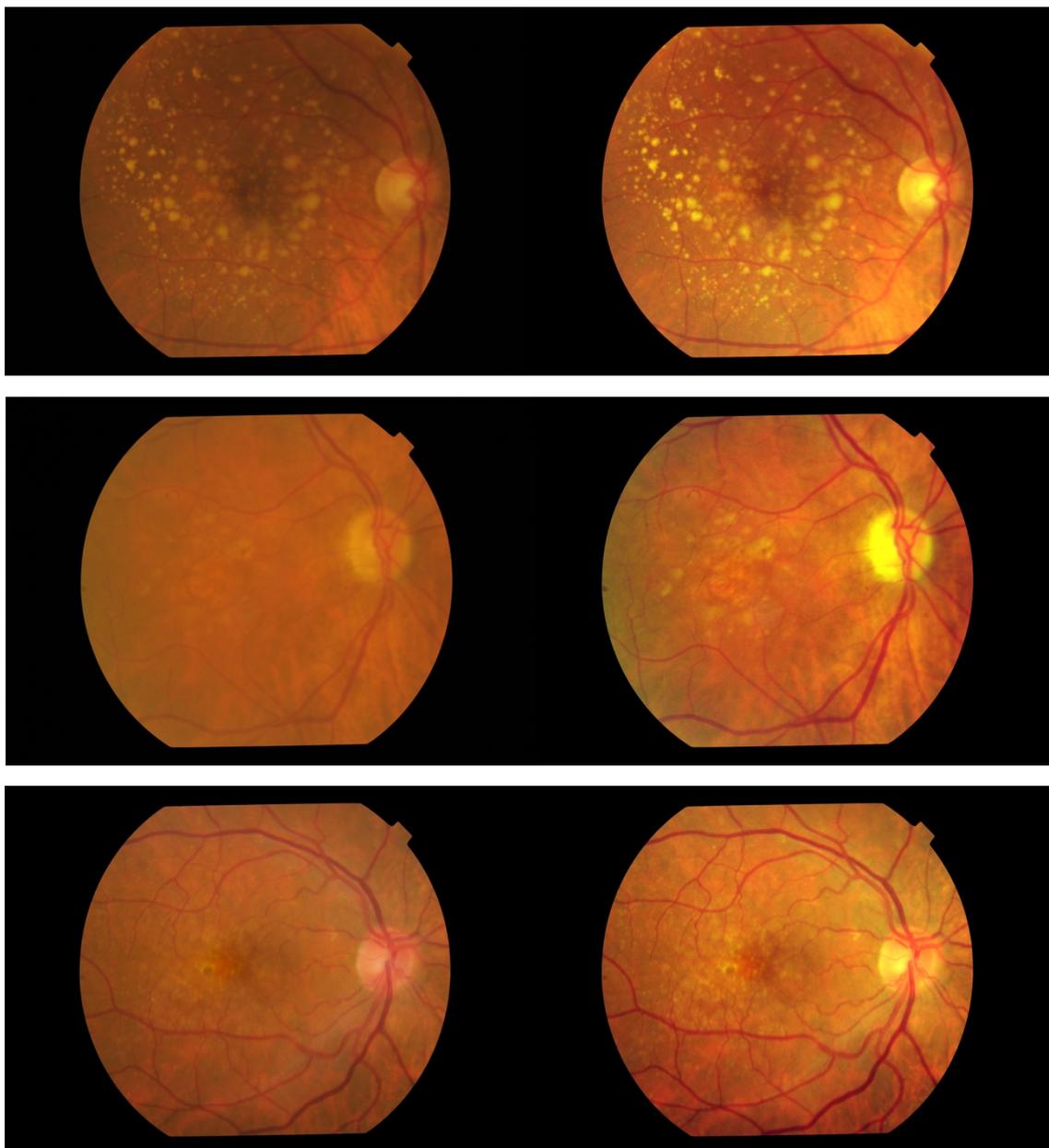


Figura 1- Retinografias de doentes com DMI antes e após tratamento com *software* de otimização de cor, brilho e contraste (18). Comparação de 3 casos. Em todos, as retinografias da esquerda representam o "antes" e as da direita o "depois" da aplicação do *software* referido. Como é facilmente observado, a utilização do *software* de otimização de imagem permite uma mais fácil identificação e delimitação das lesões fundoscópicas típicas de DMI como *drusens* e alterações pigmentares.

4. Grading das Retinografias

4.1. RetmarkerAMD

O *software RetmarkerAMD* (*Critical Health SA*) é um sistema inovador de classificação semi-automática que tem por base o protocolo internacional de classificação da DMI definido

