UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE ENGENHARIA
CÂMPUS DE ILHA SOLTEIRA

GUILHERME PAGANINI CONSTANTINO DE OLIVEIRA

RECONHECIMENTO DO ESTADO DE SAÚDE DE GLOBOS OCULARES UTILIZANDO A REDE NEURAL ARTMAP *FUZZY*

Ilha Solteira
2023

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Ilha Solteira

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

GUILHERME PAGANINI CONSTANTINO DE OLIVEIRA

RECONHECIMENTO DO ESTADO DE SAÚDE DE GLOBOS OCULARES UTILIZANDO A REDE NEURAL ARTMAP *FUZZY*

Dissertação apresentada à Faculdade de Engenharia de Ilha Solteira – Unesp como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica.

Profª. Drª. Anna Diva Plasencia Lotufo
Orientadora

Ilha Solteira
2023

FICHA CATALOGRÁFICA
Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

O48r

Oliveira, Guilherme Paganini Constantino de.
Reconhecimento do estado de saúde de globos oculares utilizando a rede neural ARTMAP Fuzzy / Guilherme Paganini Constantino de Oliveira. -- Ilha Solteira:[s.n.], 2023
90 f. : il.

Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2023

Orientador: Anna Diva Plasencia Lotufo

Inclui bibliografia

1. Anomalias oculares. 2. ARTMAP Fuzzy. 3. Codificação de atributos. 4. Extração de atributos de imagens. 5. Reconhecimento de imagens. 6. Redes neurais artificiais.

Amanda Sertori dos Santos
Bibliotecária - CRB/8-9061
Seção Técnica de Referência, Atendimento ao Usuário e Documentação
Diretoria Técnica de Biblioteca e Documentação

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Ilha Solteira

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: RECONHECIMENTO DO ESTADO DE SAÚDE DE GLOBOS OCULARES UTILIZANDO A REDE NEURAL ARTMAP *FUZZY*

AUTOR: GUILHERME PAGANINI CONSTANTINO DE OLIVEIRA

ORIENTADORA: ANNA DIVA PLASENCIA LOTUFO

Aprovado como parte das exigências para obtenção do Título de Mestre em , área: Automação pela Comissão Examinadora:

gov.br Documento assinado digitalmente
ANNA DIVA PLASENCIA LOTUFO
Data: 03/11/2023 08:31:31-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. ANNA DIVA PLASENCIA LOTUFO (Participaçao Presencial)
Departamento de Engenharia Eletrica / Faculdade de Engenharia de Ilha Solteira - UNESP

Profa. Dra. MARA LUCIA MARTINS LOPES (Participaçao Presencial)
Departamento de Matematica / Faculdade de Engenharia de Ilha Solteira - UNESP

Prof. Dr. KENJI NOSE FILHO (Participaçao Virtual)
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas / Universidade Federal do ABC - UFABC

Ilha Solteira, 27 de setembro de 2023

IMPACTO POTENCIAL DESTA PESQUISA

A pesquisa fomenta possível auxílio diagnóstico na área da oftalmologia em pacientes com eventuais patologias oculares, contribuindo para início precoce de admissível tratamento necessário e podendo favorecer as chances de recuperação. Os resultados podem ser potencializados, futuramente, para implementação em dispositivos inteligentes e estruturados para distinção de anomalias individuais.

POTENTIAL IMPACT OF THIS RESEARCH

The research provides possible diagnostic support in the field of ophthalmology for patients with possible eye pathologies, contributing to the early start of the necessary treatment and potentially improving the chances of recovery. In the future, the results could be used to implement intelligent, structured devices to distinguish individual anomalies.

À Kethelyn Pinheiro, dedico.

AGRADECIMENTOS

Agradeço a Deus por nunca me desamparar e me proporcionar inúmeras bênçãos, mesmo após minhas sucessivas falhas como ser humano. Agradeço por me instigar com ideias que nem mesmo eu achei que poderia arquitetar.

À minha futura esposa, Kethelyn, por conseguir me tornar uma pessoa melhor do que eu jamais merecia ser e me mostrar o lado bom das coisas.

À minha mãe, por ter me criado e proporcionado educação em meio a diversas dificuldades.

À professora Anna Diva, pela atenção, paciência e sugestões. Agradeço pela confiança depositada em mim para realização do trabalho.

Ao professor Minussi, pelas recomendações proporcionadas e pela prontidão em auxiliar.

À Danieli e Giovanni, por contribuírem com dicas em momentos absolutamente cruciais da execução do trabalho, meus mais sinceros agradecimentos. Ao Reginaldo, pelas sugestões de plataformas na seleção do *dataset*.

Agradeço ao amigo Bruno, de longa data, pelos incontáveis conselhos durante os cafezinhos no câmpus. Obrigado pela amizade. Agradeço também ao amigo Victor, pelo companheirismo durante o mestrado.

Ao Sintel e à FEIS, por proporcionarem a estrutura necessária para viabilização desta pesquisa.

À CNPq e à CAPES, pelo apoio financeiro concedido para realização do trabalho.

“ *I did not tell half of what I saw, for I knew I would not be believed*” - Marco Polo.

RESUMO

Com o desenvolvimento digital tão em pauta atualmente, diversas ferramentas de inteligência artificial puderam ser apropriadamente estabelecidas e passaram a fazer parte importante do cotidiano. Dentro deste tópico, o reconhecimento de imagens tornou-se bastante significativo em inúmeras aplicações, incluindo a área da saúde. Neste escopo de análise, ressalta-se a abordagem destinada à detecção do estado de saúde de globos oculares, constituindo premissa de extrema relevância e com grande possibilidade de exploração na pesquisa científica. O presente trabalho propõe a utilização da rede neural ARTMAP *Fuzzy* para reconhecimento do estado de saúde de globos oculares, tratando-se de uma empregabilidade pioneira da rede. Utilizou-se um banco de dados de imagens de fundos oculares com boa quantidade de amostras de qualidade para possibilitar o reconhecimento de imagem. Manipulações e tratamento nas imagens foram realizados a fim de permitir uma melhor padronização dos dados a serem apresentados à rede. Testou-se três diferentes processos de extração e codificação dos atributos dos dados, que incluíram: transcrição dos pixels das imagens em valores numéricos representativos; contagem de aparecimento de pixels específicos via histograma; e codificação por meio da frequência de palavras visuais existentes (método BoVW - *Bag of Visual Words*). Testes para verificação da capacidade da rede em discernir olhos saudáveis de não saudáveis foram efetuados. A métrica de taxa de classificação correta (CCR - *Correct Classification Rate*) foi empregada para comparação de eficiência dos resultados para cada uma das abordagens. Apesar dos três métodos apresentarem bom percentual de acerto geral, a utilização do BoVW demonstrou melhor distribuição de eficiência em relação aos outros pela congruência de precisão quantos aos olhos saudáveis individualmente.

Palavras-chave: anomalias oculares; ARTMAP *Fuzzy;* codificação de atributos; extração de atributos de imagens; reconhecimento de imagens; redes neurais artificiais.

ABSTRACT

With digital development so much on the agenda these days, several artificial intelligence tools have been properly established and have become an important part of everyday life. Within this topic, image recognition has become very significant in numerous applications, including the health field. Within this scope of analysis, the approach aimed at detecting the state of health of eyeballs stands out, constituting a premise of extreme relevance and with great potential for exploration in scientific research. This work proposes the use of the fuzzy ARTMAP neural network to recognize the state of health of eyeballs, which is a pioneering use of the network. A database of eye fundus images with a good number of quality samples was used to enable image recognition. The images were manipulated and processed in order to better standardize the data to be presented to the network. Three different data attribute extraction and coding processes were tested, which included: transcription of image pixels into representative numerical values; appearance count of specific pixels via histogram; and coding through the frequency of existing visual words (BoVW - Bag of Visual Words). Tests were carried out to verify the network's ability to discern healthy from unhealthy eyes. The Correct Classification Rate (CCR) metric was used to compare the efficiency of the results for each of the approaches. Although all three methods showed a good percentage of accuracy in the general case, the use of BoVW showed a better distribution of efficiency in comparison to the others due to the congruence of accuracy with regard to healthy eyes individually.

Keywords: ocular anomalies; Fuzzy ARTMAP; attribute coding; extraction of attributes from images; image recognition; artificial neural networks.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

AF	(Rede Neural Artificial) ART *Fuzzy*
AMF	(Rede Neural Artificial) ARTMAP *Fuzzy*
ART	*Adaptative Resonance Theory*
BP	*Backpropagation*
Cat C	Categorias Criadas (módulo ART_a)
CCA	*Canonical Correlation Analysis*
CCR	*Correct Classification Rate*
CNN	*Convolutional Neural Network*
D.O	Disco Óptico
E/D	Olho: Esquerdo/Direito
HOG	*Histogram of Oriented Gradients*
LTM	*Long Term Memory*
NC	Normalização Convencional
NVB	Normalização pelo Valor Base
P/B	Preto e Branco
PCA	*Principal Component Analysis*
QPV	Quantidade de Palavras Visuais
RFMiD	*Retinal Fundus Multi-Disease Image Dataset*
RGB	*Red, Green, Blue (Additive Color Model)*
RNAs	Redes Neurais Artificiais
ROC	*Receiver Operating Characteristic*
S/NS/G	Caso Ocular: Saudável/Não Saudável/Geral
SGGS	*Shri Guru Gobind Singhji Institute of Engineering and Technology*
STM	*Short Term Memory*
TSLN	*Ocular Tesselation*

SUMÁRIO

1 INTRODUÇÃO

Desde a proposição da unidade básica do neurônio artificial (MCCULLOCH; PITTS, 1943), o desenvolvimento dos preceitos vinculados à inteligência artificial ocorreram rapidamente e fundamentaram-se cada vez mais nas atividades que permeiam nosso cotidiano. Sistemas estruturados em aprendizados de máquina estão inerentes cultural e cientificamente, e as RNAs contribuíram significativamente para que este patamar fosse devidamente alcançado.

As RNAs correspondem a idealizações de estruturas de inteligência artificial embasadas no funcionamento cerebral humano. Estas arquiteturas ancoram-se na disposição específica das unidades individuais de aprendizado (neurônios artificiais) e, mediante o processo experimental vinculado a cada uma, permitem que tais unidades, via ajuste progressivo de pesos ponderadores, generalizem uma função descritora do sistema que possibilita sua utilização para análises inéditas (HAYKIN, 1999). A evolução do mundo digital promoveu o aparecimento acentuado das redes neurais artificiais nos diversos âmbitos do conhecimento, sendo comumente empregadas em inúmeras tarefas de classificação e predição de eventos.

Uma vez que as RNAs se fundamentam no funcionamento intrínseco do cérebro humano, é natural buscar que elas idealizem, cada vez mais, as características singulares do aprendizado. É necessário, portanto, que os sistemas possuam aptidão em armazenar conhecimento via treinamento e tenham capacidade de generalizar respostas razoavelmente precisas para dados novos e constituam flexibilidade de aplicação em variadas vertentes (WASSERMAN, 1989), características que se fazem presentes na estrutura cerebral humana.

Mediante a capacidade do cérebro de adequação e organização complexa dos neurônios, tarefas relacionadas a processamentos de percepção, controle motor e reconhecimento de padrões (HAYKIN, 1999) fazem parte essencial das rotinas humanas e permitem constantes interações com o mundo. Dentre estas, vale o destaque para a acentuada distinção que o cérebro pode efetuar das mais inúmeras ocorrências que se apresentam cotidianamente. Independente da entrada visual que uma determinada pessoa esteja recebendo, rapidamente ocorre o reconhecimento e a consequente associação em alguma determinada “classe neural” de conhecimento preexistente. Caso esta classe não exista, o aprendizado humano permite que um novo grupo seja criado para que este novo estímulo seja vinculado. A utilização de

sistemas inteligentes e, mais precisamente, de RNAs em aplicações cotidianas, tem como um dos principais objetivos efetivar o relatado anteriormente: estruturar um modelo que possa discernir e reconhecer padrões de imagens e estímulos com acurácia significativa, tal como nosso cérebro é capaz de fazer sistematicamente.

O reconhecimento de imagens e padrões trata-se de uma das pautas mais presentes no contexto de evolução digital, cujo desenvolvimento propiciou grandes melhoras em diversas áreas da sociedade. Efetuar adequadamente o reconhecimento de imagens, pautado em um determinado objetivo, permite otimizar e automatizar tarefas que outrora necessitavam inspeção visual (MATHWORKS, 2023a; GONZALEZ; WOODS, 2009). Além disso, sendo possível identificar padrões em imagens e usar estas informações para tomar decisões assertivas, fica assegurada alta confiabilidade do sistema e o mesmo pode ser redirecionado a ramos essenciais da vida humana.

Por intervenção desta propriedade do reconhecimento de imagens, sistemas podem monitorar ambientes e detectar situações de risco ou perigo; aparelhos móveis pessoais podem ser vinculados ao usuário e desbloqueados, garantindo potencialização da segurança; drones com câmeras podem sobrevoar regiões de matas densas e detectar possíveis focos de incêndio; consumidores podem buscar mais facilmente produtos específicos via verificação de preferências em compras online; profissionais da saúde podem utilizar ferramentas de detecção de anomalias; dentre muitas outras aplicações. Este progresso foi possível, em grande parte, à constante criação de algoritmos poderosos capazes de assimilar padrões de dados e efetuar predições de eventos (PIURI et al., 2020).

Uma das vertentes mais importantes de aplicação do reconhecimento de imagens trata-se da área da saúde. Utilizar inteligência artificial para auxiliar na detecção de anomalias fisiológicas corresponde a um tópico de extrema relevância, visto que um sistema devidamente treinado e com boa acurácia pode colaborar com o diagnóstico de um paciente e, por consequência, permite otimizar o tempo para tomada de ações no eventual tratamento (PACHADE et al., 2021). Neste contexto, a detecção de anomalias oculares e do estado de saúde do olho como um todo destaca-se pela quantidade de avarias inerentes que podem aparecer no decorrer da vida do ser humano. Além disso, muitas destas anomalias podem estar vinculadas a ocorrências raras, sendo necessária uma boa estruturação de dados e amostras a fim

de possibilitar o uso de técnicas computacionais de reconhecimento, uma vez que o leque de possibilidades é considerável (PACHADE et al., 2021).

Para estas aplicações, é comum utilizar-se de técnicas de aprendizado profundo (*deep learning*) (DENG; YU, 2014), como a rede neural convolucional (CNN). A vantagem destes métodos reside no processamento de grande quantidade de dados para melhor análise e algoritmos auto adaptativos para lidar com uma maior quantidade de amostras, possuindo boa aptidão em extrair características complexas dos componentes do banco de dados e favorecendo utilização nestas aplicações (PIURI et al., 2020).

Além das estruturas vinculadas ao aprendizado profundo, uma classe de RNAs que merece destaque pertence às redes da família ART, ancoradas na teoria da ressonância adaptativa. Estas estruturas foram altamente estudadas por Carpenter e Grossberg e, devido ao alto caráter de flexibilidade das redes, foi possível obter diversas variações de arquiteturas, cada qual destinando-se à manipulação de dados diferentes e efetuando assimilação de padrões adaptativos e auto estabilização em resposta a estímulos de entrada complexos (GROSSBERG, 1988). Além disso, por serem altamente interativas, é factível a integração de módulos e a implementação de alterações nas arquiteturas, algo que possibilita cada vez mais atingir novos objetivos.

Muito do que torna esta família de redes interessante corresponde à possibilidade de lidar com o dilema da estabilidade e plasticidade, além de possuir estruturas de memória de curto e longo prazo (GROSSBERG, 1976a). Esta capacidade de lidar com este dilema permite à rede ancorar-se nos dois preceitos simultaneamente: de um lado, garante que ocorrerá a convergência para uma determinada configuração de pesos que generalize o processo para o qual ela foi treinada; de outro, assegura que a estrutura será capaz de aprender continuamente com a inclusão de novos padrões sem se desfazer do aprendizado obtido até o momento. Esta capacidade é vital dado que nosso cérebro, ao qual as redes fundamentam-se, possui tais aptidões e frequentemente lida com estímulos ambientais variantes e não estacionários, aos quais distinguimos e lidamos continuamente (CARPENTER; GROSSBERG, 1988).

Dito isso, é interessante ressaltar que estes princípios permeiam duas grandes classificações básicas de modelos de arquiteturas destas redes: os aprendizados não-supervisionados e os supervisionados. No primeiro, caracterizado por um módulo ART independente, ocorre a categorização dos padrões de entrada sucessivos

apresentados à rede por meio de associações dos que apresentam características similares (CARPENTER; GROSSBERG, 1987a). No segundo caso, os treinamentos supervisionados exigem que se possua um par treinado para a execução do processo. A cada entrada de dados deve-se associar uma saída, justamente para permitir que a estrutura elabore um mapeamento assertivo dos pares. Neste caso, as categorias de entrada e saída serão relacionadas, e mais de um módulo ART é necessário para efetuar o procedimento. Destaca-se, portanto, as variações da rede ARTMAP (CARPENTER; GROSSBERG; REYNOLDS, 1991).

Dada toda esta conjuntura, muitos trabalhos têm sido feitos a fim de estudar as propriedades intrínsecas da AMF e como a variação de seus parâmetros atua efetivamente no desempenho da mesma (ANAGNOSTOPOULOS; GEORGIOPOULOS, 2002; GEORGIOPULOS; HUANG; HEILMAN, 1994). Por ser uma rede altamente flexível, com boa característica de convergência e possibilitar excelente tratativa de variados tipos de entradas, a AMF pode ser explorada em inúmeras aplicações, incluindo classificação, predição de eventos e até mesmo aproximação de funções (CANO-IZQUIERDO et al., 2013).

Com todo este panorama da AMF, é curioso que estudos envolvendo reconhecimento de imagens não tenham usualmente empregado suas arquiteturas. Apesar das características interessantes presentes nas técnicas de *deep learning* (DENG; YU, 2014), um questionamento específico faz-se necessário: dada a facilidade de manipulação e flexibilidade da AMF, esta poderia fornecer uma alternativa sólida a despeito da eficiência quando em tratativas de reconhecimento de imagens?

Tratando-se especificamente do reconhecimento do estado de saúde de globos oculares, é instigante a suposição de que, se as amostras forem devidamente codificadas e modeladas, o processo de treinamento e as características próprias da rede AMF devem permitir boa generalização, eficiência e acurácia em teste com amostras inéditas. Se a rede é capaz de trabalhar bem com diversos tipos de dados, basta que sejam apresentados adequadamente para que o desempenho seja satisfatório.

O grande desafio, portanto, não reside inteiramente no processo de treinamento da estrutura, que é capaz de se auto organizar e maximizar generalização com redução do erro preditivo (CARPENTER; GROSSBERG; REYNOLDS, 1991): a forma como as amostras serão codificadas e padronizadas será o principal

determinante no desempenho final da rede. Se por um lado, nas técnicas de *deep learning*, a extração de características das imagens representativas para reconhecimento posterior pode ser efetuadas diretamente nos entrelaçamentos de conexões e camadas das redes (RAN et al., 2021), por outro, no caso da utilização da rede AMF, é necessário que a extração de atributos seja realizada anteriormente e de forma individual, possibilitando o treinamento adequado da estrutura.

O presente trabalho propõe explorar tal pioneirismo de utilização da rede, analisando sua eficácia para esta aplicabilidade não usual de detecção de estado de saúde ocular por meio de reconhecimento de imagens e verificando se pode constituir alternativa sólida para tal.

1.1 OBJETIVO E PROPOSTA

A proposta de trabalho ancora-se na utilização da rede ARTMAP *Fuzzy* para o tema de reconhecimento de imagens, tal como discutido anteriormente. Almeja-se utilizar esta arquitetura para a finalidade específica de efetuar a distinção do estado de saúde de globos oculares entre saudáveis e não saudáveis. A classificação “saudável” refere-se a olhos isentos de qualquer tipo de doença; em contrapartida, olhos classificados como “não saudáveis” correspondem a exemplares acometidos por uma ou mais das 45 condições presentes no *dataset* (PACHADE et al., 2021).

A eficiência da rede será testada para diferentes formas de extração de atributos e codificação das imagens: conversão da matriz de dados em vetor linha, frequência de aparecimento de pixels via histograma de imagens, e aplicação do conceito de sacola de palavras visuais (BoVW).

Testar-se-á a eficácia recorrendo a amostras inéditas alheias às do treinamento. Se disporá de um *dataset* com imagens de fundos de retina de globos oculares (PACHADE et al., 2021) e aplicar-se-á a métrica de taxa de classificação correta (CCR).

1.2 ORGANIZAÇÃO TEXTUAL

O referido trabalho é organizado em seis capítulos, incluindo a presente introdução.

No Capítulo 2 serão expostos e discutidos alguns trabalhos da literatura relacionados ao tema de reconhecimento de imagens, assim como algumas técnicas usuais para efetivação das análises e estudos sobre parâmetros da rede AMF.

O Capítulo 3 detalha os preceitos e fundamentos que permeiam a teoria da ressonância adaptativa. Aqui, é discutido o teorema da estabilidade e plasticidade, apresentando os conceitos principais de aprendizado das redes da família ART. Posteriormente, descreve-se o funcionamento prático da arquitetura da rede ART e, em seguida, da rede ARTMAP *Fuzzy*, evidenciando como tais algoritmos de treinamento possibilitam o ajuste progressivo dos pesos.

O Capítulo 4 apresenta toda a metodologia que foi empregada para viabilizar a implementação do tema na arquitetura da rede AMF. São descritos os passos efetuados no que diz respeito ao tratamento, modelagem e extração de atributos das imagens. Além da descrição geral do *dataset*, são expostas todas as manipulações das imagens a fim de possibilitar obtenção da maior quantidade de amostras adequadas para o treinamento. Além disso o capítulo descreve, passo a passo, todos os métodos de extração de atributos e codificações das imagens empregados, mostrando como cada um foi viabilizado e suas respectivas fundamentações.

No Capítulo 5 são apresentados os resultados computacionais obtidos mediante os testes com amostras inéditas. São descritos os parâmetros utilizados para o treinamento da rede e os atributos específicos obtidos após sua conclusão. Os melhores resultados sugeridos para cada caso são comentados e analisados, verificando a eficiência de cada abordagem.

No Capítulo 6 expõe-se as conclusões a respeito da aplicabilidade da rede ARTMAP *Fuzzy* no tema de reconhecimento de imagens e detecção de estado de saúde de globos oculares, trazendo-se perspectivas e sugestões de eventuais pesquisas futuras a serem exploradas na área a fim de potencializar o estudo.

Por fim, no Apêndice A é evidenciado, de forma breve, a visualização da distribuição de patologias oculares existentes nas amostras pertencentes ao banco de dados inicialmente.

2 REVISÃO BIBLIOGRÁFICA

O presente capítulo busca apresentar um panorama a respeito de alguns estudos vinculados à área tema do trabalho, evidenciando algumas particularidades de determinadas pesquisas no ramo de reconhecimento de imagens quando centralizado o uso de RNAs. Além disso, também almeja-se fornecer um escopo de alguns estudos que tem por base o uso específico da rede AMF. Julga-se interessante, também, apresentar algumas pesquisas destinadas a análises funcionais e interpretativas da AMF mediante aprofundamento dos seus parâmetros intrínsecos, evidenciando como o campo de estudo é rico e permite exploração.

A evolução constante dos sistemas de inteligência permitiu que técnicas fundamentadas no funcionamento cerebral fossem cada vez mais desenvolvidas, culminando na popularização de RNAs e observando-se que o padrão de progresso do reconhecimento de imagens tendeu a fomentar o aprendizado profundo e o aumento de conexões entre as camadas da rede cada vez mais (AMIDI, A.; AMIDI S., 2023). No entanto, vale a ressalva de que existem diversas estruturas passíveis de exploração que ainda não são muito exploradas no tópico referido, tal como as redes da família ART.

Desde o pioneirismo da fundamentação lógica da teoria da ressonância adaptativa (GROSSBERG, 1976a) até o estabelecimento do funcionamento da rede AMF (CARPENTER et al., 1992), diversos estudos vêm sendo realizados a fim de compreender melhor o funcionamento desta arquitetura. Para tal, além de modificações específicas na estrutura, pesquisas a respeito dos parâmetros e conceitos da rede permitiram que esta idealização evoluísse progressivamente.

No âmbito do aprofundamento das especificidades da rede, Georgiopoulos, Huang e Heileman (1994) analisam propriedades de aprendizado da rede ARTMAP. Resultados preliminares evidenciam como a estrutura prioriza a escolha de nós já atribuídos em detrimento de nós sem atribuição, favorecendo a manutenção da singularidade das categorias.

Em outro estudo, realizado por Georgiopoulos et al. (1996), verifica-se como o parâmetro de escolha pode afetar a ordem de seleção das categorias da rede AF e AMF. Tais compreensões fornecem base para entendimento de algumas propriedades específicas de arquiteturas pertencentes à família ART.

Ainda no contexto de estudo imersivo das características das redes pertencentes à família ART, Anagnostopoulos e Georgiopoulos (2002) evidenciam um estudo a respeito das regiões de categoria e conceitos geométricos das redes AF e AMF. A definição dessas regiões baseadas em interpretações geométricas do teste de vigilância é vista com detalhes e o conceito de disputa dos nós da camada de reconhecimento (nós não atribuídos versus nós atribuídos) é analisado, sendo possível relacionar a progressiva diminuição do volume representativo de cada categoria com o avanço do processo de treinamento e consequente estabilização.

Tais pesquisas fornecem exemplo de como o estudo aprofundado das redes da família ART permite um acúmulo incremental de compreensão de todo o funcionamento da estrutura e possibilita caminhar, cada vez mais, numa direção de aumento da gama de aplicabilidade. Em suma, a colaboração de vários estudos com o passar do tempo promoveu um maior entendimento e assegurou o uso das redes para um leque variado de aplicações.

Já no campo específico do reconhecimento de imagens, as redes da família ART têm sido pouco exploradas. No estudo feito por Al-Obaydy e Suandi (2020), utiliza-se a rede AMF para efetuar reconhecimento facial em monitoramento de vídeo empregando apenas uma amostra por pessoa no treinamento. Os autores recorrem a dois tipos de extração de características nas imagens: histograma de gradientes orientados (HOG) (DALAL; TRIGGS, 2005) e *wavelet* de Gabor (LIU; WECHSLER, 2002). Os vetores adquiridos são reduzidos de dimensionalidade via análise de componentes principais (PCA) (JOLLIFE, 2002) e fundidos por meio de análise de correlação canônica (CCA) (KRZANOWSKI, 2000), formando os vetores finais do treinamento. O estudo expõe que a AMF se sobressai se comparada a outras técnicas testadas, mostrando que a arquitetura pode ser mais explorada como forma alternativa no âmbito de reconhecimento.

No que se refere a análises na área de detecção de anomalias oculares por reconhecimento de imagens, o uso de aprendizado profundo é o mais explorado. Na análise específica de glaucomas, por exemplo, a revisão de Ran et al. (2021) mostra um panorama de métodos de *deep learning* usados na literatura para tal e inclui a tomografia de coerência óptica para formar as imagens. Segundo os autores, a grande vantagem de utilização de *deep learning* reside na capacidade de as redes lidarem com múltiplas camadas de processamento para extração de características, facilitando posteriormente o reconhecimento de padrões similares. Em contrapartida,

outras técnicas de *machine learning* (ECML, 1995) exigem que as características principais a serem extraídas sejam muito bem definidas e determinadas. O processo de codificação é vital para a eficiência do método utilizado, algo que, por conta das variações individuais de cada *dataset*, pode exigir um trabalho formidável. Esta dificuldade é reduzida com o aprendizado profundo nas técnicas citadas pelos autores, uma vez que existe uma alta capacidade de modelagem automática das características por conta das múltiplas camadas (RAN et al., 2021). Por conta deste motivo fundamental, muitas das pesquisas na área recorrem às redes ancoradas em *deep learning* como método principal. No entanto, vale ressaltar que basta uma boa estruturação do vetor de características para que a rede AMF, por exemplo, performe com satisfatória acurácia, tal como mostrado anteriormente no trabalho de Al-Obaydy e Suandi (2020).

Outro estudo de revisão interessante de ser mencionado refere-se à pesquisa feita por Atwany, Sahyoun e Yaqub (2022). Nela, os autores discutem técnicas de *deep learning* utilizadas na literatura para classificação de olhos com retinopatia diabética. São relatadas diversas abordagens empregadas, incluindo técnicas supervisionadas e auto supervisionadas (menos sujeitas a ruído, mas necessitadas de grande quantidade de dados) utilizadas na classificação binária (olhos saudáveis/olhos com retinopatia) e na classificação de múltiplas condições (na qual determina-se o grau de severidade da doença). Resultados são apresentados para as diversas abordagens, no qual algumas obtêm acurácia média ultrapassando 80% considerando o total dos dados, ainda que em certas tratativas a precisão individual de atributos (verdadeiros positivos/verdadeiros negativos) fosse reduzida.

No estudo de Tufail et al. (2021) aborda-se o caso de retinopatia diabética utilizando-se um número limitado de amostras e um banco de dados com imagens de fundos oculares. Os autores empregam a rede convolucional 3D (3D-CNN) para classificação binária e de múltiplas classes, utilizando extensão de dados artificialmente (*upsampling*) para complementar o *dataset* limitado (criando-se cópias modificadas de imagens já existentes). Obtém-se precisão média na classificação binária em torno de 60%. Os autores trabalham com menos de 200 amostras de treinamento no caso binário e pouco mais de 1000 para o caso de classificação múltipla, evidenciando a aplicação da extensão de dados de forma artificial.

Na pesquisa de Sengar et al. (2023), busca-se efetuar o diagnóstico e classificação de múltiplas anomalias vinculadas à retina propondo o uso da rede

EyeDeep-Net, contextualizada nos preceitos de *deep learning*. Munidos do *dataset* proveniente de Pachade et al. (2021), os autores separam as etapas em três módulos principais: remoção dos dados adequados, *upsampling* e treinamento da rede, compondo a arquitetura proposta EyeDeep-Net (que basicamente consiste numa CNN de diversas camadas para investigar os padrões de características, aplicada para diagnóstico de anomalias oculares). Os autores obtêm uma média de precisão de 76% aproximadamente, evidenciando boa assertividade da rede proposta. No entanto, a extensão de dados empregada apresentou baixa variação em relação às fotos originais, favorecendo o resultado final.

Kim et al. (2022) estudam a possibilidade de retorno da miopia em pacientes que passaram pelo procedimento cirúrgico refrativo na córnea. Para tal, é utilizada a rede ResNet-50 para análise das imagens e, para agrupamento de todas as informações vinculadas aos dados pré-operatórios dos pacientes (dados demográficos, medidas pré-operatórias, método refrativo utilizado, entre outras), emprega-se a categoria de algoritmo *XGBoost* (CHEN; GUESTRIN, 2016), ancorado em árvores de decisão com gradiente descendente para minimizar a perda. Mediante treinamento do sistema com todos os dados mencionados, os autores puderam ter uma assertividade de previsão superior quando comparada com aplicação individual da ResNet-50 por meio da análise da plotagem da curva ROC.

Os estudos relatados mostram como abordar casos do ramo de saúde ocular com técnicas de inteligência artificial, notadamente, as RNAs, são de suma importância a fim de auxiliar no diagnóstico e, consequentemente, otimizar o tempo de tratamento. Apesar do uso massivo de técnicas de *deep learning*, com a devida estruturação da etapa de extração de características e formação dos vetores de treinamento, é natural imaginar que uma rede pouco usual para a aplicabilidade (como a AMF) possa ter um desempenho satisfatório. A tarefa de extração de atributos das imagens e codificação representa a principal atividade para permitir que a rede interprete os dados adequadamente e atribua uma boa generalização ao sistema, ainda que possa requerer trabalho árduo e testes prévios para implementação.

3 TEORIA DA RESSONÂNCIA ADAPTATIVA

A Teoria da Ressonância Adaptativa foi um marco que possibilitou o desenvolvimento de redes neurais fundamentadas neste conceito que, até hoje, se expandem em novas arquiteturas aplicáveis a diversas áreas. Como dito anteriormente, as RNAs se baseiam no funcionamento cerebral humano para fundamentar suas mais variadas ramificações. A tentativa de compreender como os sistemas biológicos são capazes de reter a plasticidade durante a vida, sem comprometer a estabilidade de padrões anteriormente aprendidos, motivou o desenvolvimento progressivo deste tópico (WEENINK, 1997).

Podemos definir a plasticidade de um determinado sistema, respaldado no funcionamento do cérebro, como a capacidade de adquirir conhecimento sistematicamente. Esta aptidão de aprender de maneira incremental é característica do desenvolvimento cognitivo humano, sendo abordada rotineiramente no campo das RNAs. Por outro lado, a estabilidade mencionada refere-se à capacidade do sistema em garantir o agrupamento de todos os elementos nas classes criadas, isto é: esta propriedade garante que o sistema atingirá a convergência e o aprendizado ocorrerá com o ajuste progressivo dos pesos (AMORIM, 2006). A comunhão dos dois conceitos permeia a definição prática do dilema da estabilidade/plasticidade: o sistema deve conseguir aprender novos eventos sem perder ou deteriorar informações previamente adquiridas.

É interessante ressaltar o paralelo existente com o funcionamento de aprendizado humano, uma vez que os preceitos das RNAs sempre fundamentam tais analogias. Como seres humanos, somos frequentemente expostos a estímulos visuais e sensoriais inéditos. A partir do momento que recebemos tal informação decorrente de circunstância original, podemos aprender e categorizá-la adequadamente, permitindo assimilação futura caso nos deparemos com o mesmo estímulo mencionado. Após este processo de aprendizado, todo nosso conhecimento anterior a esta influência fica preservado e não passa por deterioração. Esta analogia compete perfeitamente ao dilema da plasticidade/estabilidade: ocorreu o aprendizado de uma nova informação e, para tal, não foi preciso perder conhecimento prévio.

O desenvolvimento da teoria ART foi o produto de uma tentativa de tentar compreender esta característica de processamento de informação cognitiva humana (WEENINK, 1997). Esta troca entre aprendizado contínuo e permanência de

memórias antigas foi estudada por Grossberg e nomeada “Teoria da Ressonância Adaptativa” (GROSSBERG, 1976a, 1976b; GROSSBERG; STONE, 1986). O levantamento do dilema que centraliza esta teoria levou à reflexão e comparação com redes clássicas do tipo *feedforward*, altamente populares na época, no qual novas informações gradualmente deterioram aprendizados antigos e, portanto, não podem ser caracterizadas como estáveis em um ambiente em constante mudança. Por consequência, redes BP *feedforwards* (WASSERMAN, 1989) necessitam de uma fase separada de treinamento e performance (desempenho/testes), enquanto a AMF, por exemplo, pode trabalhar em tempo real com treinamento continuado e performance (WEENINK, 1997).

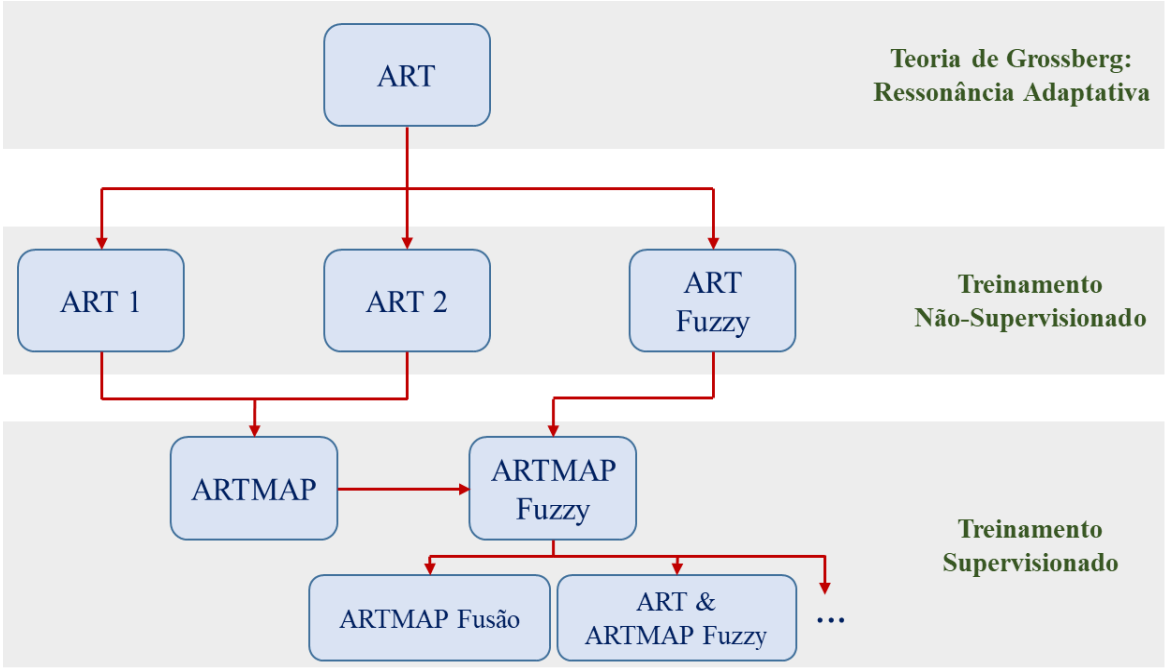
Para caráter de compreensão, pode-se fazer um paralelo com a temática de oscilações em um sistema físico. Em um sistema arbitrário, quando uma determinada entrada possui a particularidade de ter a mesma frequência de ressonância da estrutura, aumenta-se a amplitude da oscilação e se diz que ocorreu o fenômeno de ressonância. Em outras palavras, esta ocorrência apenas estará presente quando a entrada for altamente similar ao sistema no que se refere à frequência de oscilação. O mesmo pode ser parafraseado para a teoria ART: a rede neural sistematicamente aprende e direciona a entrada arbitrária a categorias específicas baseando-se no grau de similaridade e pertinência que os dois componentes possuem. Assim, quanto maior for a similaridade entre uma entrada e uma dada categoria, maior a chance de ocorrer “ressonância” e serem vinculadas (AMORIM, 2006).

A fim de estabelecer um paralelo autêntico com o comportamento biológico, a ênfase geral das redes ancoradas na teoria ART estrutura-se no aprendizado não-supervisionado e na auto-organização. Esta característica de treinamento corresponde à base e substrato para outros tipos que a sucederam, incluindo o aprendizado supervisionado (WEENINK, 1997). Em um paralelo com os sistemas biológicos, podemos afirmar que “aprender” sempre se inicia de maneira não-supervisionada durante a fase de crescimento de um ser humano. As primeiras associações correspondem à categorização de estímulos vinculados a apenas padrões de entrada, uma vez que recém-nascidos não possuem categorias de associação pré-existentes. Aos poucos, conforme ocorre o desenvolvimento, cada associação vai sendo atrelada a uma determinada correspondência (*feedback*), uma característica do aprendizado supervisionado.

Estas discussões a respeito dos conceitos mais centrais da teoria ART levaram à elaboração de uma série de modelos de redes neurais ART para reconhecimento de padrões, classificação e predição de eventos. O progresso constante dos estudos na área possibilitou o aparecimento de diversas variações de arquiteturas, cada qual com uma finalidade específica e benefício diferenciado no tratamento de dados ou análises pós-treinamento. Em suma, as redes ART têm por base o reconhecimento de padrões adaptativos de maneira estável em resposta a entradas arbitrárias em tempo real (GROSSBERG, 1988).

A Figura 1 explicita algumas das redes vinculadas à família ART, fundamentadas na teoria da ressonância adaptativa e que comumente são exploradas em tarefas de classificação, aproximação e predição.

Figura 1 - Esquematização de algumas redes vinculadas à família ART.



Fonte: Elaboração do próprio autor.

Observa-se que várias redes neurais foram elaboradas como descendentes do módulo básico. Naturalmente, outras podem ser concebidas, bastando apenas incluir uma inovação (LOPES, 2005). Destacam-se as principais abaixo:

a) **Rede Neural ART 1:** Possui treinamento não-supervisionado. Capacidade de reconhecer apenas padrões de entrada binários de maneira arbitrária (CARPENTER; GROSSBERG, 1987a);

b) **Rede Neural ART 2:** Treinamento não-supervisionado, com o acréscimo de lidar tanto com padrões de entrada binários como analógicos (CARPENTER; GROSSBERG, 1987b);
c) **Rede Neural ART *Fuzzy*:** Treinamento não-supervisionado. Ancora-se nos preceitos de lógica *fuzzy* (ZADEH, 1965) para os cálculos, admitindo entradas binárias e analógicas, portanto (CARPENTER; GROSSBERG; ROSEN, 1991);
d) **Rede Neural ARTMAP:** Possui treinamento supervisionado graças à adição de outro módulo ART para compor a estrutura de saída da arquitetura. Assim, é possível relacionar entradas e saídas e constituir um par treinado para possibilitar tarefas de classificação. Possui um campo de conexão entre os módulos chamado “inter-ART”. Pode identificar padrões de dados binários ou analógicos (CARPENTER; GROSSBERG; REYNOLDS, 1991);
e) **Rede Neural ARTMAP *Fuzzy*:** Treinamento supervisionado. Fundamentada na teoria e cálculos de lógica *fuzzy*, envolvendo questões como grau de pertinência de um elemento em um determinado conjunto (ZADEH, 1965) e possibilitando, também, padrões de entrada e saída analógicos ou binários (CARPENTER et al., 1992);
f) **Rede Neural ART&ARTMAP *Fuzzy:*** Arquitetura composta de duas redes neurais baseadas na teoria da ressonância adaptativa. Um módulo ART é adicionado na porção de entrada e captação dos dados, permitindo converter os dados de entrada em características binárias e efetivar um processo mais otimizado. Possibilita uma boa eficiência em problemas de predição (LOPES, 2005).

A seguir, detalhar-se-á os algoritmos e arquiteturas vinculados às duas redes que fundamentam o presente trabalho: ART *Fuzzy* e ARTMAP *Fuzzy*.

3.1 REDE NEURAL ARTIFICIAL ART *FUZZY*

A RNA descrita como ART *Fuzzy* representa uma arquitetura de treinamento não-supervisionado em que padrões de entrada apresentados sucessivamente são agrupados em categorias/classes. A rede, portanto, efetua uma auto-organização baseada na similaridade que os padrões possuem com determinada categoria,

fazendo com que o sistema generalize dados semelhantes por meio de regras abstratas (CARPENTER; GROSSBERG; ROSEN, 1991). Pode-se compreender cada agrupamento/categoria de forma geométrica, constituindo hiperretângulos que cobrem regiões distintas do espaço de entrada. De um modo geral, os principais componentes de uma rede ART consistem no subsistema de atenção e no subsistema de orientação (WEENINK, 1997).

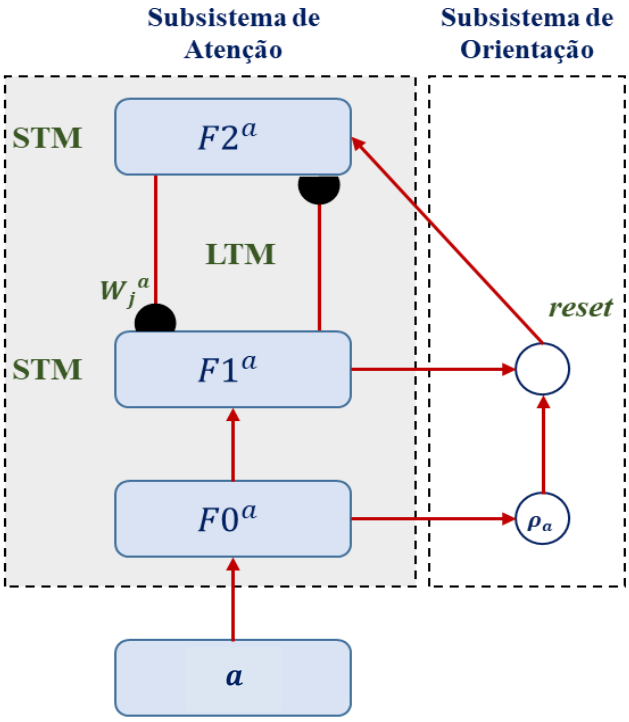
O primeiro consiste, dentre outros fatores, em dois campos (ou camadas) principais denominados *F*1 e *F*2. Estes campos são conectados por intermédio de pesos que possuem ligações do tipo *feedforward* e *feedback* (WEENINK, 1997). A conexão entre estas duas camadas forma a chamada “Memória de Longo Prazo” (LTM) do sistema, garantindo que as informações serão progressivamente armazenadas nestes componentes e permitindo perpetuar o conhecimento previamente adquirido. A camada *F*2 possui *N* unidades de processamento (neurônios) associadas às distintas categorias (*clusters*) aprendidas pela rede (AMORIM, 2006). Pode-se citar, também, a camada *F*0 adicional que corresponde ao campo vinculado ao padrão de entrada $\boldsymbol{a}$ devidamente tratado daquele determinado momento.

Por outro lado, os padrões de atividades desenvolvidos sobre os nós nas duas camadas individualmente (*F*1 e *F*2) são chamados “Memória de Curto Prazo” (STM), pois somente existem em decorrência da aplicação de uma entrada às unidades (AMORIM, 2006). Basicamente, o termo STM é associado ao padrão de atividade que se desenvolve em um campo individual conforme um novo padrão de entrada é processado, representando uma espécie de memória instantânea nas camadas (WEENINK, 1997).

Em contrapartida, o subsistema de orientação é necessário para estabilizar o processamento de STM e o aprendizado que ocorre efetivamente na LTM. É como se este subsistema fosse o responsável por gerenciar as diversas funções da arquitetura, enquanto o de atenção representa a parcela central por meio da qual a informação progride até ser efetivamente armazenada. Na porção de orientação, ocorre a presença de dois sinais de entrada e um de saída. Os de entrada são ancorados no padrão existente na camada *F*0 naquele determinado instante e na atividade correspondente em *F*1; o de saída caracteriza-se por ser o chamado *reset*. O *reset* controla o processo de “busca por coincidência” de padrões e aprendizagem na rede,

atuando como inibidor das atividades em *F*2, caso necessário (WEENINK, 1997; AMORIM, 2006). A Figura 2 esquematiza a arquitetura de uma rede neural ART.

Figura 2 - Esquematização da arquitetura de uma RNA ART *Fuzzy* típica.



Fonte: Elaboração do próprio autor.

As camadas referenciadas possuem nomeações específicas: *F*0 é o vetor de entrada atual após ser devidamente normalizado por codificação de complemento; *F*1 refere-se à camada de comparação (em que o vetor atividade será comparado com os pesos das diversas categorias presentes); e, por fim, *F*2 trata-se do nível de reconhecimento, parcela na qual os nós que constituem as categorias ficam dispostos e serão devidamente vinculados aos padrões subsequentes apresentados à rede a cada interação mediante um teste específico. Como observa-se na Figura 2, os pesos associados às conexões entre *F*1 e *F*2 são de baixo para cima (*bottom-up*) e de cima para baixo (*top-down*), fomentando uma comunicação bidirecional constante entre os níveis e constituindo a LTM apropriadamente (LOPES, 2005; AMORIM, 2006).

O teste para determinar se aquela categoria arbitrária realmente deve ser a escolhida para o padrão de entrada daquele momento é chamado **teste de vigilância**. Caso seja satisfeito, ocorre a chamada “ressonância”, e o sistema adere a nova

informação; caso contrário, força-se o chamado *reset* e uma nova categoria deve ser escolhida (CARPENTER; GROSSBERG; ROSEN, 1991).

Por fim, é interessante a ressalva de que, para fins de uso nas redes da família ART, os vetores são representados como linhas e não como colunas (habitualmente utilizados na literatura). Tal representação é empregada nas RNAs descendentes da teoria ART por se tratar de uma facilitação quanto aos cálculos e formulação das equações envolvidas no modelo (LOPES, 2005).

3.1.1 Algoritmo ART *Fuzzy*

Este tópico irá destacar, passo a passo, o panorama de como funciona o algoritmo de treinamento de uma rede neural ART *Fuzzy*. Será possível notar como cada componente se enquadra e qual sua respectiva função, além de ficar mais claro como a atualização de pesos e armazenamento de informações se sucederá.

Antes de mais nada, é interessante descrever os três parâmetros que determinam a dinâmica de uma rede AF: parâmetro de escolha, parâmetro de vigilância e a taxa de treinamento (WEENINK, 1997; CARPENTER; GROSSBERG; ROSEN, 1991):

a) **Parâmetro de escolha (α>0):** Auxilia no controle de busca das categorias presentes na camada $F2$. Assim, caso uma determinada entrada esteja vinculada a duas possíveis escolhas de categorias, este componente priorizará sempre aquela que possui menos modificações nos pesos no decorrer do treinamento (ou aquela mais específica) (CANO-IZQUIERDO et al., 2013);
b) **Parâmetro de vigilância ($\rho \in [0,1]$):** O parâmetro de vigilância verifica a combinação dos padrões de entrada e dos pesos para que ocorra a ressonância. Para um valor de ρ pequeno tem-se uma capacidade de generalização maior (mais amostras podem ser agrupadas na mesma categoria), produzindo poucas classes. Em contrapartida, se o valor de ρ é grande, a rede efetua rigorosas distinções, produzindo mais classes específicas. O estado de ressonância daquela determinada inserção se prolonga até a apresentação do novo padrão de entrada (AMORIM, 2006);

c) **Taxa de treinamento ($\boldsymbol{\beta} \in$ [0,1]):** Controla a velocidade de adaptação com que os pesos são ajustados no decorrer do treinamento. Este componente assumindo o valor 1 refere-se ao chamado “treinamento rápido”. Para tratamento com conjuntos de entrada com ruídos, comumente pode-se empregar a opção de rápido comprometimento e lenta recodificação, sendo β=1 para atribuição das primeiras categorias e, depois de tal, assumindo valor menor que 1 (CARPENTER et al., 1992).

A Tabela 1 a seguir mostra uma síntese conceitual dos parâmetros que ditam a dinâmica da rede ART *Fuzzy*.

Tabela 1 - Síntese dos parâmetros da rede ART *Fuzzy*.