por *Bird et al* (19). Os componentes do sistema permitem (i) diferenciar imagens classificadas de imagens não classificadas; (ii) avaliar as imagens em modo de ecrã inteiro; (iii) desenhar formas circulares pré-definidas (63 μm , 125 μm , 175 μm , 250 μm ou 500 μm) que possuem cores diferentes consoante o tipo de lesão correspondente (*drusen*, hiperpigmentação, hipopigmentação, atrofia geográfica ou lesão exsudativa); (iv) desenhar formas livres para todos os tipos de lesões referidos; (v) identificar confluência de *drusen*; (vi) ampliar/reduzir a imagem ou uma área específica desta; (vii) medir distâncias na imagem e alterar definições de cor (*Red, Blue, Green*), contraste e brilho; (viii) sobrepor uma grelha *standardizada* para identificar diferentes campos e (ix) introduzir manualmente os dados do doente em estudo. Cada imagem é calibrada antes de iniciar qualquer medição, usando o diâmetro do disco ótico (1,8 mm) como referência. Seguidamente, o centro da mácula (fóvea) é manualmente indicado, permitindo que o *software* obtenha os pontos de referência necessários e forneça a grelha de referência. Esta é composta por 3 círculos concêntricos e duas linhas que interseccionam a horizontal aos 45° e 135°, de acordo com a calibração prévia (Figura. 2). Os diâmetros dos círculos central, interno e externo são de 1000 μm , 3000 μm e 6000 μm , respetivamente. As alterações morfológicas que podem ser classificadas por este sistema incluem: número de *drusen* (total; <63 μm ; $\geq 63\mu\text{m}$ e < 125 μm ; $\geq 125\mu\text{m}$; $\geq 175 \mu\text{m}$; $\geq 250 \mu\text{m}$); tipo de *drusen* (duros, moles indistintos, semi-sólidos, serogranulares, cristalinos), área ocupada por *drusen* (<1%, <10%, <25%, <50%, $\geq 50\%$, total e para cada um dos círculos separadamente); área ocupada por classe de *drusen* e área de *drusen* em cada semi-campo; confluência de *drusen* (<10%, <50%, $\geq 50\%$); hipo e hiperpigmentação (ausente, <63 μm , $\geq 63 \mu\text{m}$), DMI atrófica e DMI neovascular. Além de permitir uma quantificação categórica do número (0, 1-9, 10-19, ≥ 20) e área (<1%, <10%, <25%, <50%, $\geq 50\%$) de *drusen*, o *RetmarkerAMD* permite, adicionalmente, determinar o número real de *drusen* e a área real de *drusen* (em μm^2) total ou por semi-campo. Com base no

grading efetuado, o *software* calcula automaticamente o estadio da DMI com base na Classificação de Roterdão (Tabela 2).

Além do *grading* de retinografias, o *RetmarkerAMD* permite também classificar imagens de autofluorescência, *infra-red* e *redfree*.

Tabela 2. Escala de Roterdão para classificação da DMI (20)

Estadio de DMI		Critérios
Controlo	0a	Sem sinais de DMI
	0b	<i>Drusens</i> duros <63µm
Precoce	1a	<i>Drusens</i> moles distintos
	1b	Apenas alterações pigmentares, sem <i>drusens</i> moles
Intermédia	2a	<i>Drusens</i> moles indistintos ($\geq 125 \mu\text{m}$) ou <i>drusens</i> reticulares apenas
	2b	<i>Drusens</i> moles distintos ($\geq 63 \mu\text{m}$) com alterações pigmentares
	3	<i>Drusens</i> moles indistintos ($\geq 125 \mu\text{m}$) ou <i>drusens</i> reticulares com alterações pigmentares
Avançada	4	DMI neovascular ou atrófica

Os resultados decorrentes da classificação das imagens surgem em tempo real no ecrã e podem ser exportados para um ficheiro *Excel* para posterior análise estatística.

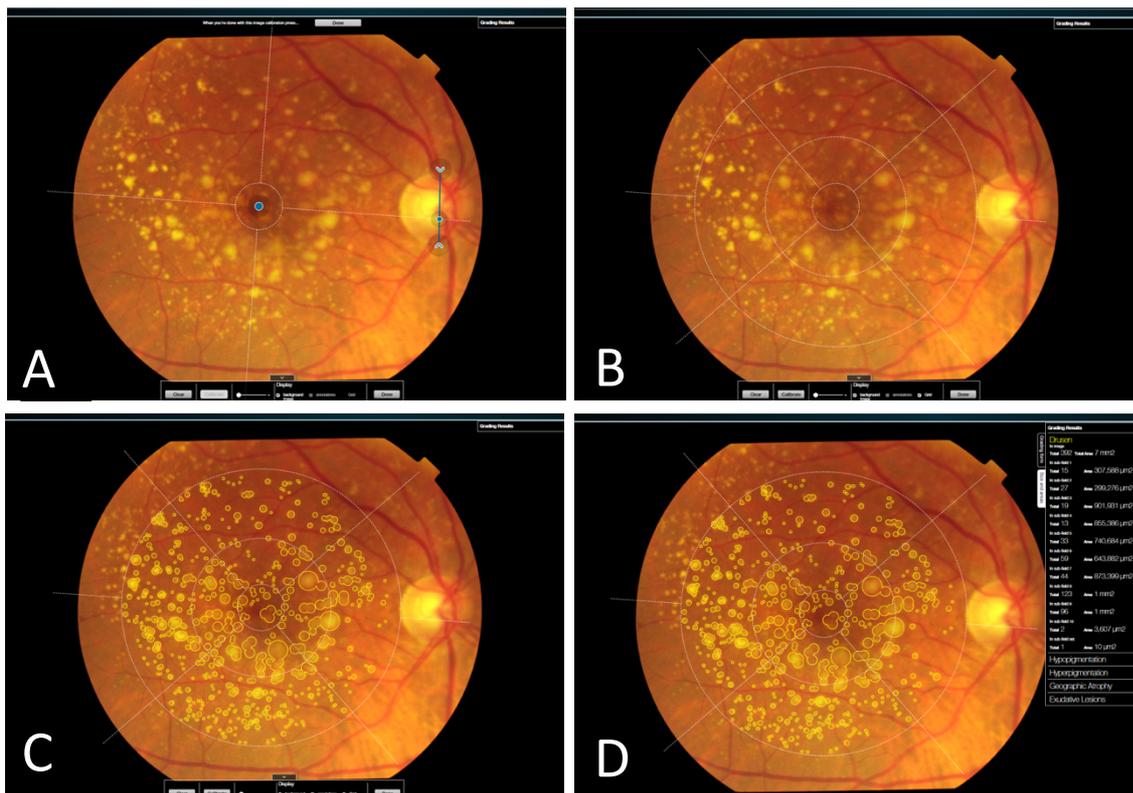


Figura 2- Demonstração em tempo real do processo de *grading* utilizando o *RetmarkerAMD* (*Retmarker SA*). **(A)** Calibração. Primeiro é necessária a identificação manual da fóvea e estabelecer o diâmetro do disco óptico (ponto azul e setas, respetivamente). O *software* gera posteriormente uma grelha **(B)** de acordo com o *International Grading System* para a DMI (19). **(C)** As alterações do fundo ocular, incluindo *drusen*, alterações pigmentares, AG e DMI neovascular podem ser quantificadas nas retinografias utilizando formas livres ou círculos pré-definidos (63, 125, 175, 250 e 500 μm). **(D)** Os resultados do *grading* são apresentados automaticamente, sendo possível a sua exportação para posterior análise estatística através de um ficheiro de *Microsoft Excel* (21).

4.2. Topcon IMAGEnet 2000

Este *software*, amplamente usado em estudos epidemiológicos, inclusivamente no estudo *AREDS* (22) permite a sobreposição de uma grelha macular (*Early Treatment Diabetic Retinopathy Study - ETDRS*) na retinografia a classificar. Além disso, permite a utilização de ferramentas de planimetria para efetuar medições calibradas (22). Para a identificação do tamanho das lesões, o *software* permite a utilização de círculos com diâmetros pré-definidos, nomeadamente C-0 (63 μm de diâmetro); C-1 (125 μm de diâmetro); C-2 (250 μm); I-2 (360 μm); O-2 (790 μm); $\frac{1}{2}$ Área de Disco (DA) (1273 μm) e 1 DA (1800 μm) (22). O *grader* tem assim a possibilidade de classificar a imagem tendo por base a classificação *AREDS* (Tabela 3). Contudo, contrariamente ao *software RetmarkerAMD*, parâmetros como o número, a área ou a

confluência de *drusens* não são automaticamente analisados, já que os círculos de referência servem apenas para auxiliar o *grader* na determinação do estadio da DMI de acordo com a classificação *AREDS*. A classificação final do estadio da DMI é, por isso, 100% dependente do *grader*.

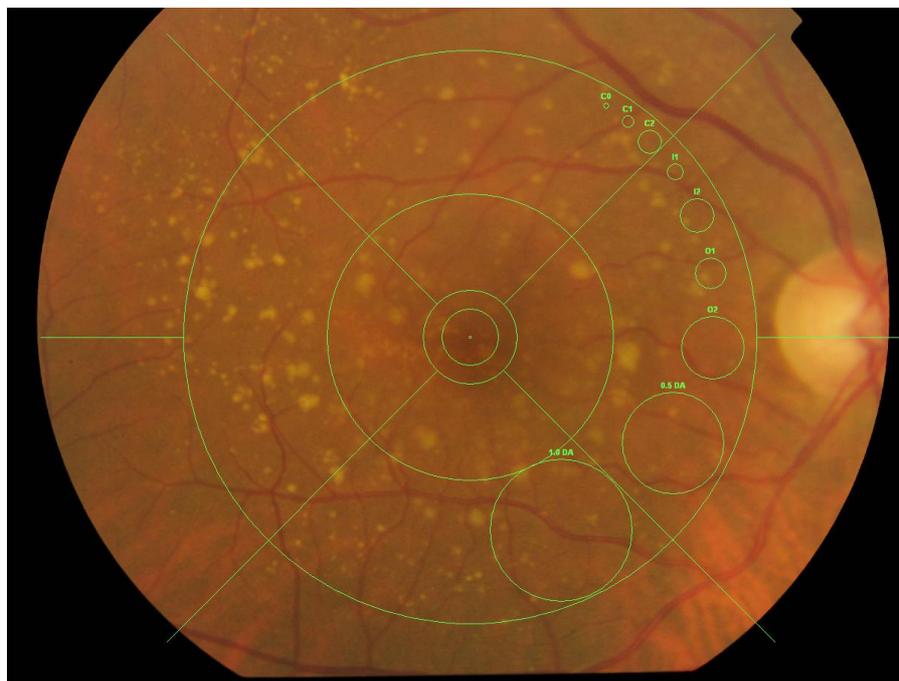


Figura 3- Retinografia analisada no *Topcon IMAGEnet 2000*, aplicando a grelha macular *ETDRS* e os círculos pré-definidos para ajudar na classificação (*vide* texto para definição dos diâmetros dos círculos e tabela 3 para a classificação do estadio da DMI segundo o *AREDS*)

Tabela 3- Classificação dos níveis de DMI grupo *AREDS* (11)

Estadio de DMI	Critérios
1 (Controlo)	Tamanho máximo de <i>drusen</i> < C-0 (63µm de diâmetro) e área total < C-1 (125 µm de diâmetro)
2 (Precoce)	Presença de 1 ou mais dos seguintes: <ol style="list-style-type: none"> a) <i>Drusen</i> de tamanho máximo ≥ que C-0 mas < C-1 b) Área total ocupada por <i>drusen</i> ≥ C-1 c) Anormalidades do epitélio pigmentar da retina consistente com DMI, definida como uma ou mais das seguintes na região central ou nas subdivisões centrais: <ol style="list-style-type: none"> a. Despigmentação