Denominação	Simbologia	Fundamento Conceitual
Parâmetro de Escolha	α>0	Controle de busca de categorias vencedoras em *F*2. Prioriza a escolha da categoria com menos modificações.
Parâmetro de Vigilância	$\rho \in$ [0,1]	Grau de especificidade para atribuição de uma amostra a uma categoria. Quanto maior ρ, mais categorias são criadas pela rede.
Taxa de Treinamento	$\beta \in$ [0,1]	Velocidade de adaptação com que os pesos são ajustados. No treinamento rápido, β = 1.

Fonte: Elaboração do próprio autor.

Também é importante destacarmos, antes da descrição efetiva do algoritmo, a representação geral que será empregada para os vetores de atividade em cada camada:

a) **Vetor atividade *F*0:** porção na qual ocorrerá a adequação do vetor de entrada daquele determinado instante. Inicialmente, tal vetor é denotado por $\boldsymbol{I} = (I_1, \dots, I_M)$, e cada componente I_i pertence ao intervalo [0,1], $i = 1, \dots, M$. A variável M representa a dimensão do vetor de entrada (quantidade de elementos). Esta dimensão irá dobrar após a codificação de complemento;

b) **Vetor atividade *F*1:** vincula-se ao vetor que representa a camada central, de comparação, e é denotado por $\boldsymbol{x} = (x_1, \dots, x_M)$;

c) **Vetor atividade *F*2:** vincula-se ao vetor que representa a camada superior, de reconhecimento, relacionando-se com o código ativo ou categoria ativa

naquele determinado instante, sendo denotado por $\boldsymbol{y} = (y_1, \dots, y_N)$ (CARPENTER et al., 1992).

O vetor de pesos da rede (assim como sua inicialização) é descrito em uma das etapas do algoritmo de funcionamento da rede ART. Explicita-se, a seguir, um passo a passo para implementação de tal.

1) Leitura dos parâmetros da rede

Os parâmetros mencionados anteriormente (Tabela 1) devem ser adotados inicialmente e lidos pela rede durante a fase inicial. Estes atributos determinarão as diversas características e o progresso do treinamento como um todo.

2) Vetor de pesos: inicialização

Associado a cada nó de categoria *j* da camada *F*2 ($j = 1, \dots, N$) vincula-se um vetor $\boldsymbol{w_j}$ de pesos adaptativos, ou os chamados "traços de memória de longo prazo". Inicialmente, como retratado na Equação (1) abaixo:

$$w_{j1}(0) = \cdots = w_{j2M}(0) = 1 \qquad (1)$$

e cada categoria é dita não estar atribuída. Depois de uma categoria ser escolhida para ser relacionada a um padrão de entrada, esta é dita "atribuída" ou "modificada". Uma vez que os pesos se iniciam em 1, o sistema acaba possuindo a característica de ser monotonicamente decrescente e, por isso, converge para um limite (CARPENTER et al., 1992).

3) Normalização da entrada por codificação de complemento

A proliferação de categorias é evitada normalizando as entradas por codificação de complemento. Este método específico vincula a cada vetor a parcela *on* e *off* correspondente, favorecendo a assimilação e preservando a amplitude

(CARPENTER et al., 1992). A normalização pode ser obtida por meio do pré-processamento de cada vetor $\boldsymbol{a}$, configurando tal como explicitado na Equação (2):

$$\overline{\boldsymbol{a}} = \frac{\boldsymbol{a}}{|\boldsymbol{a}|} \qquad (2)$$

sendo:

$|\boldsymbol{a}| = \sum_i |a_i|$;

$\overline{\boldsymbol{a}}$ = vetor de entrada normalizado.

O algoritmo utiliza a codificação de complemento para estruturar a parcela final desta normalização. O problema de proliferação de categorias pode ocorrer em certos sistemas ART quando um grande número de entradas pode "erodir" a norma dos vetores pesos. Este problema é resolvido recorrendo a esta técnica de codificação de complemento. É um mecanismo que preserva a amplitude da informação e representa tanto a resposta *on* como a resposta *off* para um certo vetor de entrada. Isso significa que a entrada é comparada a cada categoria com ambos os graus de presença e ausência das características possivelmente similares apresentadas, aumentando a assertividade e evitando proliferação (CARPENTER et al., 1992). A resposta *off* mencionada para a resposta *on* $\boldsymbol{a}$ é apresentada abaixo:

$$\overline{a}_i^c = 1 - \overline{a}_i \qquad (3)$$

Desta forma, a entrada "oficial" que corresponderá ao vetor $\boldsymbol{I}$ devidamente normalizado e com codificação de complemento será dada por:

$$\boldsymbol{I} = (\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}}^{\boldsymbol{c}}) = (\overline{a}_1, \dots, \overline{a}_M, \overline{a}_1^c, \dots, \overline{a}_M^c) \qquad (4)$$

Observa-se que a vetor passa a ser 2M dimensional e, além disso, a amplitude será preservada, já que a norma de $\boldsymbol{I}$ será dada por (CARPENTER et al., 1992):

$$|\boldsymbol{I}| = \sum_{i=1}^{M} \overline{a}_i + \left(M - \sum_{i=1}^{M} \overline{a}_i \right) = M \qquad (5)$$

4) Escolha da categoria

Para cada entrada $\boldsymbol{I}$ e nó j de *F*2, a função de escolha T_j é definida por:

$$T_j(\boldsymbol{I}) = \frac{|\boldsymbol{I} \wedge \boldsymbol{w}_j|}{\alpha + |\boldsymbol{w}_j|} \quad (6)$$

em que:

∧ = operador *fuzzy* da operação *AND*;

$(\boldsymbol{p} \wedge \boldsymbol{q})_i = \min(p_i, q_i)$.

O índice J indica a categoria selecionada naquele instante, sendo:

$$T_J = max\{T_j : j = 1, \dots, N\} \quad (7)$$

Se ocorrer a situação de mais de um T_j ser máximo, a categoria j com o menor índice será a escolhida. Em um sistema de escolha deste tipo, o vetor atividade $\boldsymbol{x}$, de *F*1, obedece à seguinte regra (CARPENTER et al., 1992):

$$\boldsymbol{x} = \begin{cases} \boldsymbol{I}, se\ F2\ está\ inativo \\ \boldsymbol{I} \wedge \boldsymbol{w}_J, se\ o\ J-ésimo\ nó\ de\ F2\ for\ escolhido \end{cases} \quad (8)$$

5) Ressonância ou *Reset*

O processo de ressonância ocorrerá se a categoria escolhida atender o chamado parâmetro de vigilância (ρ). Em outras palavras, a função de *match* tem de ser satisfeita, traduzida pela seguinte desigualdade:

$$\frac{|\boldsymbol{I} \wedge \boldsymbol{w}_J|}{|\boldsymbol{I}|} \geq \rho \quad (9)$$

A expressão (9) permite que avalie-se o quanto aquela determinada entrada e categoria pré-selecionada são parecidas, ficando o rigor desta comparação com o parâmetro de vigilância. Quanto maior ρ, maior a “rigorosidade” e, por consequência, maior será a similaridade necessária para que a desigualdade seja satisfeita (o que culminará na criação de mais categorias específicas) (WEENINK, 1997).

O *reset* ocorrerá caso a expressão (9) não seja obedecida. Neste caso, o valor da função de escolha T_J é fixado em 0 durante toda a duração de apresentação deste determinado vetor de entrada, impossibilitando a seleção da mesma categoria durante as próximas tentativas (CARPENTER et al., 1992). Escolhe-se uma nova categoria e o processo de busca permanece até um índice J atender (9). Para designar que um determinado nó foi escolhido e está ativo, utiliza-se (10) (CARPENTER et al., 1992):

$$y_j = \begin{cases} 1, se\ j = J \\ 0, se\ j \neq J \end{cases} \quad (10)$$

Vale ressaltar que **entre** as apresentações de cada padrão de entrada durante o treinamento, os vetores $\boldsymbol{y}$ e $\boldsymbol{x}$ são configurados para **0** novamente.

6) Aprendizado e atualização dos pesos

Após tais etapas, o vetor peso $\boldsymbol{w}_J$ é atualizado de acordo com a equação:

$$\boldsymbol{w}_J^{novo} = \beta\left(\boldsymbol{I} \wedge \boldsymbol{w}_J^{velho}\right) + (1-\beta)\boldsymbol{w}_J^{velho} \quad (11)$$

A Figura 3 mostra o fluxograma da síntese do algoritmo da rede AF.

3.1.2 Breve Análise da Representação Geométrica das Categorias

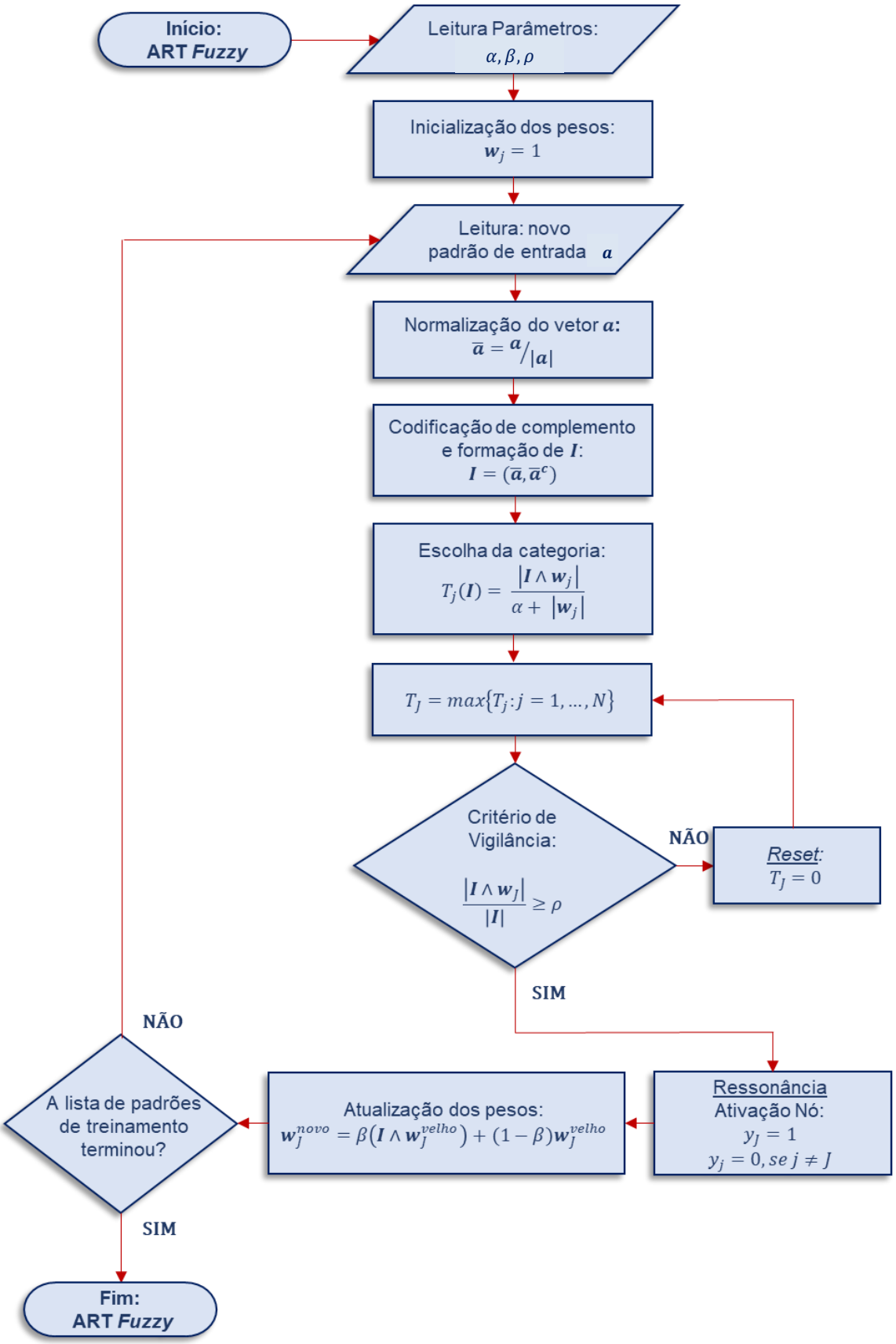
Para uma determinada categoria j, podemos representar o vetor de peso associado como (CANO-IZQUIERDO et al., 2013):

$$\boldsymbol{w}_j = (\boldsymbol{u}_j, \boldsymbol{v}_j) \quad (12)$$

Primeiramente, analisaremos $\boldsymbol{u}_j$ e $\boldsymbol{v}_j$ representando vetores unidimensionais. Para o treinamento rápido, a Equação (11) assumirá (CARPENTER et al., 1992):

$$\boldsymbol{w}_J^{novo} = \left(\boldsymbol{I} \wedge \boldsymbol{w}_J^{velho}\right) \quad (13)$$

Figura 3 - Fluxograma do algoritmo de funcionamento da ART *Fuzzy*.



Fonte: Elaboração do próprio autor.

Considerando a entrada como o exposto anteriormente ($\boldsymbol{I} = (\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}}^{\boldsymbol{c}})$), nota-se claramente que, para um nó J inicialmente sem vínculo, o primeiro ajuste de pesos se dará por:

$$\boldsymbol{w}_J^{novo} = \boldsymbol{I} = (\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}}^{\boldsymbol{c}}) \quad (14)$$

Assim, vamos considerar inicialmente que n entradas selecionaram a categoria J no treinamento. Comparando com os componentes que representam o vetor $\boldsymbol{w}_j$ na Equação (12), temos (CANO-IZQUIERDO et al., 2013):

$$\boldsymbol{u}_j = min\{\boldsymbol{a}^n\} \quad (15)$$

$$\boldsymbol{v}_j = min\{(1 - \boldsymbol{a}^n)\} = \mathbf{1} - max\{\boldsymbol{a}^n\} \quad (16)$$

em que o valor $min\{\boldsymbol{a}^n\}$ se refere ao mínimo possível para cada componente do vetor (componente *fuzzy AND*).

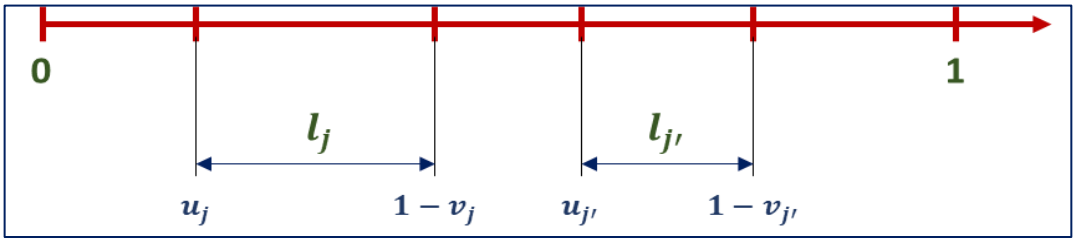
Como pode-se observar por estas últimas expressões, sugere-se que o vetor peso associado a uma determinada categoria define um **hiperretângulo** (*hyperbox*), com canto inferior definido por $\boldsymbol{u}_j$ (mínimo valor de todas as entradas $\boldsymbol{a}^n$), e canto superior definido por $1 - \boldsymbol{v}_j$ (máximo valor de todas as entradas $\boldsymbol{a}^n$). No treinamento rápido, cada interação que adeque os pesos de uma certa categoria faz com que o tamanho deste hiperretângulo se altere para o mínimo possível que englobe o novo ponto absorvido. Ou seja: conforme o valor do retângulo (ou hiperretângulo, para dimensões maiores) vai aumentando para englobar os novos pontos, o valor dos pesos vinculados àquela determinada categoria vai diminuindo monotonicamente (CARPENTER et al., 1992; CANO-IZQUIERDO et al., 2013).

Considerando o caso de uma dimensão, cada categoria será descrita por um segmento representado por $\left(u_j, 1 - v_j\right) \in [0,1]$. O tamanho do segmento será, portanto:

$$\begin{gathered} \left|l_j\right| = max_n\{a^n\} - min_n\{a^n\} = 1 - \left(u_j + v_j\right) \\ \left|l_j\right| = 1 - \left|\boldsymbol{w}_j\right| \end{gathered} \quad (17)$$

Esta última relação nos mostra como o tamanho das categorias geradas tem necessariamente vínculo com os pesos atrelados a elas. Uma representação é apresentada na Figura 4.

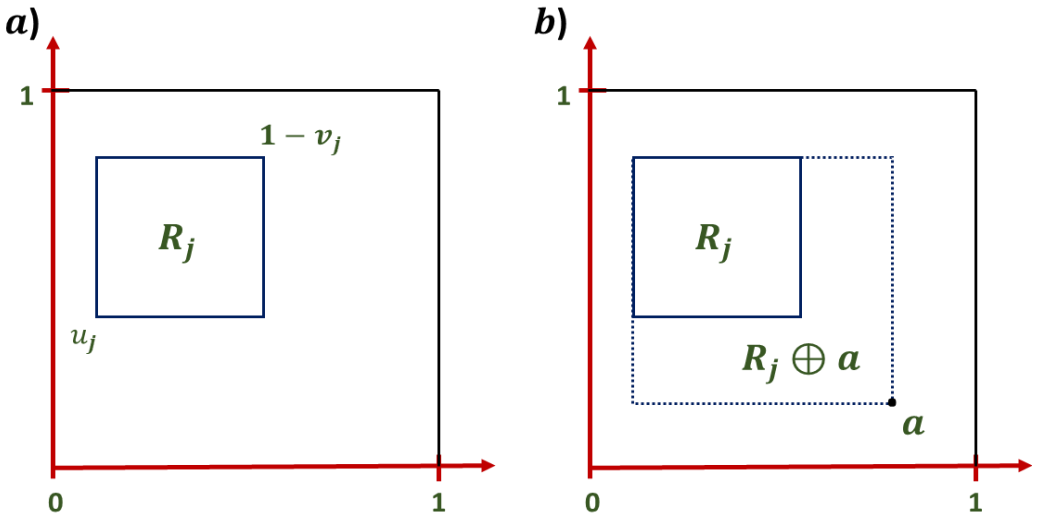
Figura 4 - Representação das categorias geradas para o caso unidimensional.



Fonte: Adaptado de Cano-Izquierdo et al. (2013).

Esta análise pode ser estendida para o caso de M dimensões, no qual o tamanho do hiperretângulo é dado por $|R_j| = M - |\boldsymbol{w}_j|$ (CARPENTER et al., 1992). Considerando o treinamento rápido e o vetor de duas dimensões, por exemplo, os cantos do retângulo $\boldsymbol{R}_j$, que caracterizará aquela determinada categoria, serão dados por $\boldsymbol{u}_j$ e $1 - \boldsymbol{v}_j$, tal como descrito anteriormente. Estes cantos referem-se ao inferior esquerdo e superior direito, respectivamente, pois dessa forma pode-se compreender claramente o tamanho da abrangência geométrica de cada um. Durante cada iteração, o tamanho de $\boldsymbol{R}_j$ se expande para o menor retângulo que contenha as informações anteriores ($\boldsymbol{R}_j$ anterior) e o novo ponto que foi "aprovado" para esta classe.

Figura 5 - Representação geométrica de uma categoria no caso bidimensional.



Fonte: Elaboração do próprio autor.

A cada iteração, utilizando as equações (5), (9) e (11) (CARPENTER et al., 1992):

$$\left|\boldsymbol{w}_j\right| \geq \rho M \qquad (18)$$

Por fim, sabendo que $\left|R_j\right| = M - \left|\boldsymbol{w}_j\right|$:

$$\left|\boldsymbol{R}_j\right| \leq (1-\rho)M \qquad (19)$$

É possível observar, por fim, como o parâmetro de vigilância pode influenciar, na prática, o tamanho do retângulo representativo gerado. Maior o parâmetro, menor o retângulo gerado e mais específico o mesmo será, fomentando a criação de mais categorias. Menor o parâmetro, uma capacidade maior de generalização será possível, e menos categorias serão geradas.

3.2 REDE NEURAL ARTIFICIAL ARTMAP *FUZZY*

ARTMAP corresponde a uma classe de RNAs que desenvolvem aprendizado supervisionado incremental de reconhecimento de categorias e mapas multidimensionais em resposta a vetores de entrada apresentados em ordem arbitrária (CARPENTER et al., 1992). As redes possuem características de serem auto organizáveis, o que significa que o sistema deve ser capaz de desenvolver categorias de reconhecimento estáveis em tempo real (WEENINK, 1997).

A rede AMF diferencia-se da ARTMAP convencional por possuir seus cálculos baseados na lógica *fuzzy* (ZADEH, 1965). Isto favorece, portanto, a capacidade deste sistema em classificar entradas mediante um conjunto *fuzzy* de características sempre com referência a valores entre 0 e 1, indicando o quanto uma determinada entrada pertence àquela categoria (conceito de função de pertinência) (ZADEH, 1965). A fim de possibilitar a implementação desta estrutura, utiliza-se dois módulos ART *Fuzzy* interconectados por uma região chamada “módulo inter-ART”. Os dois módulos ART são nomeados como ART_a e ART_b, cada um correspondendo ao processamento de uma parte do par treinado: enquanto o ART_a destina-se ao recebimento do padrão de entrada, o ART_b fica responsável por processar a associação correta de tal padrão (ou

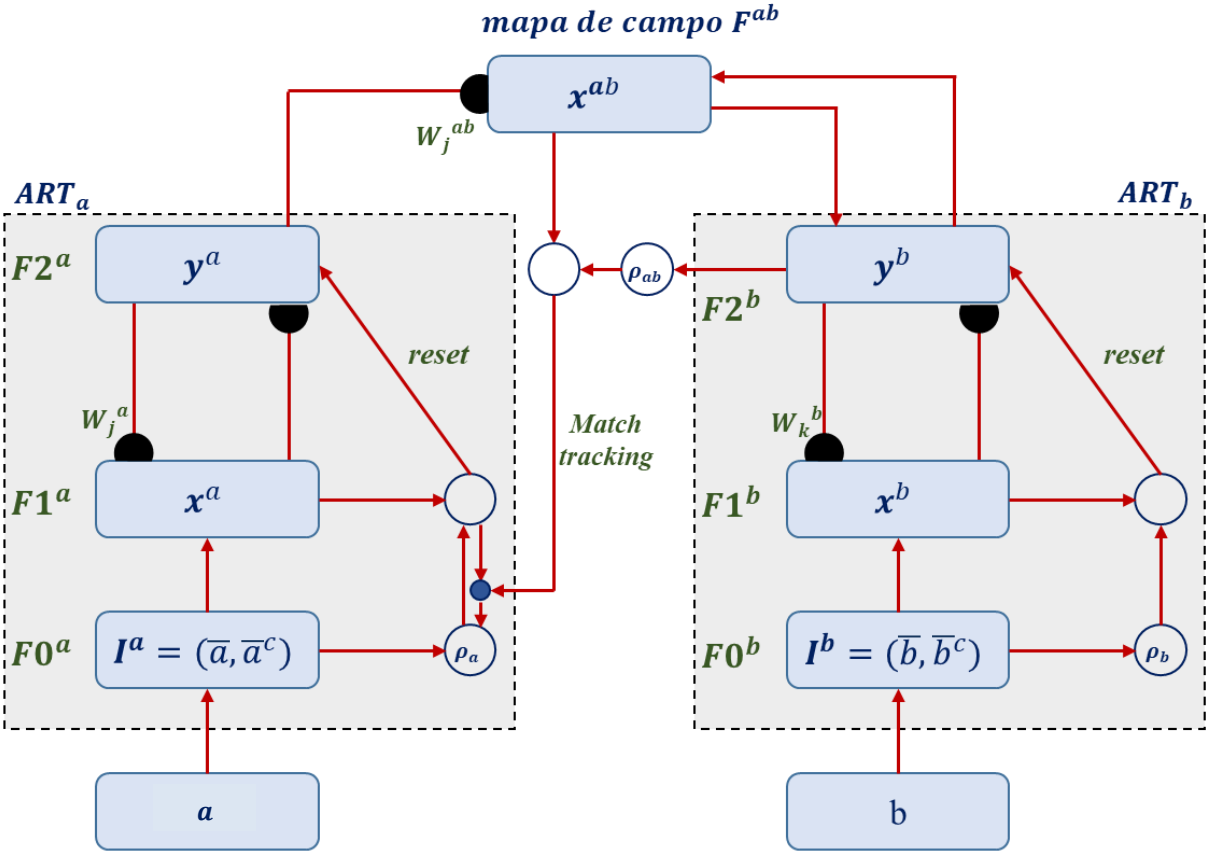
“saída desejada”). A interconexão é feita por meio de um mapeamento no módulo inter-ART por intervenção de um mecanismo conhecido como *match tracking*, no qual vincular-se-á cada categoria da entrada a uma categoria da saída (CARPENTER et al., 1992; AMORIM, 2006).

O controle interno, caracterizado pelo inter-ART, é designado para criar um número mínimo de categorias de reconhecimento ART_a, possibilitando maximizar a generalização, minimizando o erro de previsão/classificação. Sempre que a rede faz um prognóstico errado via uma conexão associativa instruída, o parâmetro de vigilância ρ_a do módulo ART_a será elevado em uma quantidade mínima necessária a corrigir o erro usando o mecanismo *match tracking*. Inicia-se uma nova busca de categoria para a entrada daquele certo instante, até que todos os critérios sejam satisfeitos e crie-se uma conexão correta entre ambos os módulos (CARPENTER et al., 1992).

Em suma, a arquitetura e operação da AMF segue a lógica do descrito para a ART *Fuzzy*, mas com o acréscimo de um novo módulo (que representará o *feedback* ou resposta associativa correta para a entrada) e do campo de mapeamento central (módulo inter-ART). Este último será responsável pelas associações preditivas entre entrada e saída, reorganizando a estrutura das categorias caso ocorra um erro preditivo e favorecendo um *match* posterior mais adequado (CARPENTER et al., 1992). Uma esquematização da arquitetura da AMF é apresentada na Figura 6. Como observado, verifica-se que o módulo inter-ART forma associações entre as categorias de ART_a com as unidades de F^{ab}. Estas unidades representam as categorias em $F2^b$, uma vez que as conexões entre $F2^b$ e F^{ab} possuem pesos com valor 1 (AMORIM, 2006).

Em suma, o módulo ART_a pré-processa, por codificação de complemento, o vetor $\boldsymbol{a}$ em um vetor $\boldsymbol{I}^a$ de dimensão $2M_a$ na camada $F0^a$. Deste modo, $\boldsymbol{I}^a$ se torna o vetor de entrada para a camada $F1^a$ de ART_a. A mesma lógica ocorre para o módulo pertencente ao processamento da saída desejada (ART_b). Quando a predição de ART_a é inviabilizada em ART_b, a inibição do vínculo ocorre pela ativação do mapa de campo, induzindo o mecanismo de *match tracking*. Este processo aumenta o parâmetro de vigilância de ART_a.até um mínimo necessário para que a categoria relacionada à entrada não seja escolhida novamente, favorecendo a seleção de outra que possa “casar” melhor com a configurada na saída (CARPENTER et al., 1992).

Figura 6 - Esquematização da arquitetura de uma RNA ARTMAP *Fuzzy* típica.



Fonte: Elaboração do próprio autor.

As camadas $F2^a$ e F^{ab}, como observado na Figura 6, são conectadas por pesos que serão ajustados via treinamento. Em contrapartida, nos módulos $F2^b$ e F^{ab}, cada nó (categoria formada) de $F2^b$ relaciona-se com seu nó equivalente em F^{ab} (LOPES, 2005). Em outras palavras, os pesos ajustados no módulo inter-ART, durante o processo de treinamento, vinculam diretamente a parcela de entrada (ART_a) com a parcela de saída (ART_b) por meio do mapa de campo. Isto ocorre porque para cada nó j da camada $F2^a$ estão relacionados os nós k existentes em $F2^b$, tudo isso traduzido pela matriz $\boldsymbol{w}_j^{ab}$.

Quando o processo ocorre com perfeita ressonância em todas as etapas, a adaptação dos pesos entre a categoria ativa J do ART_a e a categoria ativa K do ART_b é realizada de forma que a conexão entre tais terá valor 1 e todas as demais (referentes ao mesmo nó ART_a) terão valor 0 (KARTALOPOULOS, 1995). Isso significa que, quando ocorre um prognóstico correto, aquela determinada categoria selecionada em ART_a deverá ser vinculada **somente** com aquela determinada categoria selecionada em ART_b. Isto representa o coração do conceito de treinamento

supervisionado da rede: para cada entrada, é associada uma saída e, posteriormente, pode-se obter resultado para uma entrada desconhecida mediante a generalização que foi efetuada.

3.2.1 Algoritmo ARTMAP *Fuzzy*

Neste tópico será apresentado um passo a passo para utilização da rede AMF. Os conceitos seguem bastante a lógica do exposto no tópico 3.1.1, uma vez que se trata de uma composição de mais de um módulo ART. A novidade fica por conta das interações no mapa de campo, parcela importante e que também fará parte do mecanismo de adaptação dos pesos. Primeiramente, no entanto, é interessante definirmos alguns componentes e suas respectivas dimensões a fim de facilitar a compreensão do algoritmo:

a) **Entrada do ART_a (I^a):** Parcela vinculada ao módulo ART_a, devidamente normalizada por codificação de complemento. $\boldsymbol{I}^a = (\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}}^{\boldsymbol{c}})$. Representa a entrada "oficial" com dimensão $2M_a$ que vincula-se à camada $F0^a$;
b) **Entrada do ART_b (I^b):** Parcela vinculada ao módulo ART_b, devidamente normalizada por codificação de complemento. Corresponde à associação correta da entrada daquele dado instante. $\boldsymbol{I}^b = (\overline{\boldsymbol{b}}, \overline{\boldsymbol{b}}^{\boldsymbol{c}})$. Representa a entrada "oficial" com dimensão $2M_b$ que vincula-se à camada $F0^b$;
c) **Vetor atividade $F1^a$:** Corresponde ao vetor da camada de comparação do módulo ART_a. É denotado por $\boldsymbol{x}^a = (x_1^a, \dots, x_{2M_a}^a)$;
d) **Vetor atividade $F1^b$:** Corresponde ao vetor da camada de comparação do módulo ART_b. É denotado por $\boldsymbol{x}^b = (x_1^b, \dots, x_{2M_b}^b)$;
e) **Vetor de saída de F^{ab}:** Representa o vetor atividade inerente ao mapa de campo da estrutura. Denotado por $\boldsymbol{x}^{ab} = (x_1^{ab}, \dots, x_{N_b}^{ab})$;
f) **Vetor de saída ART_a:** Trata-se do vetor da camada mais extrema do módulo ART_a, designada pela camada de reconhecimento ($F2^a$). É denotada por $\boldsymbol{y}^a = (y_1^a, \dots, y_{N_a}^a)$ e possui dimensão N_a;
g) **Vetor de saída ART_b:** Vetor da camada de reconhecimento ($F2^b$) do módulo ART_b. É denotado por $\boldsymbol{y}^b = (y_1^b, \dots, y_{N_b}^b)$ e possui dimensão N_b.

1) Leitura dos parâmetros da rede

Tal como no tópico 3.1.1, os parâmetros da rede devem ser especificados a fim de iniciar o processo de treinamento. Deve-se definir o parâmetro de escolha ($\alpha > 0$), a taxa de treinamento ($\beta \in [0,1]$) e os parâmetros de vigilância dos módulos ART_a, ART_b e inter-ART (ρ_a, ρ_b *e* ρ_{ab}, respectivamente, de intervalo $[0,1]$). Outro componente se refere ao ε, que corresponderá à variação mínima que acrescentaremos ao parâmetro ρ_a inicial (também chamado *baseline*) quando não ocorrer casamento entre as categorias selecionadas (CARPENTER et al., 1992).

2) Vetores pesos e inicialização geral

Cada módulo ART possui vetores peso específicos que devem ser configurados no início do processo de treinamento. Para o módulo ART_a denota-se $\boldsymbol{w}_j^a = (w_{j1}^a, \dots, w_{j2M_a}^a)$ o vetor equivalente à j-ésima categoria. Para o ART_b, tem-se $\boldsymbol{w}_k^b = (w_{k1}^b, \dots, w_{k2M_b}^b)$. Por fim, para o mapa de campo, o vetor de peso que representa o j-ésimo nó de $F2^a$ em relação a F^{ab} é denotado por $\boldsymbol{w}_j^{ab} = (w_{j1}^{ab}, \dots, w_{jN_b}^{ab})$ (CARPENTER et al., 1992). Como pode-se observar, este último vetor de pesos possui dimensão de N_b, representando o total de categorias existentes no módulo ART_b. Ou seja: para cada nó *j* do ART_a, existe uma ligação individual vinculando cada nó *k* do ART_b.

Inicialmente, todos os pesos descritos acima devem ser configurados para assumirem valor 1, tal como descrito no algoritmo ART *Fuzzy* e na Equação (1):

$$w_{j1}^a(0) = \cdots = w_{j2M_a}^a(0) = 1$$
$$w_{k1}^b(0) = \cdots = w_{k2M_b}^b(0) = 1$$

Os pesos relacionados ao módulo inter-ART também devem ser inicializados em 1 e, conforme as categorias vão sendo associadas, os mesmos são ajustados e criam vínculo permanente entre elas (CARPENTER et al., 1992):

$$w_{j1}^{ab}(0) = \cdots = w_{jN_b}^{ab}(0) = 1 \qquad (20)$$

Ressalta-se que, **entre** as sucessivas apresentações de entrada, os vetores $\boldsymbol{x}^a$, $\boldsymbol{y}^a$, $\boldsymbol{x}^b$, $\boldsymbol{y}^b$ e $\boldsymbol{x}^{ab}$ são sempre reconfigurados para 0.

3) Escolha das categorias

Para cada módulo individual (ART_a *e* ART_b), calcula-se a categoria vencedora utilizando a mesma expressão representada pela Equação (6). Detalhando:

$$T_j^a(\boldsymbol{I}^a) = \frac{\left|\boldsymbol{I}^a \wedge \boldsymbol{w}_j^a\right|}{\alpha + \left|\boldsymbol{w}_j^a\right|} \qquad (21)$$

$$T_k^b(\boldsymbol{I}^b) = \frac{\left|\boldsymbol{I}^b \wedge \boldsymbol{w}_k^b\right|}{\alpha + \left|\boldsymbol{w}_k^b\right|} \qquad (22)$$

Tal como mostrado na Equação (7), as categorias escolhidas corresponderão aos índices que possuírem maior valor obtido nas expressões (denotados como *J* e *K*). Ressalta-se novamente que, para valores iguais obtidos em diferentes índices, o menor sempre será priorizado na escolha (CARPENTER et al., 1992).

4) Ressonância ou *Reset*

Para um determinado par de entrada/saída, a categoria vencedora obtida em ART_b (e que satisfaça ao critério de ressonância) não se alterará no decorrer do processo. A única possibilidade de alteração se dá na estrutura fundamentada no ART_a: é aqui que o parâmetro de vigilância pode ser aumentado caso não ocorra casamento, pelo mapa de campo, entre as categorias vencedoras dos dois módulos (CARPENTER et al., 1992). As expressões abaixo descrevem a condição para ressonância em cada um:

$$\frac{\left|\boldsymbol{I}^a \wedge \boldsymbol{w}_J^a\right|}{|\boldsymbol{I}^a|} \geq \rho_a \qquad (23)$$

$$\frac{\left|\boldsymbol{I}^b \wedge \boldsymbol{w}_K^b\right|}{|\boldsymbol{I}^b|} \geq \rho_b \qquad (24)$$

Passado pelo critério de vigilância, portanto, a categoria K vencedora no módulo ART_b será ativada e assim permanecerá até o fim desta iteração:

$$y_k^b = \begin{cases} 1, se\ k = K \\ 0, se\ k \neq K \end{cases} \quad (25)$$

Conforme a categoria J do módulo ART_a também passe pelo teste de vigilância e seja momentaneamente ativada (Equação (10)), é necessário que se verifique o casamento com a categoria K de ART_b. Este papel é desempenhado pelo mapa de campo atendendo ao teste de vigilância vinculado a ele:

$$\frac{\left|\boldsymbol{y^b} \wedge \boldsymbol{w}_J^{ab}\right|}{|\boldsymbol{y}^b|} \geq \rho_{ab} \quad (26)$$

em que $\left|\boldsymbol{y^b} \wedge \boldsymbol{w}_J^{ab}\right| = \left|\boldsymbol{x^{ab}}\right|$ quando existe nó ativo J e nó ativo K.

Caso a expressão (26) **não** seja obedecida, o parâmetro ρ_a deve ser acrescido até ser levemente superior ao lado esquerdo da inequação apresentada em (23):

$$\rho_a = \frac{\left|\boldsymbol{I}^a \wedge \boldsymbol{w}_J^a\right|}{|\boldsymbol{I}^a|} + \varepsilon \quad (27)$$

Isso fará com que uma nova busca seja efetuada no módulo ART_a e que torne-se impossível selecionar a vencedora anterior, ativando uma outra categoria que adeque-se de forma superior (CARPENTER et al., 1992). O valor adotado para o parâmetro no início de cada iteração é chamado "*baseline*", indicando quanto ρ_a vale antes do acréscimo mencionado.

5) Aprendizagem

Após todas as confirmações ocorrerem efetivamente, os pesos de cada módulo devem ser ajustados. As equações abaixo traduzem esta informação (AMORIM, 2006; CARPENTER et al., 1992):

$$\boldsymbol{w}_J^{anovo} = \beta\left(\boldsymbol{I}^a \wedge \boldsymbol{w}_J^{avelho}\right) + (1-\beta)\boldsymbol{w}_J^{avelho} \quad (28)$$

$$\boldsymbol{w}_K^{bnovo} = \beta\left(\boldsymbol{I}^b \wedge \boldsymbol{w}_K^{bvelho}\right) + (1-\beta)\boldsymbol{w}_K^{bvelho} \quad (29)$$

E, para o módulo inter-ART, os pesos se ajustam conforme as categorias vencedoras vão sendo obtidas:

$$w_{jk}^{ab} = \begin{cases} 1, se\ j = J\ e\ k = K \\ 0, se\ j \neq J\ ou\ k \neq K \end{cases} \quad (30)$$

O processo se repete na apresentação de um novo padrão de entrada.

3.2.2 Síntese – Algoritmo de Treinamento AMF

1) Determinação de todos os parâmetros da rede: $\rho_a, \rho_b, \rho_{ab}$, β, α e ε;
2) Inicialização de pesos: todos os componentes de $\boldsymbol{w}_j^a$, $\boldsymbol{w}_k^b$ e $\boldsymbol{w}_j^{ab}$.valem 1;
3) Verificação: a lista de padrões de treinamento terminou? Se sim, concluir o treinamento. Se não, continuar e configurar para zero $\boldsymbol{x}^a$, $\boldsymbol{y}^a$, $\boldsymbol{x}^b$, $\boldsymbol{y}^b$ e $\boldsymbol{x}^{ab}$;
4) Normalização e codificação de complemento para a entrada atual: $\boldsymbol{I}^a = (\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}}^{\boldsymbol{c}})$ e $\boldsymbol{I}^b = \left(\overline{\boldsymbol{b}}, \overline{\boldsymbol{b}}^{\boldsymbol{c}}\right)$;
5) Escolhe-se a categoria vencedora de ART_b: calcula-se cada $T_k^b\left(\boldsymbol{I}^b\right) = \frac{|\boldsymbol{I}^b \wedge \boldsymbol{w}_k^b|}{\alpha + |\boldsymbol{w}_k^b|}$ e verifica-se qual índice K refere-se ao máximo;
6) Observa-se se o critério de vigilância $\frac{|\boldsymbol{I}^b \wedge \boldsymbol{w}_K^b|}{|\boldsymbol{I}^b|} \geq \rho_b$ é satisfeito. Se sim, segue-se para o passo seguinte. Se não: *reset* em ART_b → $T_K^b = 0$ e volta-se a 5);
7) Com o nó vencedor obtém-se o vetor atividade em $F2^b$: $\boldsymbol{y}^b = (y_1^b, \ldots, y_{N_b}^b)$ → componente vale 1 apenas para a categoria vencedora;
8) Atualiza-se os pesos de ART_b: $\boldsymbol{w}_K^{bnovo} = \beta\left(\boldsymbol{I}^b \wedge \boldsymbol{w}_K^{bvelho}\right) + (1-\beta)\boldsymbol{w}_K^{bvelho}$;
9) Escolhe-se a categoria vencedora de ART_a: calcula-se cada $T_j^a(\boldsymbol{I}^a) = \frac{\left|\boldsymbol{I}^a \wedge \boldsymbol{w}_j^a\right|}{\alpha + \left|\boldsymbol{w}_j^a\right|}$ e verifica-se qual índice J refere-se ao máximo;
10) Observa-se se o critério de vigilância $\frac{|\boldsymbol{I}^a \wedge \boldsymbol{w}_J^a|}{|\boldsymbol{I}^a|} \geq \rho_a$ é satisfeito. Se sim, segue-se para o passo seguinte. Se não: *reset* em ART_a → $T_J^a = 0$ e volta-se a 9);
11) Com o nó vencedor obtém-se o vetor atividade em $F2^a$: $\boldsymbol{y}^a = (y_1^a, \ldots, y_{N_a}^a)$ → componente vale 1 apenas para categoria vencedora;

12) Observa-se se o critério de vigilância no inter-ART é satisfeito: $\frac{|\boldsymbol{y^b} \wedge \boldsymbol{w_J^{ab}}|}{|\boldsymbol{y^b}|} \geq \rho_{ab}$. Se sim, segue-se para o próximo passo. Se não, deve-se incrementar o parâmetro de vigilância por meio de $\rho_a = \frac{|\boldsymbol{I^a} \wedge \boldsymbol{w_J^a}|}{|\boldsymbol{I^a}|} + \varepsilon$ e retornar ao passo 9);
13) Atualiza-se os pesos de ART_a: $\boldsymbol{w}_J^{anovo} = \beta\left(\boldsymbol{I^a} \wedge \boldsymbol{w}_J^{avelho}\right) + (1-\beta)\boldsymbol{w}_J^{avelho}$;
14) Atualiza-se os pesos do módulo inter-ART: $w_{jk}^{ab} = \begin{cases} 1, se\ j = J\ e\ k = K \\ 0, se\ j \neq J\ ou\ k \neq K \end{cases}$;
15) Algoritmo se repete retornando ao passo 3).

O algoritmo supracitado guiará o treinamento da rede empregada no presente trabalho. As amostras de imagens devem ser, anteriormente, tratadas e padronizadas. Após tal etapa, técnicas de extração de atributos e codificação dos respectivos dados serão aplicadas a fim de possibilitar o referido treinamento. A eficiência da rede na assimilação correta em cada uma das abordagens dependerá, efetivamente, da forma com que os dados foram codificados. Isto é esmiuçado no capítulo a seguir.

4 METODOLOGIA DE IMPLEMENTAÇÃO

O presente capítulo apresentará todos os métodos selecionados e empregados para possibilitar a utilização da rede AMF na detecção do estado de saúde de globos oculares. Será necessário descrever brevemente o *dataset* utilizado, assim como as manipulações adotadas para fomentar a padronização adequada.

Vale ressaltar, novamente, que o processo de extração de atributos e codificação das imagens se trata da tarefa mais importante (podendo ser árdua, uma vez que o método ideal pode variar entre *datasets* distintos) para viabilizar confiabilidade no treinamento e assertividade eventual em amostras inéditas (RAN et al., 2021).

4.1 *DATASET* E MANIPULAÇÕES INICIAIS

A fim de possibilitar efetuar o proposto, um *dataset* contendo imagens oculares foi utilizado para treinamento e testes da rede. O referido é constituído por um conjunto de imagens contendo amostras diversas de doenças oculares (olhos não saudáveis) e amostras de olhos perfeitamente saudáveis, estruturado em fotos de fundos de retina (RFMiD) (PACHADE et al., 2021). Foi priorizada a escolha de um *dataset* de fotos bidimensionais de fundos oculares, uma vez que são mais comumente obtidas e não necessitam de técnicas excessivamente robustas para sua captura.

Antes de qualquer manipulação efetuada, o *dataset* compunha-se de 3200 imagens de fundos oculares capturadas usando três diferentes câmeras. Foram consideradas para listagem 46 condições distintas (45 referentes a diferentes patologias e 1 condição vinculada aos olhos saudáveis), que foram preenchidas com anuência de dois especialistas (PACHADE et al., 2021). A grande vantagem do *dataset* selecionado reside no fato de possuir diversas amostras de diferentes doenças oculares, incluindo algumas raras. Desta forma, um sistema classificatório baseado neste *dataset* pode auxiliar tanto na detecção de condições patológicas frequentes (como retinopatia diabética, glaucoma e degeneração macular), quanto no diagnóstico de doenças mais raras (como oclusão arterial da porção central da retina ou neuropatia óptica isquêmica anterior) (PACHADE et al., 2021). Desta forma, ressalta-se a importância de estudos na direção de detecção de patologias individuais ou do estado de saúde geral de globos oculares (saudáveis/não saudáveis).

As imagens foram obtidas usando câmeras centralizadas no disco óptico ou na mácula de pacientes visitantes de uma clínica para procedimentos cotidianos vinculados à saúde ocular. Os pacientes receberam uma gota de tropicamida para dilatação da pupila antes da captura de cada uma das imagens. A criação deste *dataset* se estendeu do período de 2009 a 2020, selecionando-se 3200 imagens dentre os milhares exames efetuados em tal hiato. Do total, 60% dos dados iniciais (1920 amostras) foram reservados para treinamento, ficando o restante para validação e testes (PACHADE et al., 2021). As informações das anomalias pertencentes a cada foto do *dataset* foram inseridas em uma planilha específica que foi utilizada constantemente para consulta. A Tabela 2 fornece uma síntese das informações mencionadas quanto ao banco utilizado.

Tabela 2 - Síntese e contextualização do *dataset* utilizado no trabalho.

Procedimento	Experimento biomédico de imagem vinculado à oftalmologia
Tema	Análise de imagem de retina para detecção do estado de saúde de globos oculares
Tipo de dado	Imagem/CSV
Obtenção dos dados	Três diferentes câmeras. Nome dos modelos:TOPCON 3D OCT-2000, Kowa VX-10 e TOPCON TRC-NW300
Pré-tratamento	Uma gota de tropicamida com concentração de 0,5% antes da captura
Localização	Hospital Sushrusha/Excelência em processamento de imagem/sinal (SGGSIE&T) - Nanded, Índia

Fonte: Adaptado de Pachade et al. (2021).

Uma vez descrito e especificado o *dataset* inicial, algumas manipulações foram efetuadas a fim de estabelecer uma melhor padronização para o processo de treinamento da rede. Inicialmente, as imagens estavam divididas em três pastas principais: uma vinculada a treinamento, outra à validação e a última para testes finais. A fim de aumentar a quantidade de amostras de treinamento da rede, todas as imagens pertencentes à pasta de validação foram integradas à pasta de treino, resultando-se em dois conjuntos principais de capturas: treinamento e testes.

Por meio desta primeira manipulação, a segunda etapa consistiu na divisão particularizada de cada uma das pastas resultantes. Inicialmente, diversas imagens de fundos de globos oculares permeavam as pastas sem qualquer tipo de distinção:

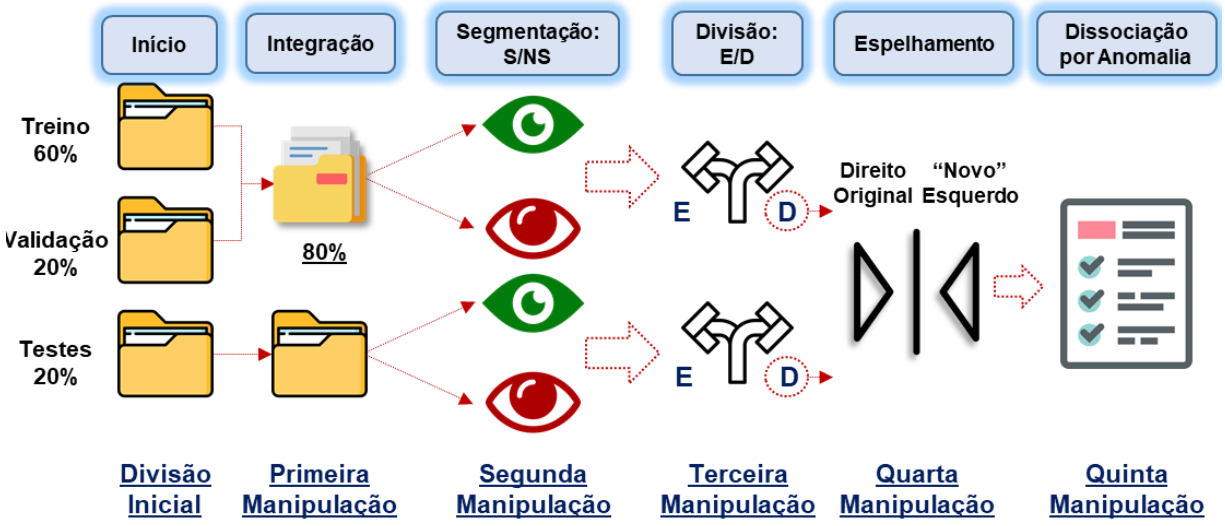
a única divisão até o momento efetuada consistia apenas na separação treinamento/testes. Para o presente trabalho, entretanto, é de interesse considerável que haja a separação de globos oculares saudáveis e não saudáveis em cada uma das pastas, uma vez que a rede precisa ser treinada para aprender a diferenciar tais condições. Aqui, olhos denominados “saudáveis” correspondem a estruturas isentas de qualquer tipo de anomalia; em contrapartida, olhos “não saudáveis” referem-se àqueles acometidos por uma ou mais das patologias existentes no *dataset* (PACHADE et al., 2021). Portanto, utilizando a planilha base do *dataset*, a segunda manipulação se deu em esmiuçar cada uma das pastas principais em duas outras subdivisões: “saudáveis” e “não saudáveis”.

Em cada um destes compartimentos, fotografias de globos oculares esquerdos e direitos apresentavam-se de forma misturada. Como forma de manter as repartições bem definidas e devidamente desagregadas, uma terceira manipulação segmentando as pastas em “olhos esquerdos” e “olhos direitos” foi empregada. A fim de permitir uma melhor padronização, é interessante que se selecione apenas uma das segmentações anteriores para apresentação de dados à rede durante o processo de treinamento. Optou-se pela escolha dos olhos esquerdos para o processo de treino da AMF, por possuírem uma quantidade de amostras levemente superior.