	<p>b. Pigmentação aumentada \geq C-1</p> <p>c. Pigmentação aumentada ou despigmentação questionáveis</p>
<p>3 (Intermédia)</p>	<p>Presença de 1 ou mais dos seguintes:</p> <p>a) à presença de <i>drusens</i> de tamanho máximo \geq C-1</p> <p>b) à presença de <i>drusens</i> de tamanho máximo \geq C-O e área total $>$ I-2 sendo o tipo de <i>drusen soft indistinct</i></p> <p>c) à presença de <i>drusens</i> de tamanho máximo \geq C-O e área total $>$ O-2 sendo o tipo de <i>drusen soft distinct</i></p> <p>d) Atrofia Geográfica dentro da grelha mas não no centro da mácula</p>
<p>4 (Avançada)</p>	<p>Presença de 1 ou mais dos seguintes:</p> <p>a) Atrofia Geográfica nas subdivisões centrais com pelo menos um questionável envolvimento da região central da mácula.</p> <p>b) Existência de alterações compatíveis com DMI neovascular</p> <p>a. Descolamento seroso ou fibrovascular do epitélio pigmentar</p> <p>b. Descolamento seroso ou hemorrágico da retina sensorial</p> <p>c. Hemorragia subretiniana</p> <p>d. Fibrina ou fibrose subretiniana</p> <p>e. Fotocoagulação para a DMI</p>

5. Análise Estatística

Tendo em conta que os sistemas usados para classificação do estadió de DMI são diferentes nos dois *softwares*, foi feita a seguinte conversão para possibilitar a comparação entre o *RetmarkerAMD* e o *Topcon IMAGEnet 2000*:

- 0a e 0b da classificação de Roterdão considerado como Controlo (Sem DMI)
- 1a e 1b da classificação de Roterdão considerado como DMI precoce

- 2a, 2b, 3a e 3b da classificação de Roterdão considerado como DMI intermédia
- 4 da classificação de Roterdão considerado como DMI Avançada

A concordância entre *graders* foi atingida através do cálculo da percentagem de concordância (i.e., o número de imagens com a mesma classificação) e o coeficiente *Kappa* foi ponderado tendo em conta ponderações lineares.

A percentagem de concordância foi calculada considerando o número de imagens com a mesma classificação tendo em conta o número total das mesmas, ou seja, 202 imagens.

O valor de *Kappa* foi calculado através do *Stata version 12.1*.

Para a interpretação dos valores de *Kappa* foi utilizado o critério de *Landis e Koch (17)*, ou seja, com valores abaixo de 0,20 a concordância seria classificada como *slight*; valores entre 0,21 e 0,40 seria classificada como *fair*; entre 0,41 e 0,60 seria *moderate*; entre 0,61 e 0,80 *substantial* e entre 0,81 e 1 seria *almost perfect* (Ver Tabela 5).

Tabela 4. Valores de *Kappa* de acordo com *Landis & Koch and Fleiss (17)*

<i>Kappa</i>	<i>Agreement</i>
	<i>Landis. J.R.; & Koch. G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics 33 (15): 159–174. Doi: 10.2307/2529310</i>
<0.00	<i>Poor</i>
0 – 0.20	<i>Slight</i>
0.21–0.40	<i>Fair</i>
0.41–0.60	<i>Moderate</i>
0.61-0.80	<i>Substantial</i>
0.81–1.00	<i>Almost Perfect</i>

Resultados

Apresentam-se os resultados de concordância para o *software RetmarkerAMD* (*grader JPM vs grader JP*), para o *software Topcon IMAGEnet 2000* (*grader JPM vs grader JP*; *grader JPM vs goldstandard*; *grader JP vs goldstandard*) e entre ambos os *softwares* (*grader JPM vs grader JPM e grader JP vs grader JP*).

1. *RetmarkerAMD* (JPM vs JP)

Como descrito na secção de Material e Métodos, o *software RetmarkerAMD* coloca à disposição do *grader* várias variáveis passíveis de ser analisadas e que ultrapassam em muito a simples classificação da DMI.

A tabela 5 apresenta os resultados de concordância entre os *graders* JPM e JP para todos os parâmetros do *RetmarkerAMD*. De ressaltar que a concordância média registada foi de 86,2%. Considerando apenas o estadio de DMI, a concordância entre *graders* foi *almost perfect* (concordância observada = 93.0%; *Kappa*=0.95, $p<0.001$).

Tabela 5. Concordância JPM vs JP na análise com o *RetmarkerAMD*

<i>Variable</i>	<i>% Agreement</i>	<i>Kappa coefficient</i>	<i>Strength of agreement</i>	<i>P-value</i>
Número de <i>drusens</i>	85.2%	0.86	<i>Almost Perfect</i>	<0.001
Número de <i>drusens</i> <63µm	82.7%	0.83	<i>Almost Perfect</i>	<0.001
Número de <i>drusens</i> 63-125µm	76.7%	0.78	<i>Substantial</i>	<0.001
Número de <i>drusens</i> >125µm	82.7%	0.80	<i>Substantial</i>	<0.001
Tipo de <i>drusen</i> predominante na parte externa	83.2%	0.25	<i>Fair</i>	<0.001
Área total ocupada por <i>drusens</i>	90.0%	0.75	<i>Substantial</i>	<0.001
Área ocupada por <i>drusens</i> na subdivisão 1	85.2%	0.84	<i>Almost Perfect</i>	<0.001
Área ocupada por <i>drusens</i> no círculo central	83.0%	0.69	<i>Substantial</i>	<0.001
Área ocupada por <i>drusens</i> no círculo externo	86.5%	0.66	<i>Substantial</i>	<0.001
Confluência	48.8%	0.32	<i>Fair</i>	<0.001
Hiperpigmentação	97.0%	0.93	<i>Almost Perfect</i>	<0.001
Hipopigmentação	99.5%	0.98	<i>Almost Perfect</i>	<0.001
Atrofia Geográfica	99.5%	0.96	<i>Almost Perfect</i>	<0.001
DMI Neovascular	100.0%	1.00	<i>Perfect</i>	<0.001
Estadio DMI	93.0%	0.95	<i>Almost Perfect</i>	<0.001

Topcon IMAGEnet 2000 (JPM vs JP vs Gold Standard)

A concordância observada entre *graders* utilizando o *software* Topcon IMAGEnet 2000 no que diz respeito ao estadio da DMI foi *almost perfect* (concordância observada = 90.1%; *Kappa* = 0.87, $p < 0.001$). A concordância entre os *graders* JPM e JP e o *goldstandard* foi igualmente *almost perfect*:

- JPM vs *goldstandard*: concordância observada=90.1%; *Kappa*=0.88, $p < 0.001$;
- JP vs *goldstandard*: concordância observada=87.1%; *Kappa*=0.86, $p < 0.001$.