Ainda com o objetivo de aumentar o banco de exemplares para treinamento da rede, analisou-se a constituição das amostras pertencentes aos olhos direitos (que até então seriam “descartadas” do processo de treinamento, uma vez que se deliberou pela escolha dos esquerdos). Observou-se que os olhos direitos poderiam constituir um bom acréscimo ao banco caso fossem aproveitados. A alternativa utilizada nesta etapa foi a de “espelhar” as fotografias de globos oculares direitos, permitindo que se comportassem como amostras inéditas de olhos esquerdos. Este método mostrou-se interessante, visto que foi plausível a inserção de uma boa quantidade de dados que a rede compreenderá como exemplares individuais e inéditos durante o processo de treinamento. Ao invés de criar amostras artificiais por meio da extensão de dados (TUFAIL et al., 2021; SENGAR et al., 2023), por meio do espelhamento das imagens pôde-se ampliar de forma consistente o *dataset* para utilização final. Para tal artifício utilizou-se o software *XnConvert,* capaz de espelhar arquivos de imagens em lote.

A Figura 7 ilustra como as imagens foram inicialmente organizadas. Na fase de “dissociação por anomalias”, agrupou-se algumas amostras de olhos não saudáveis que pertenciam à mesma condição patológica em pastas individuais.

Figura 7 - Síntese das manipulações iniciais efetuadas no dataset.



Fonte: Elaboração do próprio autor.
Legenda: S/NS: Saudável/Não Saudável - E/D: Esquerdo/Direito.

4.2 APURAÇÃO E PRÉ-PROCESSAMENTO DAS IMAGENS

A partir das manipulações iniciais efetuadas no banco de dados, partiu-se para uma etapa mais técnica no que diz respeito à adequação das imagens. Neste subitem serão relatados todos os processos de seleção realizados, assim como todo e qualquer tratamento e conciliação que se julgou interessante produzir a fim de aprimorar o processo de treinamento da rede AMF. Subdivide-se em duas seções principais, tal como direcionado a seguir.

4.2.1 Seleção e Amostras Resultantes

Após todas as adequações iniciais no *dataset*, sintetizadas pela Figura 7, passou-se a analisar as fotos resultantes individualmente a fim de verificar a condição de cada uma. A ideia principal era, inicialmente, fomentar a padronização das fotografias pertencentes ao banco, eliminando aquelas que destoassem significativamente do padrão geral.

Como primeiro passo, buscou-se retirar as capturas imperfeitas que compunham o conjunto de exemplares, realizando uma espécie de exclusão manual de fotos com qualidade ruim. Muitas imagens, ou por iluminação destoante no momento da captura ou por descentralização repentina do paciente (gerando ruído na

fotografia) apresentaram distorções relevantes. A fim de suprimir dados que se enquadravam nesta descrição, uma exclusão manual foi efetuada nas fotos de cada uma das pastas resultantes. O processo, apesar de trabalhoso, visava fomentar a confiabilidade geral do *dataset* e permear um eventual processo de treinamento íntegro. A Figura 8 apresenta, em caráter de exemplo, duas amostras de imagens retiradas durante o processo de exclusão manual.

Figura 8 - Dois exemplares retirados durante o processo de exclusão manual.

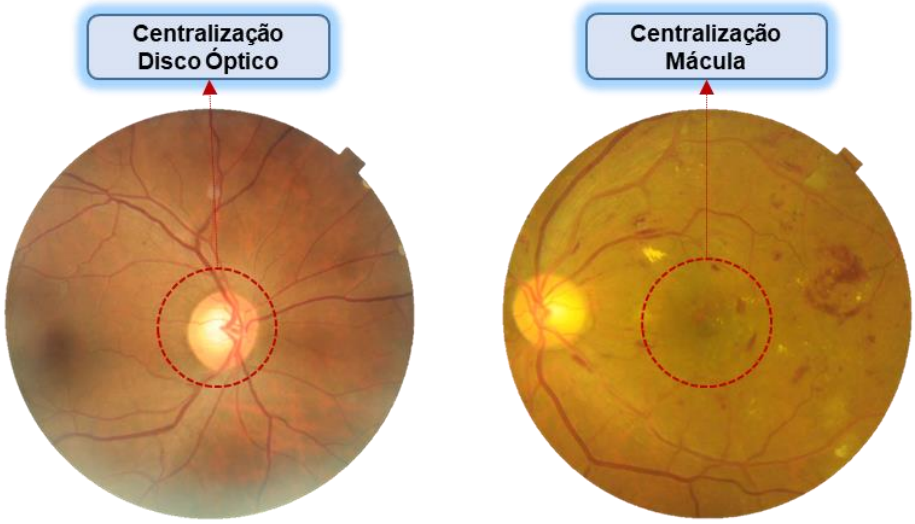
Fonte: Elaboração do próprio autor.

Na figura, observa-se um caso de possível descentralização do olho do paciente no momento da fotografia, provocando a entrada de uma iluminação destoante que deve ser encarada como um ruído do dado disponível. No outro caso, provavelmente devido à incorreta estabilização da câmera ou à presença de impurezas na lente (PACHADE et al., 2021), observou-se uma amostra resultante com acentuada opacidade. Ambas caracterizam capturas que foram retiradas do conjunto de dados final ao qual a rede seria submetida a treinamento e posterior teste, julgando-se desinteressante permanecerem.

Por fim, a última etapa vinculada à seleção das imagens foi com relação à centralização específica das capturas. As fotos foram tiradas em duas posições principais: concentradas ou na mácula ou no disco óptico dos pacientes. Verificando que a grande maioria dos exemplares correspondia à focalização na mácula, priorizou-se tal lote de capturas. Assim, visando a estandardização do conjunto de

dados, retirou-se do processo de treinamento e eventuais testes as imagens atreladas à centralização do disco óptico ocular.

Figura 9 - Exemplos dos dois tipos de centralização adotadas nas capturas.



Fonte: Elaboração do próprio autor.

Como verificado na Figura 9, as fotografias com referência à porção central da retina (mácula) permitem uma ampla verificação de boa parte da região ocular e, além disso, estavam vinculadas à maioria das anomalias presentes no *dataset* (PACHADE et al., 2021). Desta forma, além da seleção de apenas as capturas centralizadas na mácula, manteve-se somente as amostras que possuíam patologias alheias ao disco óptico, visto que a absoluta maioria das condições se apresentava no restante do globo ocular, fazendo com que a permanência de amostras com condições apenas vinculadas ao disco não fosse primordial. Além disso, a retirada dos poucos exemplares vinculados a anomalias do disco óptico permitirá efetuar o corte de tal parcela na etapa de procedimentos finais pré-codificação, que será relatada no tópico 4.2.2 seguinte.

Por conta de todas as exclusões empregadas e acima descritas, o banco de dados final utilizado para treinamento da rede AMF e posteriores testes foi reduzido em quantidade de exemplares. Apesar disso, a padronização favoreceu a confiabilidade das amostras presentes e eventual fidedignidade dos procedimentos, mantendo ainda um ótimo acervo.

A Tabela 3 sumariza a contagem remanescente para o teste objetivo do trabalho.

Tabela 3 - Síntese dos exemplares remanescentes.

Pasta \ Condição	Saudável	Não Saudável
Treinamento	226	845
Teste	15	216

Fonte: Elaboração do próprio autor.

4.2.2 Procedimentos Finais Pré-Codificação

Com a formação final do conjunto de dados para cada um dos testes designados, partiu-se para a última preparação das amostras antes da extração de atributos propriamente utilizadas.

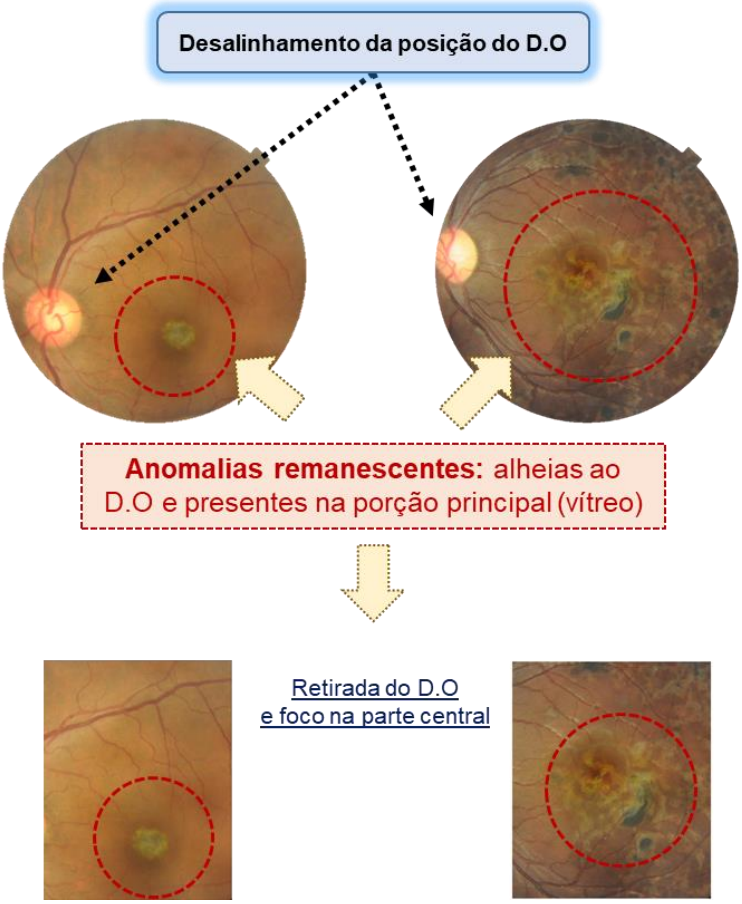
A primeira consideração referiu-se à análise quanto ao disco óptico dos globos oculares. Como relatado no tópico 4.2.1, no *dataset* final foram mantidas as fotografias centralizadas na mácula e com anomalias alheias ao disco óptico individualmente. Desta forma, priorizou-se os exemplares vinculados a doenças existentes na porção central das imagens de fundo ocular (região de existência do corpo vítreo) por conta da maior quantidade e melhores características distintivas. Dito isso, independente do processo de codificação a ser aplicado para representatividade dos dados, seria desnecessário manter a parcela da imagem pertencente ao disco óptico.

Ora, se todas as amostras finais que irão compor o banco de dados estão centralizadas na mácula e nenhuma anomalia vinculada ao disco óptico foi mantida para treinamento e teste da rede, não faz sentido manter esta porção nas imagens durante o processo de codificação. É como se a porção das fotos vinculada ao disco óptico fosse uma ramificação adicional. Além disso, observando as fotos resultantes minuciosamente, nota-se um desalinhamento da posição do disco óptico de foto para foto, dificilmente ocorrendo uma ideal centralização desta região entre as imagens. Mediante todas estas justificativas, julgou-se interessante retirar o disco óptico das fotos antes de efetuar a codificação. A Figura 10 ilustra tais ocorrências.

Ressalta-se que a retirada do disco óptico de todas as fotos foi feita de forma precisa e automatizada utilizando a função de recorte (“*imcrop*”) do Matlab. Nos processos de treinamento e testes da rede AMF, inicialmente três abordagens quanto aos recortes foram utilizadas: a retratada na Figura 10, com focalização no centro da imagem; uma menos acentuada, mantendo mais da região do corpo vítreo e retirando

apenas as extremidades; e uma geral, preservando o máximo possível das fotografias e excetuando-se apenas o disco óptico. Tais abordagens serão mostradas no item 5.

Figura 10 - Indicação da retirada do disco óptico.



Fonte: Elaboração do próprio autor.
Legenda: D.O – Disco Óptico.

O último procedimento estabelecido antes do início da implementação dos processos de extração de atributos e codificação referiu-se à conversão das imagens para escala de cinza/preto e branco. Para se compreender a necessidade de aplicação de tal medida, é preciso primeiramente entender como o software Matlab faz a leitura de imagens de formato “png”, característico das amostras do banco.

Qualquer imagem é composta por unidades individuais conhecidas como pixels. A quantidade de pixels em uma imagem está diretamente atrelada à sua resolução: quanto maior o número de tais unidades, maior a qualidade da foto e vice-versa. No Matlab, cada pixel representativo de uma imagem é atrelado geralmente a um número de 0 a 255. Tal número, cuja unidade corresponde a “inteiro sem sinal de 8 bits” (uint8), representa basicamente o grau de tonalidade presente naquela

determinada região da imagem (MATHWORKS, 2023b). Quando se efetua a leitura de uma imagem colorida em png, três matrizes de dimensões equivalentes à resolução da imagem são lidas: uma representando a cor vermelho, outra a cor verde e, por último, uma vinculada ao azul. Em cada posição de pixel de cada uma dessas matrizes aparece o valor da referida tonalidade das cores individuais. A união e composição das três matrizes garante a tonalidade que dá a cor final para a imagem lida. Esta estruturação é conhecida como RGB (HIRSCH, 2004), sistema de cores aditivas no qual o vermelho, verde e azul são combinados de modo a reproduzir qualquer cor do espectro cromático.

A existência das três matrizes que compõem cada imagem representa uma grande quantidade de elementos para se trabalhar vinculada a cada amostra. Como verificado no *dataset*, algumas imagens possuem diferença de iluminação em relação às outras, provavelmente devido aos distintos períodos do dia (ou características do ambiente e câmeras) em que as fotos foram tiradas. Mesmo que o *dataset* tenha passado pelos filtros iniciais relatados anteriormente, estas diferenças de iluminação podem ser prejudiciais ao treinamento. Além disso, as saliências que distinguem as imagens entre si são perceptíveis sem a necessidade da coloração. Por fim, vale mencionar que os processos de detecção de pontos de interesse e consecutivas extração de características são funcionais apenas com as imagens em escala de cinza.

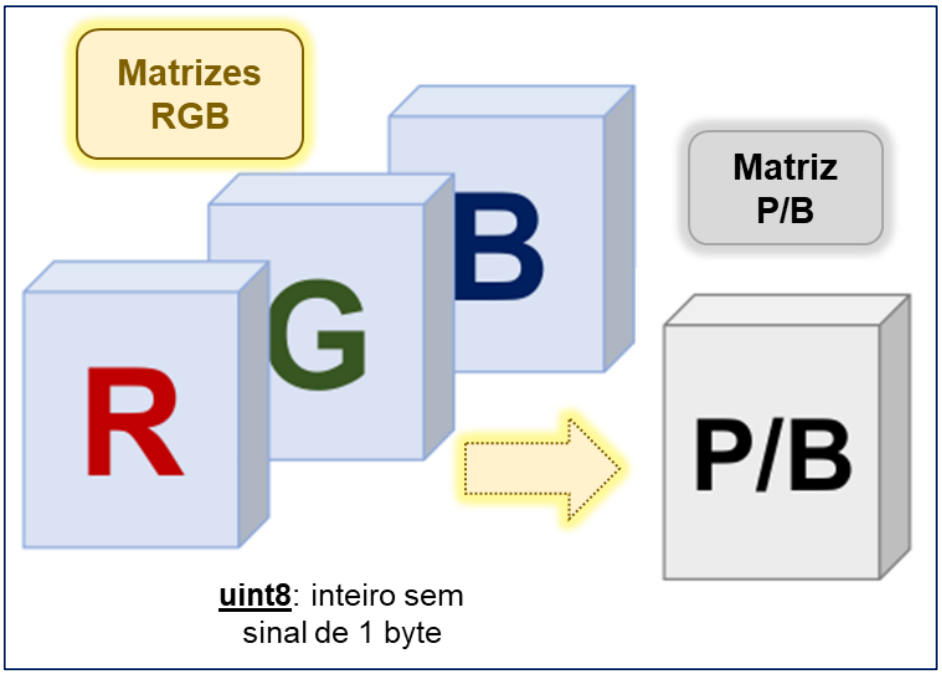
Dado todo este panorama, fez-se essencial converter as imagens para preto e branco antes do início do processo de codificação, reduzindo-as para apenas uma matriz (ao invés das três da estrutura RGB). A Figura 11 esquematiza esta medida. A partir de todas as adequações efetuadas, prosseguiu-se com os processos de codificação a fim de promover o início do treinamento da rede.

4.3 EXTRAÇÃO DE ATRIBUTOS E CODIFICAÇÃO

A seção atual retrata os métodos de extração de atributos empregados e testados no presente trabalho. A maneira de efetuar a representação das imagens antes da etapa de treinamento da rede AMF corresponde ao principal trunfo para garantir fidedignidade do sistema na detecção do estado de saúde de olhos. Os métodos empregados foram: matriz em vetor linha (proposto pelo autor), codificação por histograma de imagens (sugerido para esta aplicação médica) e frequência de

palavras visuais existentes (BoVW). Desta forma, será possível comparar a eficiência de cada método de extração de atributos e codificação para o presente tema.

Figura 11 - Esquematização da conversão para escala de preto e branco.



Fonte: Elaboração do próprio autor.

4.3.1 Matriz da Imagem em Vetor Linha

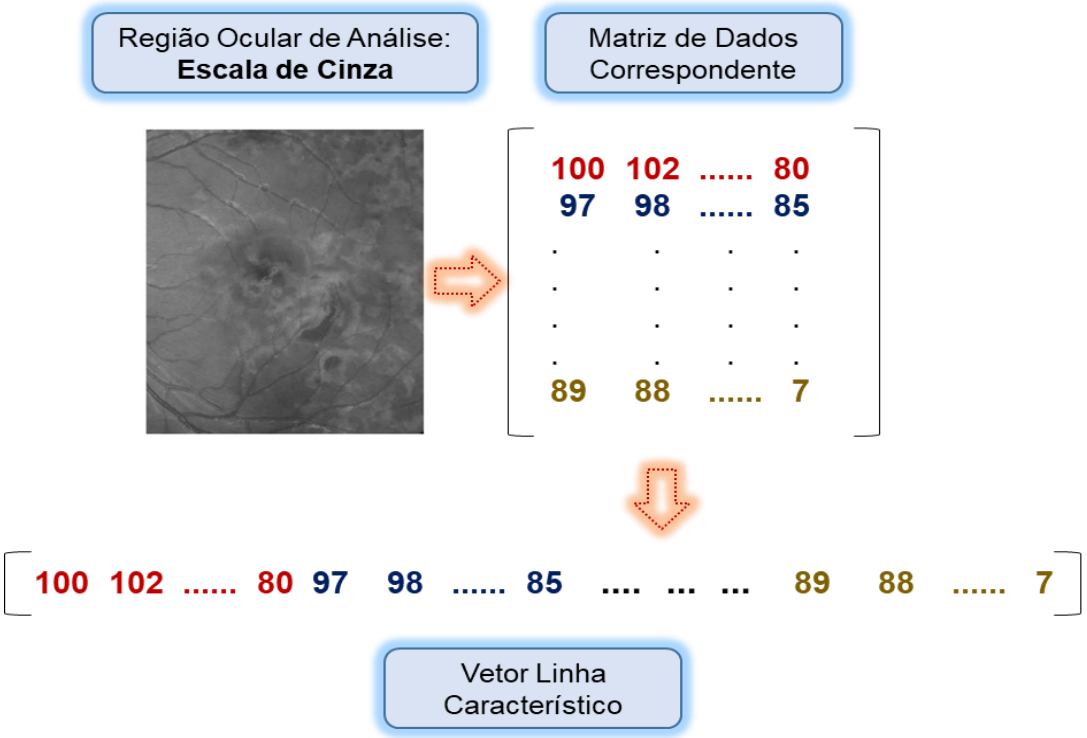
Como dito no tópico 4.2.2, a leitura de uma imagem em preto e branco é efetuada mediante estabelecimento de valores numéricos sem sinal (de 0 a 255) para os pixels correspondentes, cada qual representando o grau de tonalidade naquela determinada região. A primeira ideia foi utilizar exatamente a matriz equivalente de cada imagem como estrutura de codificação, sendo sugerida pelo próprio autor.

Cada imagem em escala de cinza possuirá uma matriz vinculada com os valores numéricos em cada pixel. Esta matriz é inerente a cada foto, representando a respectiva tonalidade/grau de cor em cada porção mínima. Isso significa que cada matriz conterá as informações específicas de cada fotografia individualmente, traduzida pelo valor dos elementos que farão sua composição. De foto para foto esses elementos podem variar até drasticamente, dependendo da região da anomalia ou do grau de agressividade acometido ao globo ocular. Sabendo disso, a ideia seria utilizar

estas informações como forma de codificação, haja visto a individualidade inerente a cada captura.

No entanto, como cada foto é representada por uma matriz, é de interesse converter esta informação bidimensional em unidimensional a fim de inserir na rede AMF e vinculá-la a uma respectiva saída. Desta forma, a sugestão inicial é empregar as informações da matriz referente a cada captura ajustada e convertê-las para um vetor linha característico. A Figura 12 ilustra este procedimento. Para as três abordagens de recorte realizadas nas fotos, o tamanho dos vetores variou entre 7700, 8400 e 8772 elementos, dado o redimensionamento efetuado nas fotografias e o grau de severidade do recorte. No item 5 tais informações serão descritas.

Figura 12 - Estratégia adotada na primeira codificação empregada.



Fonte: Elaboração do próprio autor.

4.3.2 Histograma de Imagens

A segunda forma de codificação foi estruturada nos histogramas representativos de cada imagem. Este basicamente conta quantos valores de tonalidade aparecem em uma determinada imagem que foi lida. Como dito, a leitura de uma fotografia com profundidade de 8 bits gera valores de pixels entre 0 e 255. Um

histograma da imagem em escala de cinza lido refere-se a quantas vezes cada um desses valores de tonalidade apareceu naquela determinada captura, representando uma boa ferramenta para avaliar distribuição de tons (GONZALEZ; WOODS, 2009).

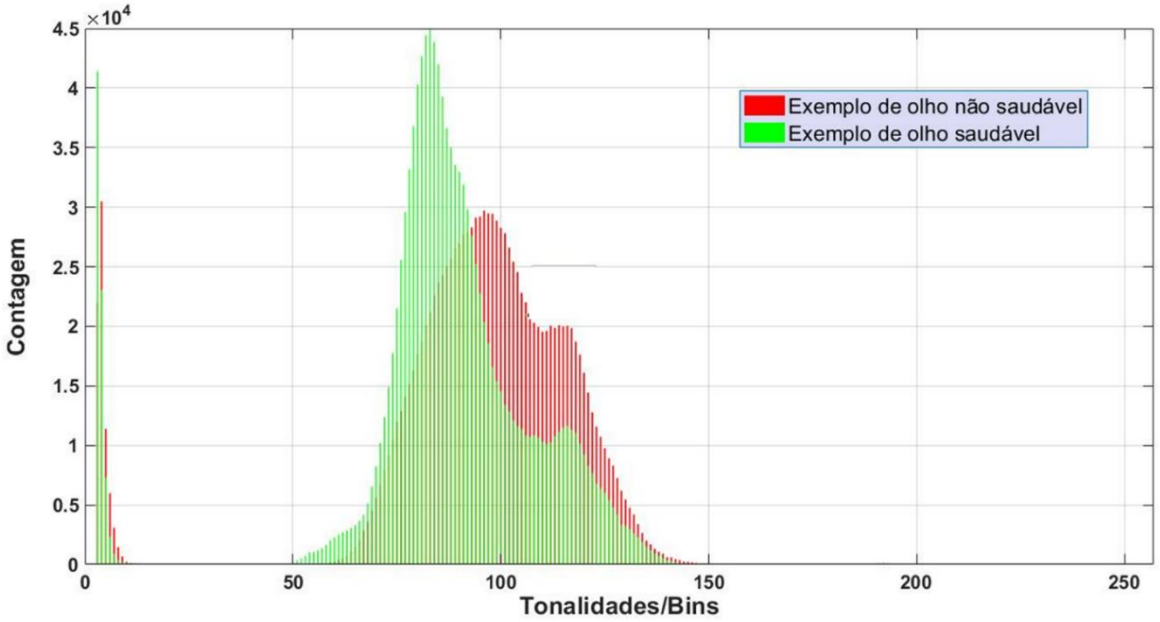
Quanto mais próximo de 255, mais escuro representa o tom; quanto mais próximo de 0, mais claro. Geralmente, a região na qual a maioria dos valores tonais se encontra pode variar drasticamente de uma imagem para outra. Isso representa uma informação interessante dado o *dataset* disponível: dependendo da condição ocular, a faixa de tons principais pode ser diferente (depende do tamanho da anomalia do referido globo e da patologia em si).

Além disso, supondo que a pasta pertencente aos olhos saudáveis possui amostras relativamente parecidas entre si, é natural supor que uma diferenciação no histograma possa existir entre estes e os não saudáveis. No entanto, é importante ressaltar que o histograma efetuará apenas a contagem do aparecimento de cada tonalidade dentre as existentes em cada fotografia: isso significa que, ainda que um determinado globo ocular possua um patógeno agressivo que acometa uma grande região, mas por acaso sua distribuição de tonalidades seja parecida com a de um olho saudável, certamente a rede pode encontrar dificuldade em diferenciar as duas amostras.

Ressalta-se que, também, a condição de iluminação das imagens pode afetar a representação do histograma. Qualquer entrada de luminosidade em alguma fotografia, ainda que todas estejam em escala de cinza para confecção do histograma, pode afetar sua composição, uma vez que tal alteração será traduzida numa contagem diferente de tonalidade de pixel. A Figura 13 retrata um exemplo comparativo de diferenciação de histograma em olho saudável e não saudável, baseado em duas amostras aleatórias do *dataset*.

Como observado, no caso dessa amostra de olho saudável específica observa-se um pico de aparecimento de tonalidades vinculadas à faixa de 75-100. No caso da amostra não saudável, verifica-se uma distribuição mais uniforme em comparação ao exemplar saudável, não possuindo nenhum pico muito acentuado. Desta forma, fica simples codificar as imagens fundamentadas nesta segunda proposta: basta montar um vetor característico, para cada foto, representado pela contagem de tonalidades (histograma) existente. Assim, nesta abordagem, cada vetor representativo das fotos possuirá tamanho fixo de 256, traduzindo o aparecimento de cada tonalidade na respetiva captura.

Figura 13 - Comparativo de histogramas de dois globos oculares do *dataset*.



Fonte: Elaboração do próprio autor.

4.3.3 BoVW – *Bag of Visual Words*

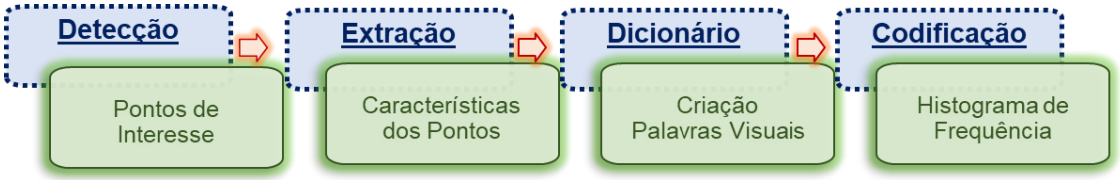
Neste item, abordar-se-á o BoVW para extração de atributos e codificação das imagens dos olhos. Se comparado aos outros, este representa a abordagem com mais etapas para implementação. Tal artifício corresponde a um método relativamente usual para extração de características de imagens em diferentes áreas de aplicação.

1) Definição

BoVW representa, ao pé da letra, o conceito de “Sacola de Palavras Visuais”. A ideia foi desenvolvida como forma de obter uma linguagem mais compreensiva e interpretativa dos dados aos quais se está trabalhando. Se suponha a criação de um certo número de categorias baseadas em pontos de interesse que sejam capazes de representar as características de um conjunto de dados. Para cada uma dessas categorias, um “nome” (ou “palavra”) específico será atribuído. A partir da criação de todas estas palavras (o chamado “dicionário), é possível representar uma determinada amostra do *dataset* como a **frequência de vezes de aparecimento de cada palavra**. Ou seja: a codificação por BoVW descreve um exemplar como um conjunto de palavras que serão criadas baseadas nas características do *dataset* (OKAWA, 2018).

Para utilização deste método é necessário o estabelecimento de algumas etapas. No caso das imagens, primeiramente deve-se implementar alguma ferramenta que detecte pontos de interesse das fotografias e, posteriormente, extraia as características correspondentes. Com todas as características disponíveis, cria-se o dicionário de palavras visuais. Estas atuarão como centroides: cada característica poderá, assim, ser atribuída ao centroide mais próximo. Por fim, codifica-se as imagens individualmente quantificando-as quanto à frequência de cada palavra visual (OKAWA, 2018; VIMINA; JACOB, 2019).

Figura 14 - Sumarização do método BoVW.



Fonte: Elaboração do próprio autor.

2) Detecção de pontos de interesse

Para utilizar o BoVW será preciso empregar alguma técnica para detecção de pontos de interesse nas fotografias. Será por meio destes pontos que as características serão extraídas e poderá ser formado o dicionário de palavras visuais.

Recentemente, o uso das características KAZE tem sido proposto para detectar a orientação principal dos pontos chaves das amostras e para obter um descritor invariante em escala e rotação em espaços de escala não lineares (OKAWA, 2018). Além disso, fundamentar o método BoVW nestes chamados descritores invariantes tem mostrado ótimo desempenho em classificação de objetos e reconhecimentos (VIMINA; JACOB, 2019).

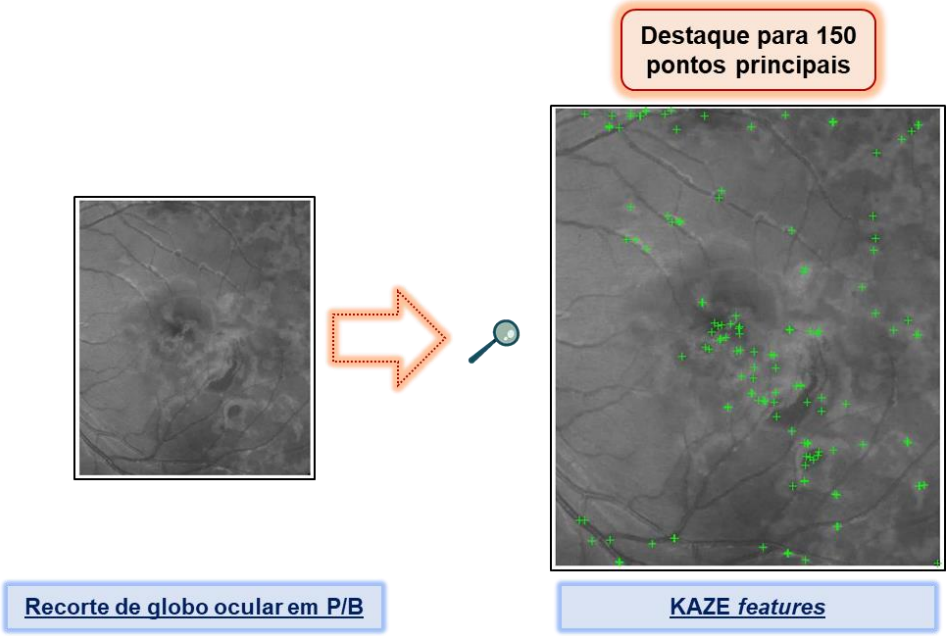
O KAZE basicamente representa um algoritmo de detecção e descrição de características, ancorados em análises bidimensionais e em espaços de escala não lineares. Sua principal diferença quanto aos outros descritores comumente empregados refere-se à abordagem: pelo uso de filtragem difusa não linear é possível promover a focalização dos dados das imagens de forma localmente adaptativa, preservando melhor os limites da mesma. Esta propriedade favorece a diferenciação mais sutil de ruídos e alguns detalhes (como cantos e realces), obtendo acurácia

superior de localização e acentuada distinção de regiões (ALCANTARILLA; BARTOLI; DAVISON, 2012).

No caso do presente trabalho, pelos atributos mencionados anteriormente, julgou-se interessante a utilização da ferramenta KAZE como detector de pontos de interesse dos recortes das capturas em escala de cinza. Além disso, mediante testes experimentais, percebeu-se um desempenho superior desta abordagem em relação às SURF (BAY et al., 2008) e SIFT (LOWE, 2004) para este conjunto de dados. No caso de aplicação da KAZE, geralmente uma maior quantidade de pontos de interesse era localizada em cada fotografia, permitindo a seleção dos pontos mais fortes ou com realce mais acentuado para compor a eventual matriz de características.

Em algumas fotografias nas quais testou-se o método SURF, por exemplo, menos de 10 saliências eram detectadas, desfavorecendo sua utilização para criação do dicionário de palavras. Esta ocorrência deve estar presente, provavelmente, devido às características matemáticas e de abordagem teórica que distinguem os métodos. Alcantarilla, Bartoli e Davison (2012) descrevem como os métodos que empregam aproximação do espaço de escala Gaussiano (tal qual o SURF) de uma imagem podem não respeitar os limites naturais dos objetos e suavizar ruídos e detalhes, prejudicando o desempenho. A Figura 15 ilustra os pontos de interesse KAZE obtidos em uma amostra arbitrária do *dataset*, ampliados para visualização.

Figura 15 - Pontos de interesse de amostra do *dataset* via características KAZE.



Fonte: Elaboração do próprio autor.

Como observado, o método pode destacar regiões de interesse do recorte de uma captura arbitrária, incluindo saliências, alterações na textura e distinções variadas nas imagens. Para o trabalho, utilizou-se um máximo de 150 pontos para cada exemplar, visto que o método poderia detectar muitos pontos em algumas fotografias.

3) Extração de Características

Após a detecção dos pontos mencionados acima, é necessário utilizar alguma ferramenta que permita efetuar a descrição de tais regiões. Uma vez que se possui a localização geométrica dos pontos e suas características, é preciso que as informações relacionadas a cada um sejam descritas de alguma maneira. Deve ser necessário, portanto, extrair características e sintetizar descritores das regiões vinculadas a cada ponto de interesse obtido.

Para tal, foi empregada a função “extractFeatures”, própria do Matlab. Com ela é possível obter o retorno destes descritores intrínsecos a cada imagem. O método de extração empregado nesta função foi o KAZE (baseado numa combinação de informações de gradientes e padrões locais), uma vez que também se tratou da abordagem de detecção. Esta forma descritiva é robusta a mudanças geométricas e rotações, sendo eficiente computacionalmente e bastante versátil (ALCANTARILLA; BARTOLI; DAVISON, 2012). Vale frisar que a maneira de detectar os pontos e descrevê-los representa duas etapas distintas e separadas: é possível que se utilize o detector KAZE de início e, para extração de características, recorra-se ao SURF ou HOG, por exemplo. Cada forma de extração irá resultar em descritores diferenciados.

A aplicação da função “extractFeatures” ancorada nos descritores KAZE culminou na obtenção final de um vetor descritivo para cada ponto de interesse colhido das imagens. Cada vetor ficou com 64 colunas e cada coluna referenciava uma determinada informação a respeito de algum ponto-chave e sua respectiva vizinhança. O descritor KAZE combina informações para construir uma representação descritiva da região em torno desses pontos (ALCANTARILLA; BARTOLI; DAVISON, 2012).

4) Criação do Dicionário e Vetor Característico das Imagens

Com todas as características extraídas após o processo de detecção dos pontos de interesse de cada fotografia, foi necessária a elaboração do dicionário de

palavras para permitir a codificação pelo método BoVW. Com a matriz de características elaborada (formada pela descrição de cada ponto mediante a extração de informações), partiu-se para a criação do dicionário.

A fim de permitir sua implementação, é preciso que se utilize algum método de *clustering* que possibilite a criação de vínculos entre cada ponto e a respectiva palavra ao qual será atribuído. Na formulação *k-means* (ARTHUR; VASSILVITSKII, 2007), por intervenção da estipulação de K centros (ou palavras), almeja-se direcionar os *n* pontos existentes de tal forma a minimizar a soma das distâncias Euclidianas ao quadrado entre cada ponto e o centro mais próximo. Desta forma, basta especificar a quantidade de palavras desejadas K que é possível empregar o algoritmo *k-means* para efetuar o agrupamento das características nos respectivos *clusters* mais próximos.

No caso do presente trabalho, obteve-se uma contabilização de 132221 pontos de interesse extraídos considerando todas as imagens. Este valor representa a soma de todos os pontos detectados levando em conta o conjunto completo de treinamento. Na grande maioria das imagens, o método KAZE detectou uma grande quantidade de pontos, sendo configurado para escolha os 150 principais; no entanto, em algumas figuras específicas, o método detectou menos de 150 pontos, sendo todos eles considerados nestas ocasiões. A soma total dos pontos, portanto, foi de 132221 e estes serão destinados aos centroides mais próximos durante a aplicação do método *k-means* mediante a escolha da quantidade de palavras visuais (e por meio de todo o processo iterativo que visa reduzir progressivamente o erro).

O método de agrupamento *k-means* se inicia distribuindo K centros/palavras de forma arbitrária no espaço dos dados existentes. Cada ponto é designado ao centro mais próximo, e cada centro é recalculado como o centro de massa de todos os pontos vinculados a ele. Este processo se repete até a estabilização, podendo-se ressaltar que o erro total é monotonicamente decrescente: os centros vão sendo recalculados e a soma das distâncias Euclidianas ao quadrado entre os pontos e os centros a que pertencem reduzem progressivamente. No estudo de Arthur e Vassilvitskii (2007) evidencia-se a importância de inferir os centros iniciais mediante probabilidades específicas, ao invés de apenas gerá-los randomicamente. Desta forma, o algoritmo funciona melhor e não corre o risco de cair em centros ruins. Esta forma melhorada de concebê-los inicialmente é chamada *k-means*++ (ARTHUR; VASSILVITSKII, 2007).

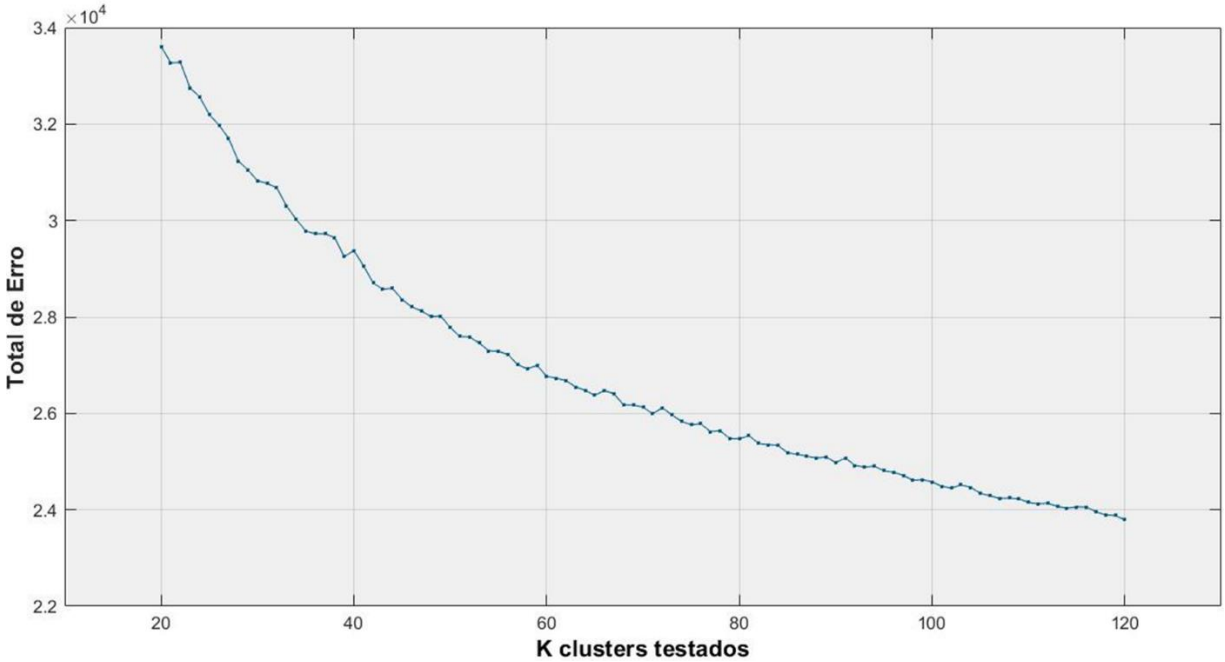
No Matlab a função *k-means* emprega, por padrão, a escolha inicial dos centros utilizando a métrica *k-means++.* Deste modo, pôde-se aplicar tal função na determinação dos agrupamentos. O algoritmo citado pode ser sintetizado nos passos abaixo (ARTHUR; VASSILVITSKII, 2007):

a) Determinação dos centros iniciais via abordagem *k-means++;*
b) Para todos os centro c_i elaborados no item anterior, considerar o *cluster* C_i correspondente ao conjunto de pontos/dados que estão mais próximos de c_i do que dos outros centros;
c) Reconfigure cada c_i como o centro de massa de todos os pontos em C_i;
d) Repetir até que não haja mudança de posição de cada c_i.

Para aplicação do algoritmo, inicialmente foi preciso determinar qual a quantidade de palavras visuais seria escolhida. Não existe uma regra rígida que direcione para a escolha ideal de *clusters* no algoritmo. Portanto, optou-se por escolher o *range* de 20-120 para testes de aplicabilidade do método. Como cada fotografia possuía no máximo 150 pontos de interesse detectados, julgou-se interessante configurar um limite de palavras visuais para utilização. Assim, para cada um dos casos, analisou-se qual a soma total do erro, ou seja, qual a soma total das distâncias Euclidianas quadradas dos pontos aos centros respectivos criados pelo algoritmo. Desta forma, inicialmente foi preciso rodar o algoritmo 101 vezes, uma para cada quantidade K de palavras, a fim de estabelecer o melhor valor de K dentre os testados. Para cada um dos 101 testes, toda a matriz de características elaborada foi utilizada (que representa todas as imagens de treinamento do *dataset*). A Figura 16 ilustra a relação da soma total do erro para cada K.

Como verificado pela figura referida, o erro diminui conforme aumento da quantidade de palavras visuais estipulada. No entanto, não era desejado utilizar um número muito grande de palavras, dado que cada captura possuí no máximo 150 características detectadas. Além disso, a seleção de uma quantidade não muito grande de palavras permite-nos analisar melhor as discrepâncias comparativas entre a frequência de aparecimento de palavras em cada uma das fotos, dado que poucas delas devem deixar de aparecer na contagem dos histogramas finais (para cada captura individual).

Figura 16 - Comparação da quantidade de clusters e erro total no *k-means.*



Fonte: Elaboração do próprio autor.

Por fim, com o dicionário elaborado, cada fotografia foi codificada mediante a contagem de aparecimento de cada palavra. Assim, em um dicionário de K palavras, cada imagem possuirá vetor representativo com K colunas. Esta modelagem será usada para treinar a rede AMF. Inicialmente buscou-se efetuar os testes com 100 palavras visuais. Análises adicionais para 300 palavras foram implementadas apenas para caráter de comparação quando da utilização de um dicionário três vezes maior.

5 RESULTADOS EXPERIMENTAIS E DISCUSSÃO

O desempenho da AMF para cada uma das codificações será testado neste capítulo. O *dataset* foi manipulado tal como relatado no capítulo anterior. A rede foi implementada no software Matlab R2019b, seguindo as diretrizes estabelecidas por Carpenter et al. (1992). Para a avaliação da performance da rede utilizando as abordagens sugeridas de extração de atributos e codificação, utilizou-se a métrica de taxa de classificação correta (CCR) (AL-OBAYDY; SUANDI, 2020). Tal parâmetro foi aplicado na verificação do resultado geral e em cada classe individualmente.

Em cada um dos métodos de codificação descritos anteriormente, testou-se a rede para diferentes valores de redimensionamento das imagens. Quanto menos agressivo o redimensionamento, maior a quantidade de informações da fotografia; no entanto, maior o número de elementos representativos e densidade do vetor característico. Buscou-se verificar, na base de sucessivos testes, qual o redimensionamento ideal para cada caso, preservando um número interessante de dados sem prejudicar o treinamento com o grande aumento de informações nos vetores de entrada.

O processo de treinamento da rede também foi efetuado mediante vários testes do parâmetro *baseline* $\overline{\rho_a}$, partindo-se de 0 e indo até 0,9 com passo de 0,1. Os melhores resultados obtidos serão apresentados no decorrer do capítulo. Além disso, para cada cômputo superior realizou-se o treinamento de uma época e o de *n* épocas, sendo esta última técnica empregada até que a precisão da rede com o conjunto de dados do treinamento fosse de 100%, tal como corroborado em um dos testes de Carpenter et al. (1992). A Tabela 4 sintetiza algumas informações e parâmetros gerais do treinamento.

5.1 ESTADO DE SAÚDE DE GLOBO OCULAR – MATRIZ EM VETOR

Como relatado no tópico 4.2.2, três abordagens distintas quanto a recortes nas imagens foram testadas no treinamento da rede. Cada uma delas fornece uma quantidade distinta de informações para compor o vetor de características intrínseco a cada imagem. O objetivo é verificar qual a melhor abordagem para a rede, analisando em qual delas ocorre superioridade da eficiência. A Figura 17 mostra os tipos de recorte efetuados para um mesmo globo ocular.

Tabela 4 - Informações e parâmetros gerais empregados.

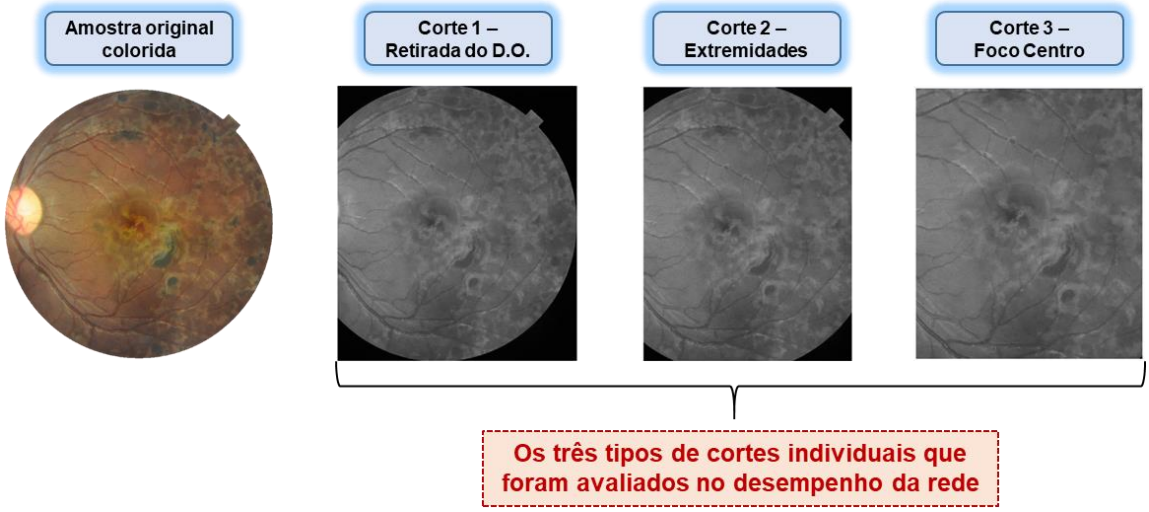
Designação	Equivalência
Dados de entrada	*Dataset* RFMiD (PACHADE et. al, 2021)
Tamanho dos dados	Originalmente 2144x1424x3
Adequações	Recortes distintos excluindo o disco óptico
Redimensionamento	Testes com redução das imagens na faixa de 50% - 93%
Codificações	Matriz em vetor, histograma de imagens, BoVW
Classificador	Rede neural artificial ARTMAP *Fuzzy*
Baseline $\overline{\rho_a}$	Testes na faixa [0,1 0,2 0,3 ... 0,9]
ρ_b	0,95
ρ_{ab}	0,90
β_a e β_b	1
ε	0,0001
α	0,001
Número de épocas	1 época / *n* épocas (até 100% conjunto de treinamento)
Análise principal da performance	CCR
Software de implementação	MatLab R2019b

Fonte: Elaboração do próprio autor.

A Tabela 5 sumariza os melhores resultados obtidos para cada um dos recortes neste método sugerido, apresentando o redimensionamento empregado em cada um dos casos. Como a dimensão das imagens originais é naturalmente grande, é necessário aplicar uma boa redução do tamanho a fim de obter um vetor não muito extenso, mas que contenha informações suficientes que possibilitem diferenciá-los. A dimensão resultante (elementos) em cada recorte é explicitada no rodapé da tabela.

Para o recorte 1 (ver Figura 17), o redimensionamento empregado permitiu que se obtivesse uma redução nas imagens em 93%. Esta redução possibilitou manter o vetor de características (após a codificação de complemento) com menos de 16000 elementos para cada fotografia. Os melhores resultados logrados foram para os parâmetros *baseline* de 0 e 0,9. Para o primeiro, o treinamento de 4 épocas permitiu que a rede obtivesse 100% de precisão com os dados de treinamento, ampliando levemente a taxa de classificação geral (em relação ao de uma época). No caso do *baseline* em 0,9, o treinamento de 1 época já possibilitou 100% de precisão com as amostras de treinamento. Percebe-se que, neste último, a classificação dos olhos saudáveis individualmente teve melhora significativa, muito provavelmente devido à

Figura 17 - Recortes empregados para testes no método matriz em vetor.



Fonte: Elaboração do próprio autor.

criação de mais categorias discriminativas na rede AMF durante o processo de treinamento.

No recorte 2 aplicou-se uma taxa de redimensionamento um pouco menor. Isso foi possível devido ao fato deste recorte ser mais acentuado em relação ao primeiro, obtendo-se aproximadamente o mesmo tamanho resultante do vetor de características por fotografia com a redução efetuada. Os melhores resultados encontrados foram para os *baselines* de 0,7 e 0,9, respectivamente.

Tabela 5 - Resultados principais do método de codificação Matriz em Vetor.

Método	Taxa Redim.	*Baseline*	Épocas	Cat C	CCR S	CCR NS	CCR G
Recorte 1	0,07 – 93% redução	0	1	8	0,00%	84,26%	78,79%
			4	8	0,00%	85,19%	79,65%
		0,9	1	226	**40,00%**	80,56%	77,92%
Recorte 2	0,08 – 92% redução	0,7	1	29	0,00%	83,33%	77,92%
			2	31	0,00%	83,33%	77,92%
		0,9	1	255	20,00%	77,31%	73,59%
Recorte 3	0,1 – 90% redução	0	1	10	26,67%	**98,15%**	**93,51%**
			3	10	26,67%	**98,15%**	**93,51%**
		0,5	1	18	33,33%	93,06%	89,18%
		0,9	1	251	**40,00%**	78,70%	76,19%

Fonte: Elaboração do próprio autor.