RetmarkerAMD vs Topcon IMAGEnet 2000

Como previamente descrito na secção de Materiais e Métodos, o sistema de classificação da DMI utilizado pelos 2 *softwares* é distinto. Para tornar possível a comparação, a classificação dos *graders* no RetmarkerAMD (Roterdão) foi convertida em Controlo (sem DMI), DMI Precoce, Intermédia ou Avançada (*vide* ponto 5 de Materiais e Métodos)

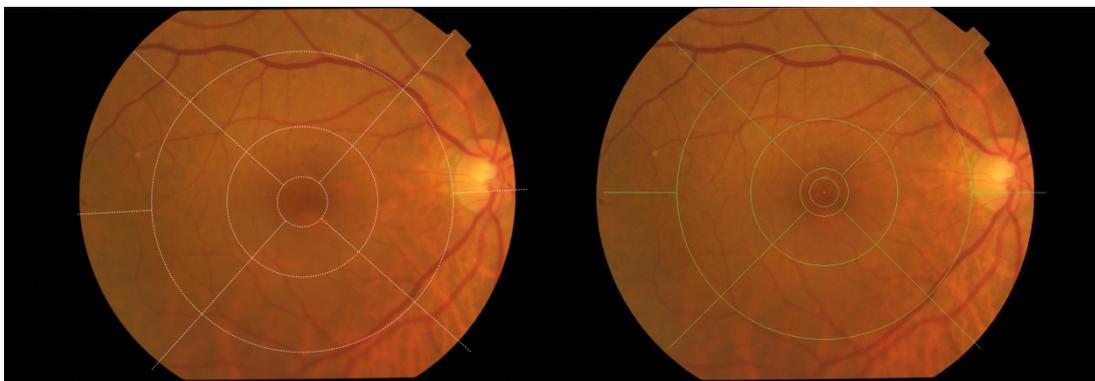


Figura 4- Retinografias com grelha automática gerada em ambos os *softwares*. A retinografia da esquerda corresponde ao RetmarkerAMD e a retinografia da direita corresponde ao Topcon IMAGEnet 2000.

JPM obteve uma concordância *substantial* no parâmetro “Estadio de DMI” entre o RetmarkerAMD e o Topcon IMAGEnet 2000 (concordância observada = 76.8%; *Kappa* = 0.73, $p < 0.001$).

JP obteve também uma concordância *substantial* para a mesma análise (concordância observada = 70.8%; *Kappa* = 0.67, $p < 0.001$).

Nas tabelas 6 e 7 podem observar-se as diferenças em relação aos estadios de DMI entre *softwares*, para JPM e JP, respetivamente. Como pode ser observado, a maior discrepância ocorreu entre estadios precoce (*early*) e intermédio (*intermediate*), tanto para o *grader* JPM como para o *grader* JP.

Tabela 6- Concordância entre Topcon IMAGENet 2000 e RetmarkerAMD feita por JPM

JPM						
		RetmarkerAMD				
		Control	Early	Intermediate	Late	Can't Grade
Topcon IMAGENet 2000	Control	25	5	0	0	0
	Early	8	18	4	1	1
	Intermediate	0	13	81	4	4
	Late	0	0	5	26	0
	Can't Grade	1	0	0	0	2

Tabela 7- Concordância entre Topcon IMAGENet 2000 e RetmarkerAMD feita por JP

JP						
		RetmarkerAMD				
		Control	Early	Intermediate	Late	Can't Grade
Topcon IMAGENet 2000	Control	23	4	0	0	0
	Early	15	11	2	1	2
	Intermediate	2	21	84	5	2
	Late	0	0	5	25	0
	CG	0	0	0	0	0

Discussão

A importância da quantificação das lesões fundoscópicas no desenvolvimento de formas avançadas de DMI tem levado ao desenvolvimento e aperfeiçoamento de *softwares* de *grading*, tendo como principal objetivo a deteção precoce dos olhos com maior risco de progressão (13,14,16,23,24). Espera-se que a aplicação destes sistemas à escala global permita quebrar barreiras geográficas e oferecer cuidados de saúde a um cada vez maior número de doentes. A popularidade dos sistemas de *grading* automáticos (*deep learning*) (13,23,25) e semi-automáticos (14,16,24,26) tem vindo a crescer pela facilidade/comodidade de utilização, rapidez, transversalidade e precisão, principalmente quando comparada com métodos de análise convencionais, como a análise manual (14,15).

Assim, este trabalho teve como principal objetivo validar um *software* semi-automático desenvolvido especificamente para o *grading* de imagens de doentes com DMI – *RetmarkerAMD*. Tal como no *AREDS*, usou-se a retinografia, uma vez que oferece uma imagem bastante fiel à que é observada no exame fundoscópico, tornando fácil a extrapolação dos resultados para a prática clínica (12). Além disso, a retinografia garante um registo permanente, possibilitando a reprodutibilidade da análise (11). Para oferecer aos *graders* imagens com a maior qualidade possível, todas as retinografias foram sujeitas a processamento com um *software* automático de otimização de cor, brilho e contraste, já validado para *grading* (18).

As imagens foram classificadas por 2 *graders* independentes em dois *softwares* diferentes – *RetmarkerAMD* e *Topcon IMAGEnet 2000*. As mesmas imagens tinham já sido classificadas no *software Topcon IMAGEnet 2000* por um *grader* sénior (RS), tendo essa classificação servido como *goldstandard*. Procedeu-se a uma análise intra- e *intergrader* para avaliação da concordância.

1. *RetmarkerAMD* (JPM vs JP)

Como podemos ver em “Resultados”, os valores de concordância JPM vs JP são bastante elevados (*almost perfect*) no que respeita ao estadió da DMI (concordância observada = 93.0%; $Kappa=0.95$, $p<0.001$) e mesmo quando avaliados todos os parâmetros oferecidos pelo *RetmarkerAMD*, (concordância média observada de 86.2%). Contudo, observaram-se algumas diferenças significativas, a mais notória na variável “confluência de *drusen*” – obtendo uma classificação de *Fair* (concordância foi de 48.8% e $Kappa = 0,32$; $p<0.001$). São várias as explicações possíveis para a discrepância, entre as quais se destaca a subjetividade inerente a uma análise deste tipo no que diz respeito a grandes áreas de *drusens* moles e confluentes e também a experiência dos *graders* (25). O *grader* JPM é oftalmologista, dedicando-se à área da retina e contando já com 5 anos de experiência como *grader*. Pelo contrário, o *grader* JP é estudante de medicina e teve o seu primeiro contacto com o *grading* da DMI durante o presente trabalho. Para minimizar as discrepâncias, o *grader* JP só pôde iniciar funções de *grader* neste estudo após certificação pelo *reading centre*. Ainda assim, para um *grader* jovem, é totalmente natural que surjam dúvidas relativamente às reais dimensões dos *drusens* em áreas de confluência, bem como dificuldades na sua identificação pela semelhança com focos de hipopigmentação ou *pseudodrusens* reticulares, cujo padrão fundoscópico pode mimetizar *drusens* confluentes. Apesar de tudo, e tendo em conta a curta experiência do *grader* JP, os resultados de concordância são amplamente satisfatórios, demonstrando que o *RetmarkerAMD* é uma plataforma de interface simples que permite o *grading* de lesões típicas de DMI mesmo por *graders* menos experientes.

Também *Scholl et al* (16) exploraram as potencialidades dos métodos semi-automáticos no *grading* da DMI. O seu estudo contou com a avaliação da concordância *inter-grader* e *intra-grader*, sendo que o mesmo *grader* faz as medições com pelo menos com 2 semanas de intervalo. Os resultados de concordância por eles obtidos foram *fair to moderate* (variando entre

62% e 75% com $Kappa=0.30-0.47$) para a concordância *inter-grader*. Para *drusens* duros, esta foi de 70–89% ($Kappa = 0.26-0.63$), parâmetro que neste estudo foi de 82,7% usando o *RetmarkerAMD*. Já para casos de DMI avançada, os autores destacam valores de concordância na ordem dos 88–98% ($Kappa = 0.60-0.95$) para a AG e de 84–100% ($Kappa = 0.62-1.00$) para lesões neovasculares. Na nossa análise com o *RetmarkerAMD*, os valores de concordância foram de 99,5% e 100%, respetivamente para casos de AG e DMI neovascular. Podemos constatar que os valores acima citados são bastante mais baixos do que aqueles que foram obtidos no presente estudo, para a mesma análise (concordância *inter-grader*). Para este resultado poderá ter contribuído o facto de a amostra do estudo de *Scholl et al* ser bastante mais pequena ($n=50$), tornando-a mais sujeita a variações estatísticas. Pelo contrário, neste trabalho, o cálculo do tamanho da amostra permitiu dimensionar o estudo, conferindo-lhe a robustez necessária para a validação dos *outcomes*. Apesar de reportarem uma deteção eficaz e bem-sucedida, o método usado por *Scholl et al* (16) não permite a classificação em *drusens* duros/moles e distintos/indistintos, contrariamente ao *Retmarker AMD*, onde estas variáveis podem ser avaliadas. Foi também destacada por estes (16) e outros autores (24) a dificuldade na definição dos limites dos *drusens* bem como a dificuldade na sua identificação, pela sua semelhança com outras lesões (nomeadamente focos de hipopigmentação e manchas algodinosas). Utilizando o *RetmarkerAMD*, o *grader* indica qual o tipo de *drusen* mais prevalente, tornando possível a classificação automática da DMI de acordo com a escala de Roterdão. Além disso, a utilização de um *software* de otimização de cor, brilho e contraste, facilitou a correta identificação das lesões típicas de DMI neste estudo.

Assim, conclui-se que o *RetmarkerAMD* permite uma análise precisa, completa e simples de todas as alterações típicas de DMI. A sua plataforma *user friendly* permite que a curva de aprendizagem seja curta, algo que é facilmente demonstrado pela elevada concordância observada entre 2 *graders* com níveis de experiência diferentes. Outra vantagem

deste *software*, é o facto de permitir uma análise multimodal, uma vez que além do *grading* de retinografias, também de imagens de autofluorescência, *infra-red* e *red free* podem ser classificadas.

Apesar das vantagens referidas, o *RetmarkerAMD* também apresenta limitações, a principal das quais inerente ao facto de não ser um sistema automático mas sim semi-automático. Isto implica que um *grader* faça a marcação das lesões, um processo que, dependendo dos casos, pode ser moroso. Além disso, não permite a classificação de imagens de tomografia de coerência ótica (OCT), um método de imagem com resultados promissores, especialmente no que respeita à classificação e medição da área ocupada por *drusens* (22).