Legenda: Cat C = Categorias Criadas ART_a, S = Saudáveis, NS = Não Saudáveis, G = Geral.
Dimensões resultantes (elementos): Recorte 1 = 7700, Recorte 2 = 8400, Recorte 3 = 8772.

Novamente, para o caso de $\overline{\rho_a}$ valendo 0,9, a rede criou mais classes e conseguiu discernir melhor olhos saudáveis. No caso de classificação geral (incluindo saudáveis e não saudáveis conjuntamente), nos dois primeiros métodos de recorte a rede atingiu eficiência de pouco menos de 80%.

Para o terceiro recorte, três casos mereceram destaque: $\overline{\rho_a}$ de 0, 0,5 e 0,9. No primeiro, a rede atingiu a maior eficiência geral em relação a todos as outras abordagens citadas anteriormente, além da maior eficiência dos não saudáveis individualmente. O fato de $\overline{\rho_a}$ valer 0 permite que todas as categorias vencedoras da função de escolha passem pelo critério de vigilância do módulo ART_a , sendo desconsideradas apenas mediante atuação do mecanismo de *match tracking*. Este artifício faz com que o fluxo de categorias selecionadas no módulo de entrada seja muito maior, ficando o módulo inter-ART isoladamente responsável por “barrar” uma vencedora, sendo retratado em um dos testes comprobatórios de Carpenter et. al (1992). Nos casos subsequentes ($\overline{\rho_a}$ de 0,5 e 0,9), nota-se como a criação de mais categorias no módulo ART_a aumenta a capacidade de detecção de olhos saudáveis, ainda que promova uma pequena redução da porcentagem de acerto geral.

No caso da última abordagem de recorte, fica nítido a superioridade em relação às outras duas, obtendo a melhor eficiência em cada uma das classes analisadas (destaque em verde na Tabela 5). A melhor classificação dá-se, provavelmente, pela manutenção apenas da porção mais importante do globo ocular, excluindo-se totalmente as extremidades, o disco óptico (cujas anomalias vinculadas não são analisadas neste trabalho) e a porção escura remanescente dos cantos das fotos. Em outras palavras, trabalhar apenas com a porção mais central de cada captura permite que o sistema identifique e represente melhor cada caso por este processo de codificação. Além disso, manter apenas esta região favorece o redimensionamento menos acentuado das imagens para o mesmo tamanho aproximado do vetor de características, possibilitando sustento de uma maior quantidade de informações.

Como o terceiro recorte apresentou melhor desempenho, um novo tipo de treinamento foi efetuado apenas a efeito de teste para esta abordagem. Neste teste adicional, treinou-se a rede algumas vezes sequencialmente mantendo a mesma proporção de amostras de olhos saudáveis e não saudáveis a cada iteração de treino. O objetivo era verificar se a apresentação igualitária da quantidade de amostras no processo de treinamento ocasionava uma melhora da taxa de classificação correta. Para o melhor caso geral obtido no recorte 3 ($\overline{\rho_a}$ de 0), efetuou-se o treinamento

descrito por 4 vezes seguidas, apresentando 226 amostras de cada classe a cada iteração. Obteve-se eficiência geral de 91,34% nos testes com exemplares inéditos, incluindo nenhum acerto dos saudáveis. Como tal resultado mostrou-se muito desbalanceado, tal técnica não foi mais empregada no decorrer do trabalho.

Devido à evidente superioridade do recorte 3 em relação aos outros, os próximos métodos de extração de atributos e codificação que serão apresentados receberam enfoque principal nesta abordagem. O método de transferência das informações da matriz de cada fotografia em vetores característicos, apesar de simples, traduziu boa interpretação e eficiência geral da rede em até 93,51%. Esta abordagem, sugerida pelo autor, pode ser potencializada em trabalhos futuros mediante aplicação de técnicas de redução de dimensionalidade do vetor de características (como análise de componentes principais - PCA), posto que seria possível manter ainda mais dados informativos por imagem analisada.

5.2 ESTADO DE SAÚDE DE GLOBO OCULAR – HISTOGRAMA DE IMAGENS

Para a análise da eficiência da rede empregando a técnica estruturada no histograma das imagens, aderiu-se a duas abordagens distintas. Na primeira, efetuou-se a normalização convencional estabelecida por Carpenter et al. (1992) quando do fundamento da rede AMF, descrita na Equação (2). Na segunda, testou-se a rede para uma normalização uniforme na qual cada elemento dos vetores era dividido pelo valor máximo da matriz completa de dados. Em outras palavras, identificou-se o maior valor da matriz de dados de treinamento (a que será referido por “valor base”) e dividiu-se cada elemento da mesma por tal, permitindo manter a proporção adequada entre eles e os valores resultantes entre 0 e 1.

A ideia principal de empregar outro método de normalização era comparar a eficiência da rede em cada diretiva. Isso porque, no caso da normalização convencional (CARPENTER et al.,1992), cada elemento pertencente a um dado vetor de entrada é dividido pela soma total dos valores do mesmo vetor. Em algumas situações, esta divisão resulta em elementos de ordem de magnitude extremamente baixa, podendo desequilibrar a estruturação das respostas *on* e *off* da rede AMF. Por este motivo, buscou-se testar as duas abordagens.

Os testes realizados foram ancorados apenas no recorte 3 de cada imagem, dado que apresentou os melhores resultados gerais no método anterior e

potencialmente representa a melhor síntese de informações para extração de atributos e codificação. Os melhores resultados são sumarizados na Tabela 6 abaixo.

Tabela 6 - Síntese dos melhores resultados para o histograma de imagens.

Método	*Baseline*	Taxa Redim.	Épocas	Cat C	CCR S	CCR NS	CCR G
NC	0	0,3 – 70% redução	1	16	6,67%	90,28%	84,85%
			4	20	13,33%	92,59%	87,45%
	0,7	0,5 – 50% redução	1	18	13,33%	94,91%	89,61%
			3	19	13,33%	94,91%	89,61%
NVB	0	0,5 – 50% redução	1	18	6,67%	92,59%	87,01%
			4	21	6,67%	93,98%	88,31%
	0,5	0,1 – 90% redução	1	6	0,00%	95,37%	89,18%
			3	13	6,67%	95,37%	89,61%
	0,9	0,1 – 90% redução	1	25	13,33%	91,20%	86,15%
			3	29	13,33%	92,13%	87,01%
		0,3 – 70% redução	1	41	0,00%	**96,30%**	**90,04%**
			3	45	0,00%	**96,30%**	**90,04%**

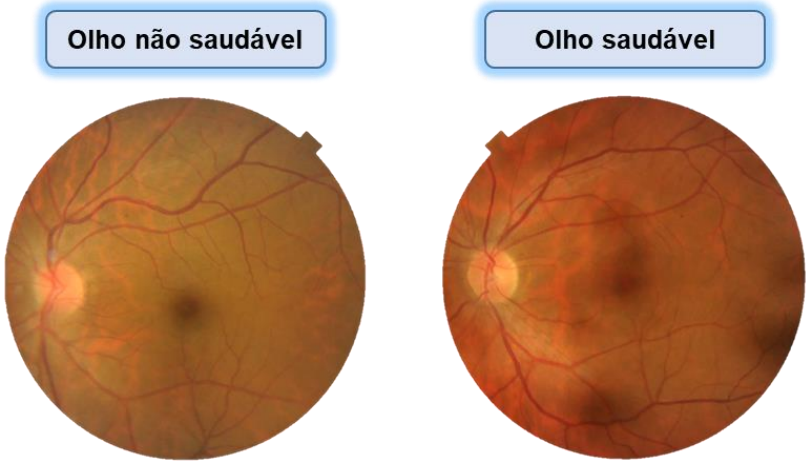
Fonte: Elaboração do próprio autor.
Legenda: NC = Normalização Convencional, NVB = Normalização pelo Valor Base.

Esta nova técnica de codificação sugerida apresentou resultados de taxa de classificação correta para o caso geral (saudáveis e não saudáveis em conjunto) próximos de 90%. Olhando pelo enfoque dos globos oculares não saudáveis separadamente, a eficiência na classificação se destacou, chegando à marca de 96,3% no caso de 70% de redução das imagens no redimensionamento e 0,9 *baseline*. Percebe-se que a média de tais resultados, em relação à codificação anterior proposta, foi superior em diversos dos testes apresentados.

No entanto, apresentou-se grande dificuldade na classificação dos olhos saudáveis individualmente. Assim como no caso anterior, esta seção de classificação foi a que demonstrou desempenho inferior em relação às outras. No melhor dos casos, a classificação correta dos saudáveis atingiu apenas 13,33%. Isso se deve, provavelmente, à quantidade inferior de amostras de olhos saudáveis disponível para treinamento da rede AMF, representando aproximadamente 26% do total existente para os olhos não saudáveis. Além disso, algumas doenças presentes no banco final de treinamento distinguiam-se pouco dos olhos absolutamente saudáveis. Um

exemplo disso diz respeito à presença de vasos coroides devido à redução da densidade dos pigmentos (TSLN) (PACHADE et al., 2021). Quando esta condição é leve, devido (majoritariamente) à idade do paciente, pouca diferença entre estas imagens e as imagens de olhos saudáveis é perceptível, mesmo analisando criteriosamente as duas fotografias postas lado a lado. Por conta deste motivo, é possível que a rede tenha dificuldade em categorizar adequadamente algumas amostras saudáveis.

Figura 18 - Caráter complexo de distinção em algumas amostras.



Fonte: Elaboração do próprio autor.

A Figura 18 evidencia uma amostra de olho não saudável, retirada do banco resultante de treinamento da rede, e uma amostra de olho saudável, retirado do banco de testes. Nela, é possível compreender o relatado anteriormente: o globo ocular à esquerda apresenta condição moderada de TSLN, indicando que deve ser categorizado como não sadio (PACHADE et al., 2021). Já o olho à direita na figura apresenta condição considerada regular e leve, pertencente ao avanço da idade do paciente e deve ser categorizado como saudável e em conformidade. Esta pequena distinção pode resultar na dificuldade de classificação correta da rede AMF em alguns casos, haja visto que até mesmo a análise por inspeção visual humana pode ser complexa.

Para o método de extração de atributos por histograma das imagens, verifica-se também que a rede não criou muitas categorias nos testes. Mesmo nos bons resultados, quando treinados em mais épocas até atingir a precisão de 100% no conjunto de treinamento, a rede criou no máximo 45 categorias no módulo ART_a. Os

testes apresentados com os melhores resultados variaram em questão de redimensionamento aplicado e parâmetro *baseline*, mas a porcentagem de acerto sempre se manteve basicamente na mesma ordem, não se distanciando muito. As duas normalizações aplicadas apresentaram eficiência geral satisfatória, ficando os melhores resultados absolutos a cargo da ancorada no valor base máximo da matriz de treinamento, mas diferenciando-se por pouco.

Apesar da média geral dos melhores resultados demonstrarem-se levemente superior à primeira codificação testada, o valor máximo absoluto obtido foi melhor transferindo as matrizes para representação de vetor. Além disso, o desempenho individual da primeira codificação no quesito dos olhos saudáveis isoladamente apresentou superioridade nos testes. O uso da extração de atributos por histograma de frequência de aparecimento de tonalidade de pixels é uma boa alternativa, mas a contagem de aparecimento pode não ser a melhor opção tratando-se de classes tão distintas como neste caso. Olhos não saudáveis englobam diversos casos com anomalias que podem se distinguir de forma acentuada. A generalização pelo histograma pode dificultar a isonomia e similaridade entre estes casos, promovendo um distanciamento entre diferentes imagens pertencentes à mesma categoria.

No caso do primeiro método de extração de atributos e codificação, como todas as informações contidas nos pixels são transferidas para um vetor unidimensional, as imagens pertencentes à classe dos olhos saudáveis (por serem relativamente semelhantes) apresentam distribuição de informações uniforme em cada vetor, favorecendo o agrupamento e identificação nos testes (daí a porcentagem de acerto superior em relação ao histograma de imagens). No caso da segunda codificação, se por coincidência a contagem de aparecimento de pixels em uma captura de olho não saudável se assemelhar à de um olho saudável, ainda que as imagens sejam visivelmente distintas, o treinamento pode direcionar à associação destas informações e promover classificação errônea no momento dos testes.

Ainda assim, pelos resultados apresentados na Tabela 6, percebe-se que os testes da segunda codificação apresentaram eficiência no caso geral com média superior aos testes da primeira codificação (mesmo que o máximo resultado absoluto tenha sido obtido na primeira). Para trabalhos futuros, pela eficiência no caso geral observada, sugere-se o acréscimo de algum tipo de variável distinta ao vetor de características proveniente do histograma das imagens para compor os vetores finais de treinamento. Atrelando alguma outra variável (vinculada às imagens) aos

histogramas representativos, o método pode se potencializar e melhorar a precisão da classificação dos olhos saudáveis. Uma sugestão seria a extração de alguns pontos de interesse das figuras e complemento dos histogramas com tais características extraídas, promovendo possivelmente mais assertividade e melhora dos resultados de cada classe isolada.

5.3 ESTADO DE SAÚDE DE GLOBO OCULAR - BOVW

O método de extração de atributos por BoVW também possuiu enfoque no recorte 3 com focalização no centro de cada captura. Os pontos de interesse foram extraídos desta região e depois traduzidos em características, tal como explicitado no tópico 4.3.3. Como dito anteriormente, optou-se pela seleção de 100 palavras visuais para compor o dicionário. Ainda assim, a caráter de teste e comparação, efetuou-se algumas análises para 300 palavras, uma quantidade 3 vezes superior.

A normalização empregada para estes casos foi a convencional estabelecida em Carpenter et al. (1992) (Equação (2)), dado que os vetores de características possuem dimensões baixas (dependendo da quantidade de palavras visuais) e elementos com valores numéricos pequenos. Além disso, optou-se pela utilização do redimensionamento em 50% nas imagens, procurando reduzir o mínimo possível e manter a maior quantidade de informações nos processos de treinamento da rede e sucessivos testes. Os melhores resultados encontrados são dispostos na Tabela 7 a seguir.

Tabela 7 - Síntese dos melhores resultados para codificação BoVW.

QPV	*Baseline*	Épocas	Cat C	CCR S	CCR NS	CCR G
K=100	0	1	5	46,67%	94,91%	**91,77%**
		3	7	53,33%	93,52%	90,91%
	0,9	1	11	**66,67%**	85,65%	84,42%
		3	12	46,67%	89,81%	87,01%
K=300	0,5	1	3	46,67%	94,91%	**91,77%**
		3	4	40,00%	**95,37%**	**91,77%**
	0,9	1	7	40,00%	94,44%	90,91%
		3	9	40,00%	94,44%	90,91%

Fonte: Elaboração do próprio autor.
Legenda: QPV = Quantidade de Palavras Visuais.

Analisando individualmente cada uma das taxas corretas de classificação (caso geral, não saudáveis e saudáveis), o terceiro método de codificação apresentou a melhor distribuição de eficiência de todos os outros. A menor porcentagem encontrada de acerto para os olhos saudáveis foi de 40%, o que **representa a maior porcentagem de acerto para esta mesma classe quando comparado com os métodos anteriores.** Mesmo com a dificuldade de distinção que pode ocorrer em algumas ocasiões (ver Figura 18), a detecção de pontos de interesse em cada imagem e a extração consecutiva das características trata-se de um artifício valioso para identificação de informações de cada captura. A associação de dados pela rede passa a ser mais robusta e padronizada com este método, uma vez que independente da imagem a ser analisada apenas os pontos principais de destaque serão utilizados para codificação em cada caso. Desta forma, as informações pertencentes aos pontos que o algoritmo KAZE detector considera menos importantes são desconsiderados, favorecendo apenas a preservação de informações mais relevantes.

Isso não ocorre com os outros métodos: neles, todos os pontos existentes no recorte empregado nas imagens são considerados na montagem do vetor de características. No caso da conversão da matriz de dados em vetor, qualquer pixel pertencente à matriz final é transferido para a estrutura unidimensional, enquanto no histograma de imagens todas as repetições de valores de uma mesma tonalidade são contabilizadas.

Fica evidente a robustez do BoVW e a assertividade consequente obtida pela rede AMF. No melhor caso, mesmo com a grande similaridade entre as classes diferentes em alguns casos, a rede conseguiu classificar corretamente 67% dos saudáveis. Para o caso geral e o dos não saudáveis, boa parte dos testes apresentou eficiência superior a 90% de acertos. A utilização de mais palavras visuais pode discriminar melhor as informações de cada vetor representativo das imagens; no entanto, pode deixá-los desnecessariamente maiores e zerar a contagem de várias palavras por imagem. Além disso, empregar um dicionário com três vezes mais palavras aumenta o custo computacional consideravelmente no momento de criação do dicionário, quase triplicando o tempo para conclusão de tal etapa pré-treinamento.

Apesar da melhor distribuição de eficiência no método BoVW, os maiores valores absolutos no quesito de classificação geral (saudáveis e não saudáveis em conjunto) e classificação de globos não saudáveis isoladamente ainda pertenceram às outras formas de extração de atributos, mesmo que por uma diferença percentual

trivial. Também é possível notar que na abordagem BoVW a rede criou a menor quantidade de categorias no módulo ART_a em relação às outras duas.

5.4 ANÁLISE COMPARATIVA

Os três métodos de extração de atributos e codificação possuem particularidades vantajosas distintas, assim como alguns pontos de enfoque.

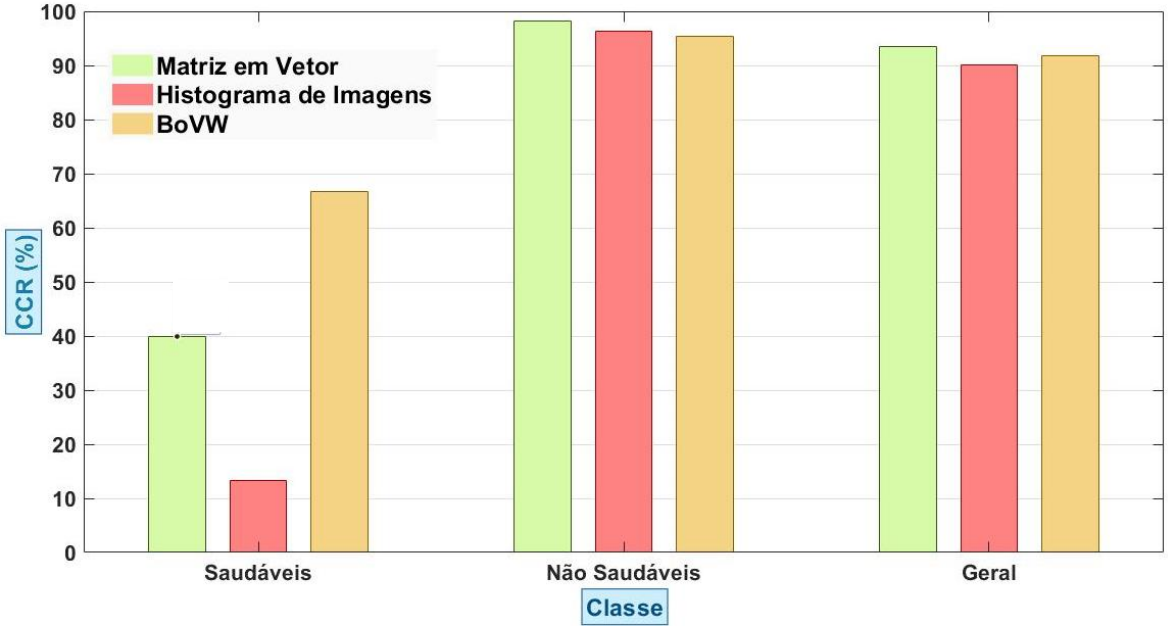
No caso da matriz em vetor, a facilidade de criação dos vetores de características representa a grande vantagem da abordagem, uma vez que basta efetuar a leitura dos dados e transferi-los para um formato unidimensional. Ainda, este método armazena grande quantidade de informações de cada imagem, permitindo que todas as nuances da captura sejam devidamente representadas. O maior percalço desta codificação reside no tamanho resultante de cada vetor representativo, possuindo diversos elementos, mesmo com uma alta taxa de redimensionamento e redução. Dito isto, é essencial otimizar a taxa de redução aplicada a cada fotografia com a quantidade ideal de armazenamento de dados: assim, este é um método que se atrela muito na necessidade de fazer sucessivos testes com mudança de valores a fim de obter o melhor resultado possível.

No segundo caso, o histograma de imagens também armazena todas as informações das capturas, mas de forma indireta. A contagem do aparecimento da mesma tonalidade de pixels representa uma vantagem no ponto de vista da facilidade de codificação, igualmente ao primeiro. Além disso, nesta abordagem os vetores possuem tamanho fixo de 256, o que potencializa o processamento computacional e permite trabalhar com uma quantidade muito menor de elementos. O ônus do método reside na inexistência de dados espaciais das imagens, sendo contabilizada apenas o aparecimento de cada valor e nenhuma informação a respeito de sua respectiva disposição física no globo ocular (algo que não ocorre no primeiro, já que as informações são armazenadas mediante o aparecimento espacial de cada tonalidade, traduzidas na matriz lida de cada fotografia).

O terceiro método é o mais robusto, identificando pontos de interesse de destaque de cada imagem e trabalhando com as características extraídas de tais. A criação do dicionário de palavras permite que se fixe os vetores em um tamanho igual ao número de palavras visuais escolhido. Assim, além de trabalharmos com um vetor de pequena dimensão em relação aos outros métodos, ele é capaz de sintetizar as

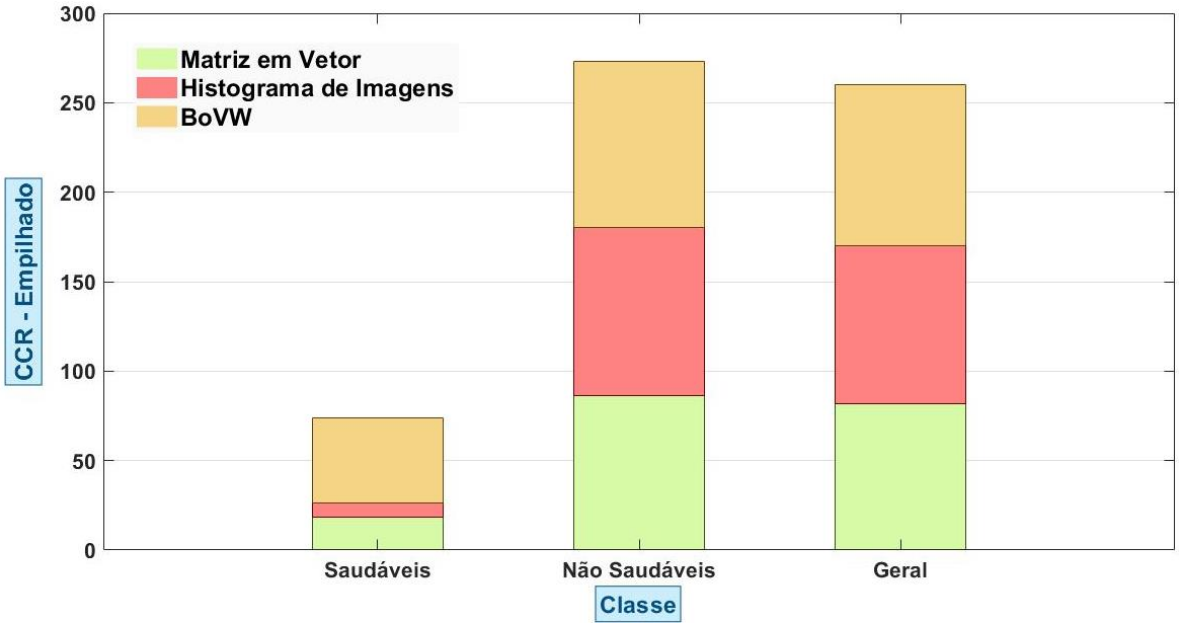
informações mais importantes de cada imagem. Por outro lado, essa codificação é mais complexa do ponto de vista computacional, passando por várias etapas e destoando da facilidade dos outros métodos. O custo computacional é maior, uma vez que é necessário usar o algoritmo *k-means* para agrupamento e criação dos centroides que representarão as palavras visuais do método.

Figura 19 - Comparação da melhor taxa de classificação obtida em cada método.



Fonte: Elaboração do próprio autor.

Figura 20 - Comparativo das médias dos melhores resultados das codificações.



Fonte: Elaboração do próprio autor.

A Figura 19 e a Figura 20 permitem analisar visualmente a eficiência de cada um dos métodos de codificação no que tange a melhor taxa de classificação obtida e a média dos melhores resultados (evidenciados nas tabelas do decorrer do capítulo). No caso da média, nota-se como a ferramenta BoVW se sobressai em relação às outras, destoando-se principalmente nos acertos relacionados aos olhos saudáveis. O artifício do histograma de imagens se destaca levemente na precisão de acerto para os globos oculares com algum tipo de patologia. No caso da análise geral, pelo gráfico empilhado nota-se que as três abordagens apresentam resultados médios semelhantes, sobrepujando-se o BoVW.