Topcon IMAGEnet 2000 (JPM vs JP vs Gold Standard)

A concordância entre *graders* mostrou valores *almost perfect* (concordância observada = 90.1%; *Kappa* = 0.87, $p < 0.001$), tal como para a análise *RetmarkerAMD* (*JPM vs JP*), vista anteriormente. O *grader* JPM obteve uma classificação superior à do *grader* JP quando efetuada a análise de concordância com o *goldstandard* (concordância observada=90.1% vs concordância observada=87.1%, respetivamente). A diferença de experiência mencionada anteriormente pode ter influenciado os resultados. De facto, a experiência do *grader* é um fator explicativo transversal a vários estudos para justificar discrepâncias na análise de concordância (16,24). O *Topcon IMAGEnet 2000* é um *software* que não indica automaticamente o estadió de DMI existente, sendo essa responsabilidade inteiramente projetada no *grader*. Ainda assim, e tendo em conta os diferentes níveis de experiência, ambos os *graders* obtiveram valores de concordância extremamente satisfatórios.

RetmarkerAMD vs Topcon IMAGEnet 2000

A análise *intragrader* revelou, para os *graders* JPM e JP, uma concordância de 76,8% e de 70,8%, respetivamente. Assim, e apesar da diferença no valor absoluto, ambos obtiveram a classificação de *substantial* para a comparação das suas análises entre *RetmarkerAMD* e *Topcon IMAGEnet 2000*.

Através da análise das tabelas 6 e 7, podemos constatar que a maior discrepância ocorreu entre estadios *early* e *intermediate*, tanto para o *grader* JPM como para o *grader* JP. Pode haver relação com o facto de as classificações utilizadas serem diferentes, uma vez que o *RetmarkerAMD* utiliza a escala de Roterdão, bastante mais pormenorizada e o *Topcon IMAGEnet 2000* utiliza uma classificação mais simplificada, baseada na escala *AREDS*. Desta forma, a classificação baseada na escala de Roterdão teve de ser convertida nos estadios de DMI definidos pela escala *AREDS* (*vide Material e Métodos - Análise Estatística*) para assim permitir a comparação entre os dois *softwares*, o que pode ter introduzido um viés de classificação. Outro aspeto contribuidor pode ter sido, como já foi referido, o facto de o *Topcon IMAGEnet 2000* não indicar automaticamente o estadio de DMI existente, ao contrário do que acontece no *RetmarkerAMD*. Desta forma, a análise feita no *Topcon IMAGEnet 2000*, revela-se muito mais *grader dependent*, sendo assim altamente influenciável por fatores como o nível de experiência entre *graders*.

Existe um esforço crescente para a validação de métodos automáticos na classificação da DMI. Contudo, quando comparados com a análise manual, as concordâncias parecem ainda não ser suficientemente robustas para uma aplicação em larga escala, principalmente no que respeita à análise de imagens mais complexas ou com pouco contraste (14,24). *Kirkpatrick et al* (23) observaram uma sensibilidade de 60% e especificidade de 90% no que diz respeito à deteção de *drusens*. Não obstante, *Burlina et al* (13) concluíram que os métodos automáticos podem ter a mesma eficácia que métodos manuais, atingindo valores de *Kappa* entre 0.764 e

0.829. Já *Shin et al.*(14) salientaram que os métodos automáticos permitem uma análise robusta, eficiente e bastante mais rápida.

Uma vez que não existem estudos de comparação de *softwares de grading* automáticos vs semi-automáticos não é possível comparar o estudo feito neste trabalho com outros. Ainda assim, os sistemas semi-automáticos parecem, por enquanto, ser mais precisos e apesar de *time-consuming*, garantem excelentes resultados de concordância, tal como demonstrado neste estudo.

Conclusão

Em suma, este estudo avaliou e validou a capacidade do *RetmarkerAMD* no que respeita ao *grading* da DMI, a partir da análise de retinografias, tendo obtido resultados bastante animadores quando comparado com um dos *softwares* tido como *goldstandard* para esta análise, *Topcon IMAGEnet 2000*. Além da classificação do estadio da DMI, o *RetmarkerAMD* permite avaliar vários parâmetros como o número, localização e área de *drusens*, tendo-se observado uma concordância média *intergrader* de 86,2% para todos os parâmetros analisados. O *RetmarkerAMD* revelou-se um *software* credível, preciso e útil na classificação das lesões típicas de DMI, podendo ajudar na monitorização das lesões ao longo do tempo e contribuir para a identificação de biomarcadores imagiológicos precoces das formas avançadas da doença.

Agradecimentos

Dr. João Pedro Marques,

Por toda a orientação, paciência, apoio e disponibilidade demonstrados e por mostrar sempre que seria possível. Um muito obrigado por tudo!

Prof. Doutor Rufino Silva,

Pela disponibilidade e apoio dado. Obrigado!

Prof. Doutor Joaquim Murta,

Por como diretor do Serviço de Oftalmologia do CHUC ter permitido que realizasse o meu projeto neste serviço

Dra. Inês Laíns,

Por ter contribuído com as imagens provenientes do *Massachusetts Eye and Ear Infirmary*

AIBILI (na pessoa do Prof. Doutor José Cunha-Vaz, Dra Cecília Martinho e todo o staff)

Por terem permitido a realização de parte do meu estudo nesta instituição. Obrigado!

Agradecimento especial ao CORC (na pessoa da Professora Doutora Conceição Lobo e staff - Catarina Neves, Raquel Branco, Sofia Gomes e Christian Schwartz)

Por terem permitido a realização do meu estudo na instituição, bem como por toda a ajuda e disponibilidade demonstradas. Obrigado!

4C (na pessoa da Doutora Sandrina Nunes e Dalila Alves)

Pela colaboração essencial em todo o processo, em especial na análise estatística

Aos meus pais

Por toda a ajuda e apoio incondicional, conselhos e por estarem sempre presentes, disponíveis, mesmo quando tudo pareceu intransponível. Obrigado!

À Daniela

Por todo o apoio, por estar sempre presente, pelos conselhos, pela ajuda, por toda a dedicação e carinho. Obrigado!

Amigos e família

Por me acompanharem ao longo de todo este percurso

Referências

1. Mehta S. Age-Related Macular Degeneration. *Prim Care - Clin Off Pract.* 2015;42(3):377–91.
2. Cachulo M da L, Laíns I, Lobo C, Figueira J, Ribeiro L, Marques JP, et al. Age-related macular degeneration in Portugal: prevalence and risk factors in a coastal and an inland town. *The Coimbra Eye Study Report 2. Acta Ophthalmol.* 2016;94(6):e442–53.
3. Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng CY, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis. *Lancet Glob Heal.* 2014;2(2):e106–16.
4. Klein R, Klein BEK, Knudtson MD, Meuer SM, Swift M, Gangnon RE. Fifteen-Year Cumulative Incidence of Age-Related Macular Degeneration. *The Beaver Dam Eye Study. Ophthalmology.* 2007;114(2):253–62.
5. Klein R, Klein BEK, Jensen SC, Meuer SM. The Five-year Incidence and Progression of Age-related Maculopathy. *Ophthalmology.* 1997;104(1):7–21.
6. Stein JD, Brown MM, Brown GC, Hollands H, Sharma S. Quality of life with macular degeneration: perceptions of patients, clinicians, and community members. *Br J Ophthalmol.* 2003;87(1):8–12.
7. Cachulo MDL, Lobo C, Figueira J, Ribeiro L, Laíns I, Vieira A, et al. Prevalence of age-related macular degeneration in Portugal: The Coimbra eye study - Report 1. *Ophthalmologica.* 2015;233:119–27.
8. Klaver CCW, Wolfs RCW, Assink JJM, et al. Genetic Risk of Age-related Maculopathy. 1998; 116:1646–51.
9. Seddon JM, Mitchell BD. Familial Aggregation of Age-related Maculopathy: Author Reply. *Am J Ophthalmol.* 1997;124(6):862–3.
10. Davis MD, Gangnon RE, Lee L-Y, Hubbard LD, Klein BEK, Klein R, et al. The Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration. *Arch Ophthalmol.* 2005;123(11):1484.
11. Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study Report Number 6. *Am J Ophthalmol.* 2001;132(6):668–81.
12. Age-Related Eye Disease Study Research Group. A Simplified Severity Scale for Age-Related Macular Degeneration. AREDS Report No. 18. *Arch Ophthalmol.* 2005;123(11):1570.
13. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmol.* 2017;1–7.
14. Shin DS, Javornik NB, Berger JW. Computer-assisted, interactive fundus image processing for macular drusen quantitation. *Ophthalmology.* 1999;106(6):1119–25.
15. Marques JP, Costa M, Melo P, Oliveira CM, Pires I, Cachulo ML, et al. Ocular Risk Factors for Exudative AMD: A Novel Semiautomated Grading System. *ISRN Ophthalmol.* 2013;2013(3):464218.
16. Scholl HPN, Peto T, Dandekar S, Bunce C, Xing W, Jenkins S, et al. Inter- and intra-observer variability in grading lesions of age-related maculopathy and macular degeneration. *Graefe's Arch Clin Exp Ophthalmol.* 2003;241(1):39–47.
17. Landis JR KG. The measurement of observer agreement for categorical data. *Biometrics.*; 1977. 159-74 p.
18. Tsikata E, Laíns I, Gil J, Marques M, Brown K, Mesquita T, et al. Automated Brightness and Contrast Adjustment of Color Fundus Photographs for the Grading of Age-Related Macular

- Degeneration. *Transl Vis Sci Technol.* 2017;6(2):3.
19. Bird a. C, Bressler NM, Bressler SB, Chisholm IH, Coscas G, Davis MD, et al. An international classification and grading system for age-related maculopathy and age-related macular degeneration. *Surv Ophthalmol.* 1995;39(5):367–74.
 20. van Leeuwen R, Klaver CCW, Vingerling JR, Hofman A, de Jong PTV. The Risk and Natural Course of Age-Related Maculopathy. *Arch Ophthalmol.* 2003;121(4):519.
 21. Marques JP, Ms C, Láins I, Ms C, Costa MÁ, Ms C, et al. RETINAL ANGIOMATOUS PROLIFERATION A Quantitative Analysis of the Fundoscopic Features of the Fellow Eye. *Retina.* 2015;35:1985–91.
 22. Danis RP, Domalpally A, Chew EY, Clemons TE, Armstrong J, SanGiovanni JP, et al. Methods and reproducibility of grading optimized digital color fundus photographs in the Age-Related Eye Disease Study 2 (AREDS2 Report Number 2). *Investig Ophthalmol Vis Sci.* 2013;54(7):4548–54.
 23. Kirkpatrick JN, Spencer T, Manivannan A, Sharp PF, Forrester J V. Quantitative image analysis of macular drusen from fundus photographs and scanning laser ophthalmoscope images. *Eye.* 1995;9 (Pt 1)(0950–222X (Print)):48–55.
 24. Smith RT, Sohrab MA, Pumariega NM, Mathur K, Haans R, Blonska A, et al. Drusen analysis in a human-machine synergistic framework. *Arch Ophthalmol.* 2011;129(1):40–7.
 25. Niemeijer M, Van Ginneken B, Russell SR, Suttorp-Schulten MSA, Abramoff MD. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investig Ophthalmol Vis Sci.* 2007;48(5):2260–7.
 26. Smith RT, Nagasaki T, Sparrow JR, Barbazetto I, Klaver CCW, Chan JK. A Method of Drusen Measurement Based on the Geometry of fundus reflectance. 2003;13:1–13.