Figura 21 – Evolução percentual média com o treinamento em várias épocas.

Fonte: Elaboração do próprio autor.

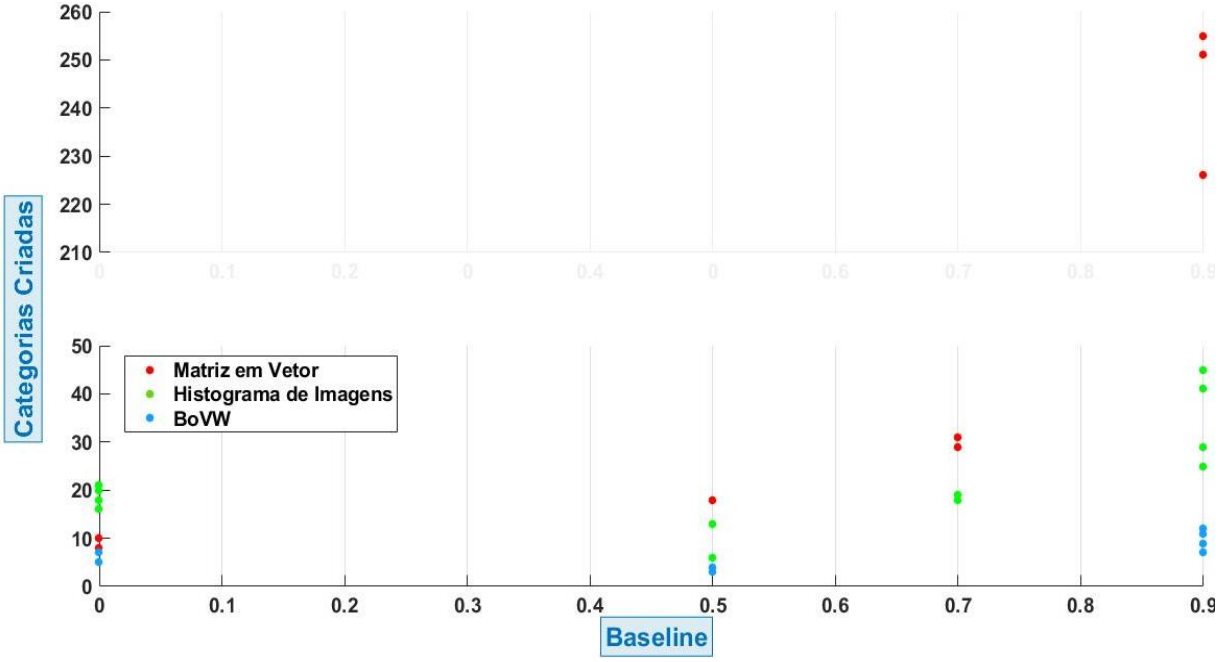
A Figura 21 evidencia como o treinamento sucessivo da rede, em várias épocas até atingir 100% da precisão nas amostras de treinamento, pode influenciar no desempenho da taxa de classificação em cada caso. Empregou-se as médias dos melhores resultados para cada abordagem de codificação. Os gráficos representam a evolução percentual de acertos em dois casos: o primeiro (à esquerda de cada figura) representa a referência (centrada em 0) dos resultados obtidos com 1 época de treinamento; o segundo (à direita de cada figura, o “término” do gráfico) representa o quanto a taxa de classificação correta evoluiu (ou regrediu) com o treinamento efetuado em diversas épocas. Facilmente nota-se que, de uma maneira geral, a

eficiência da taxa de classificação tende a subir com o treinamento em várias épocas sucessivas. De fato, isto foi estabelecido por Carpenter et al. (1992) em um dos testes que fundamentaram a proposição da rede AMF pela primeira vez.

No caso do presente trabalho, é perceptível a melhora em alguns casos particulares. Para a codificação por histograma de imagens, alcançou-se um aumento médio superior a 2% de classificação no caso dos olhos saudáveis. No caso do BoVW, apesar da melhora no caso geral e dos não saudáveis ser possível, observa-se um decréscimo quando da análise dos globos oculares saudáveis. A generalização pelo treinamento sucessivas vezes pode readequar os valores dos pesos da rede e é possível haver a ocorrência de situações nas quais aumente-se a quantidade de falsos negativos em detrimento da melhora percentual de outras classes. De um ponto de vista mais amplo, é possível notar que a técnica de treinamento sucessivo se mostra como boa alternativa para potencializar alguns acertos, mas exige um maior custo computacional (pelas iterações sucessivas). A análise que deve ser determinada é se este custo adicional compensa o avanço na eficiência proporcionado pelo método.

A Figura 22 mostra a quantidade total de categorias criadas em cada um dos resultados apresentados de cada método, para cada *baseline* destacado nas tabelas demonstrativas anteriores (Tabelas 5, 6 e 7).

Figura 22 - Número de categorias criadas nos métodos variando-se o *baseline*.



Fonte: Elaboração do próprio autor.

Como não houve nenhum valor no intervalo central (de 50 a 210), esta porção foi retirada do gráfico a fim de permitir melhor visualização geométrica dos pontos.

De uma maneira geral, é possível notar como o método que possui a maior quantidade de elementos por vetor de características (matriz em vetor) se sobressai na quantidade de categorias criadas no módulo ART_a após o treinamento da rede, com alguns casos figurando até na faixa de 250. Por outro lado, nota-se a baixa quantidade de categorias distintas concebidas no caso do BoVW (menor tamanho do vetor de características) ainda que com parâmetros *baseline* mais elevados, como 0,9.

Tabela 8 - Comparativo descritivo dos métodos de codificação.

Método	**Dificuldade para Implementação**			**Desempenho**		
	Codificação	Custo Computac.	Tamanho do Vetor Final	G	S	NS
Matriz/Vetor	Fácil	Alto	Muito grande	Bom	Ruim	Ótimo
Histograma	Fácil	Pequeno	Razoável	Ótimo	Muito Ruim	Excelente
BoVW	Laboriosa	Muito Alto	Razoável	Ótimo	Razoável	Excelente

Fonte: Elaboração do próprio autor.

A Tabela 8 compara os três métodos por um ponto de vista da complexidade de implementação e desempenho resultante, estabelecendo critérios de rotulação mediante os resultados encontrados para cada um. Nesta tabela, é possível ver que os resultados obtidos pelo método BoVW, com relação ao seu desempenho, não apresentaram nenhum ponto ruim, mesmo com as particularidades do *dataset*. No entanto, este método possui uma dificuldade/complexidade maior na obtenção dos atributos, exigindo um custo computacional superior aos demais. Trata-se de uma codificação robusta para as amostras; todavia, possui diversas etapas que fazem com que a codificação seja laboriosa e o custo computacional muito alto, principalmente na determinação da quantidade de palavras visuais a ser utilizada (e consequente implementação) por meio do método de agrupamento *k-means clustering*.

6 CONCLUSÕES

O presente trabalho apresentou uma metodologia pioneira para detecção de estado de saúde de globos oculares baseada no reconhecimento de imagens. Almejou-se empregar a rede ARTMAP *Fuzzy* para efetuar as referidas classificações, correspondendo a uma abordagem inédita para tal tema. Utilizou-se um banco de dados estruturado com imagens de fundos de retina de globos oculares. Diversas seleções manuais e manipulações foram efetuadas no *dataset* até obter o banco resultante e padronizado.

Três processos diferentes de extração de atributos das imagens foram aplicados a fim de testar e comparar as respectivas eficiências por meio da métrica de taxa de classificação correta. O primeiro método, proposto no presente trabalho, referiu-se à conversão completa da matriz de dados equivalente a cada captura em um vetor unidimensional. No segundo, as frequências de aparecimento de uma determinada tonalidade representativa de pixel foram contabilizadas e estruturadas na forma de vetor de características para treinamento. Por último, a terceira abordagem abrangeu a utilização do conceito de *Bag of Visual Words*, elaborando-se um dicionário de palavras visuais mediante detecção de pontos de interesse KAZE, extração de características e aplicação do algoritmo de agrupamento *k-means*.

Para o treinamento da rede no caso do método Matriz em Vetor, três tipos de recortes foram efetuados nas imagens do *dataset* mencionado com diferentes graus de incidência: um excetuando-se apenas o disco óptico, outro as porções mais externas e, por último, um com focalização central nas capturas. Pelos resultados obtidos, priorizou-se a análise dos métodos subsequentes atrelada na terceira abordagem de recorte.

Pelos resultados obtidos com as amostras pertencentes ao banco de testes, os três métodos atingiram percentuais satisfatórios para o caso geral (analisando conjuntamente olhos saudáveis e não saudáveis), com desempenho médio ligeiramente superior no caso do BoVW, atingindo aproximadamente 90% e ficando muito próximo do modelo do histograma. No caso do desempenho absoluto máximo obtido para um teste individual, obteve-se 93,51% na análise com *baseline* 0 no caso da codificação matriz em vetor. No entanto, pelo desbalanceamento dos dados, é importante se analisar o comportamento médio de todos os testes efetuados a fim de garantir a confiabilidade do sistema, e não apenas o maior valor obtido isoladamente

em um dos testes. Por este enfoque, o método BoVW apresentou claro melhor desempenho.

Observando individualmente a média de classificação correta dos olhos não saudáveis, fica evidente novamente a assertividade dos métodos, com pequena superioridade da abordagem do histograma (atingindo quase 94%) seguida muito próxima pelo BoVW. No caso da Matriz em Vetor, atingiu-se média de 86% para este caso.

Por fim, analisando o panorama dos olhos saudáveis individualmente é que se percebe uma discrepância maior dos métodos de extração de atributos empregados. O histograma de imagens apresentou uma média de resultados abaixo dos 10% de classificação correta, representando o método com maior dificuldade de discernir corretamente este aspecto. Isso se deve provavelmente à forma de codificação em si, na qual a contagem de aparecimento de tonalidades estruturada no vetor de características não traz informações espaciais diretas de cada captura, apenas informações qualitativas que podem provocar associações incorretas. Olhando pelo enfoque dos saudáveis exclusivamente, percebe-se como o método BoVW se sobressai e mantém a melhor distribuição de acertos em relação aos outros, chegando a um valor máximo de acerto de quase 70%.

Os resultados também mostram como, de uma maneira geral, a média de taxa de classificação correta dos métodos cresce ligeiramente mediante o treinamento da rede em sucessivas épocas até atingir 100% de precisão nas amostras de treinamento. No entanto, a pequena melhoria percentual obtida não justifica o custo computacional superior decorrente do treinamento sucessivo para tal implementação. É possível analisar, por fim, como os métodos detentores da maior quantidade de elementos por vetor de características acarretam uma maior criação de categorias no módulo ART_a com o aumento do *baseline*, chegando a 255 no método da Matriz em Vetor e apenas 12 no BoVW (que possui vetores com apenas 100 ou 300 elementos).

A comparação das formas de extração de atributos e codificações evidencia como a robustez do BoVW permite uma melhor distribuição de acertos em cada uma das análises efetuadas, ainda que o processo para criação dos vetores representativos de cada imagem seja mais árduo e com maior custo computacional, envolvendo uma boa quantidade de etapas para implementação. Método como o histograma de imagens, que envolve codificação mais facilitada e vetor não muito extenso, permite acerto geral satisfatório e custo computacional reduzido, mas acaba

deixando a desejar em assertividades individuais e carece de algum complemento na codificação. O modelo de matriz em vetor acaba figurando entre os dois supracitados, exigindo um certo custo computacional e envolvendo uma codificação simples, mas com muitos elementos.

6.1 PROPOSTAS PARA TRABALHOS FUTUROS

Os métodos demonstraram desempenho satisfatório no que tange a eficiência geral da taxa de classificação do estado de saúde dos globos oculares, com destaque para a codificação BoVW por apresentar também boa assertividade no que diz respeito à classificação dos saudáveis individualmente. Com a implementação de alguns tópicos, é possível potencializar cada uma das abordagens para resultados ainda mais promissores.

Algumas sugestões para eventuais melhorias são listadas a seguir:

1) **Matriz em Vetor:** Redução da taxa de dimensionalidade aplicada e inserção de mais elementos no vetor representativo de cada fotografia. Empregar técnicas de preservação dos componentes principais, como o PCA, para trabalhar com uma menor quantidade de elementos;
2) **Histograma de Imagens:** Complementar a codificação com algum outro método que relacione características espaciais das figuras, podendo até ser considerada detecção de pontos de interesse e extração de características;
3) **BoVW:** Teste com outras estruturas de detecção de pontos de interesse além das características KAZE. Integrar as imagens centralizadas no disco óptico no *dataset* para treinamento, testando também abordagem sem recortes;
4) **Classificação de Doenças:** Empregar as técnicas do presente trabalho para classificação de vasta gama de patologias oculares em olhos não saudáveis;
5) **Dataset:** Integrar a utilização de outro *dataset* a fim de aumentar a quantidade de amostras sem ruído existentes para treinamento e consequente melhoria da eficiência, dado que inúmeras exclusões foram necessárias para eliminar exemplares ruins do banco empregado;
6) **Implementação:** Futuramente, pode-se buscar integrar o presente sistema em dispositivos inteligentes que façam escaneamento imediato da retina e permitam a detecção de possíveis anomalias em pacientes.

REFERÊNCIAS

ALCANTARILLA, P. F.; BARTOLI, A.; DAVISON, A. J. KAZE Features. *In*: EUROPEAN CONFERENCE ON COMPUTER VISION – ECCV, 21., 2012, Florença. **Anais** [...] Heidelberg: Springer, 2012. p. 214-227.

AL-OBAYDY, W. N. I.; SUANDI, S. A. Open-set single-sample face recognition in video surveillance using fuzzy ARTMAP. **Neural Computing and Applications,** Londres, v. 32, p. 1405-1412, 2020.

AMIDI, A.; AMIDI, S. **The evolution of image classification explained**. Stanford Educational Blog. Disponível em: https://stanford.edu/~shervine/blog/evolution-image-classification-explained. Acesso em: 02/06/2023.

AMORIM, D. G. **Redes ART com categorias internas de geometria irregular**. 2006. 352 p. Tese (Doutorado em Física) – Departamento de Eletrônica e Computação, Universidade de Santiago de Compostela, Santiago de Compostela, 2006.

ANAGNOSTOPOULOS, G.; GEORGIOPOULOS, M. Category regions as new geometrical concepts in Fuzzy-ART and Fuzzy-ARTMAP. **Neural Networks**, Oxford, v. 15, n. 10, p. 1205-1221, 2002.

ARTHUR, D.; VASSILVITSKII, S. k-means++: The Advantages of Careful Seeding. *In*: PROCEEDINGS OF THE EIGHTEENTH ANNUAL ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS (SODA '07), 18., 2007, Nova Orleans. **Anais** [...] Philadelphia: Society for Industrial and Applied Mathematics, 2007. p. 1027-1035.

ATWANY, M. Z.; SAHYOUN, A. H.; YAQUB, M. Deep Learning Techniques for Diabetic Retinopathy Classification: A Survey. **IEEE Access,** Piscataway, v. 10, p. 28642-28655, 2022.

BAY, H; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Speeded-Up Robust Features (SURF). **Computer Vision and Image Understanding,** Nova Iorque, v. 110, n. 3, p. 346-359, 2008.

CANO-IZQUIERDO, J. M.; SÁNCHEZ, E. G.; BRAVO, M. J. A.; PINZOLAS, M.; IBARROLA, J. How well Fuzzy ARTMAP approximates functions? **Journal of Intelligent & Fuzzy Systems,** Amsterdam, v. 25, n. 2, p. 335-350, 2013.

CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. **Computer Vision, Graphics, and Image Processing,** San Diego, v. 37, n. 1, p. 54-115, 1987.

CARPENTER, G. A.; GROSSBERG, S. ART2: Stable self-organization of pattern recognition codes for analog input patterns. **Applied Optics,** Washington, v. 26, n. 23, p. 4919–4930, 1987.

CARPENTER, G. A.; GROSSBERG, S. Self-organizing neural network architectures for real-time adaptive pattern recognition. *In*: INTERNATIONAL SYMPOSIUM ON

NEURAL AND SYNERGETIC COMPUTERS, 1., 1988, Krün. **Anais** [...] Heidelberg: Springer-Verlag, 1988. p. 42-74.

CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. **Neural Networks**, Oxford, v. 4, n. 6, p. 759–771, 1991.

CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. **Neural Networks,** Oxford, v. 4, n. 5, p. 565-588, 1991.

CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H.; ROSEN, D. B.; MARKUZON, N. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. **IEEE Transactions on Neural Networks,** Piscataway, v. 3, n. 5, p. 698-713, 1992.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. *In*: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, San Francisco. **Anais** [...] Nova Iorque: Association for Computing Machinery, 2016. p. 785-794.

DALAL, N.; TRIGGS, B. Histograms of Oriented Gradients for Human Detection. *In*: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR'05), 1., 2005, San Diego. **Anais** [...] [S.l]: IEEE, 2005. p. 886-893.

DENG, L.; YU, D. **Deep Learning:** methods and applications. Boston: Now Publishers, 2014. v. 7. 195 p.

EUROPEAN CONFERENCE ON MACHINE LEARNING – ECML-95, 8., Iráclio. **Anais** [...] Heidelberg: Springer, 1995. v. 912.

GEORGIOPOULOS, M.; HUANG, J.; HEILEMAN, G. Properties of learning in ARTMAP. **Neural Networks,** Oxford, v. 7, n. 3, p. 495-506, 1994.

GEORGIOPOULOS, M.; FERNLUND, H.; BEBIS, G.; HEILEMAN, G. L. Order of Search in Fuzzy ART and Fuzzy ARTMAP: Effect of the Choice Parameter. **Neural Networks,** Oxford, v. 9, n. 9, p. 1541-1559, 1996.

GONZALEZ, R. C; WOODS, R. E. **Processamento Digital de Imagens**. 3. ed. São Paulo: Pearson Universidades, 2009. 624 p.

GROSSBERG, S. Adaptive Pattern Classification and Universal Recoding: I. Paralel development and coding of neural feature detectors. **Biological Cybernetics,** Heidelberg, v. 23, p. 121-134, 1976.

GROSSBERG, S. Adaptive Pattern Classification and Universal Recoding: II. Feedback, Expectation, Olfaction, Illusions. **Biological Cybernetics,** Heidelberg, v. 23, n. 4, p. 187-202, 1976.

GROSSBERG, S. Nonlinear neural networks principles, mechanisms and architectures. **Neural Networks,** Oxford, v. 1, n. 1, p. 17-61, 1988.

GROSSBERG, S.; STONE, G. Neural Dynamics of Word Recognition and Recall: Attentional Priming, Learning, and Resonance. **Psychological Review,** Washington, v. 93, n. 1, p. 46-74, 1986.

HAYKIN, S. **Neural networks:** a comprehensive foundation. 2. ed. Upper Saddle River: Prentice-Hall, 1999. 842 p.

HIRSCH, R. **Exploring colour photography:** a complete guide. Londres: Laurence King Publishing, 2004. 360 p.

JOLLIFE, I. T. **Principal component analysis**. 2. ed. Nova Iorque: Springer, 2002. 488 p.

KARTALOPOULOS, S. V. **Understanding neural networks and fuzzy logic:** basic concepts and applications. 1. ed. Hoboken: Wiley-IEEE Press, 1995. 232 p.

KIM, J. et al.; Machine learning predicting myopic regression after corneal refractive surgery using preoperative data and fundus photography. **Graefe's Archive for Clinical and Experimental Ophthalmology**, Heidelberg, v. 260, n. 11, p. 3701-3710, 2022.

KRZANOWSKI, W. J. **Principles of multivariate analysis**. Revised edition. Oxford: Oxford University Press, 2000. 608 p.

LIU, C.; WECHSLER, H. Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. **IEEE Transactions on Image Processing**, Nova Iorque, v. 11, n. 4, p. 467-476, 2002.

LOPES, M. L. M. **Desenvolvimento de redes neurais para previsão de cargas elétricas de sistemas de energia elétrica**. 2005. 169 p. Tese (Doutorado em Engenharia Elétrica) — Faculdade de Engenharia, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Ilha Solteira, 2005.

LOWE, D. G. Distinctive Image Features from Scale-invariant Keypoints. **International Journal of Computer Vision**, Amsterdam, v. 60, n. 2, p. 91-110, 2004.

MATHWORKS Image Recognition. What is Image Recognition? Disponível em: https://www.mathworks.com/discovery/image-recognition-matlab.html. Acesso em: mai. 2023.

MATHWORKS Help Center. Imread: Read image from graphics file. Disponível em: https://www.mathworks.com/help/matlab/ref/imread.html. Acesso em: mar. 2023.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, Nova Iorque, v. 5, n. 4, p. 115–133, 1943.

OKAWA, M. From BoVW to VLAD with KAZE features: Offline signature verification considering cognitive processes of forensic experts. **Pattern Recognition Letters**, Amsterdam, v. 113, p. 75-82, 2018.

PACHADE, S. et al. Retinal Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research. **MDPi Data**, Basel, v. 6, n. 2, 14, 2021.

PIURI, V.; RAJ, S.; GENOVESE, A.; SRIVASTAVA, R. **Trends in deep learning methodologies:** algorithms, applications, and systems. 1. ed. Cambridge: Academic Press, 2020. 306 p.

RAN, A. R et al. Deep learning in glaucoma with optical coherence tomography: a review. **Eye: The Scientific Journal of the Royal College of Ophtalmologists**, Londres, v. 35, n. 1, p. 188-201, 2021.

SENGAR, N.; JOSHI, R. C.; DUTTA, M. K.; BURGET, R. EyeDeep-Net: a multi-class diagnosis of retinal diseases using deep neural network. **Neural Computing and Applications**, Londres, v. 35, p. 10551-10571, 2023.

TUFAIL, A. B. et al. Diagnosis of Diabetic Retinopathy through Retinal Fundus Images and 3D Convolutional Neural Networks with Limited Number of Samples. **Wireless Communications and Mobile Computing**, Londres, v. 2021, 15 p, 2021. Disponível em: https://doi.org/10.1155/2021/6013448. Acesso em: 02 fev. 2023.

VIMINA, E. R.; JACOB, K. P. Feature fusion method using BoVW framework for enhancing image retrieval. **IET Image Processing,** Londres, v. 13, p. 1979-1985, 2019.

WASSERMAN, P. D. **Neural computing**: theory and practice. Nova Iorque: Van Nostrand Reinhold, 1989. 230 p.

WEENINK, D. Category ART: a Variation on Adaptative Resonance Theory Neural Networks. *In*: IFA PROCEEDINGS, 21, 1997, Amsterdam. **Anais** [...]. Amsterdam: [s.n], 1997. p. 117-129.

ZADEH, L. A. Fuzzy sets. **Information and control**. Pequim, v. 8, n. 3, p. 338-353, 1965.

APÊNDICE A – INFORMAÇÕES ADICIONAIS DO *DATASET*

Como relatado anteriormente, antes de qualquer manipulação e exclusão de fotografias com qualidade ruim, o conjunto de dados possuía 3200 imagens divididas em três pastas principais: treinamento, validação e testes. Constituindo tal conjunto existiam amostras de olhos saudáveis, isentos de qualquer tipo de doença, e olhos não saudáveis (ou anormais), acometidos por uma ou mais das patologias existentes.

A classificação das imagens do banco de dados foi realizada com anuência de dois especialistas da área (PACHADE et al., 2021). Foram consideradas para listagem 46 condições oculares distintas para preenchimento da composição do *dataset*. Dentre estas, 45 referiam-se a condições vinculadas a patologias e 1 a olhos perfeitamente saudáveis. Com o auxílio dos especialistas, as 3200 imagens utilizadas foram atreladas a alguma destas 46 condições listadas, sendo que determinados globos oculares poderiam ser acometidos por mais de uma condição. No *dataset* inicial, antes de qualquer alteração e manipulação efetuada, 60% das imagens eram destinadas ao treinamento, 20% à validação e 20% à fase de testes.

A Tabela 9 apresenta a distribuição de patologias na pasta de treinamento do *dataset* utilizado, relacionando as doenças (ou condições oculares) com suas respectivas quantidades iniciais de amostras (PACHADE et al., 2021). Desta forma, é possível analisar qual a quantidade de anomalias existentes para treinamento no *dataset* inicial, verificando quais condições apresentam-se mais frequentemente e quais caracterizam-se por sua raridade de ocorrência.

Pode-se notar que nem todas as condições listadas pelos especialistas que analisaram as capturas apareceram nos pacientes que visitaram a clínica referência para coleta de dados (PACHADE et al., 2021). Além disso, nota-se que algumas patologias se manifestaram em pouquíssimas ocasiões, como no caso do descolamento epitelial hemorrágico (uma ocorrência apenas). Em contrapartida, outras condições aparecem com um percentual mais assíduo, como a retinopatia diabética (376 casos registrados para treinamento inicialmente). Este panorama da distribuição de frequência de aparecimento dos casos serve de base para um eventual trabalho, ancorado neste *dataset*, que almeje classificar patologias distintas (classificação multi-classe). No caso desta possível abordagem, sugerida inclusive no tópico relacionado a propostas para trabalhos futuros, é interessante considerar apenas as anomalias mais frequentes, visando perpetuar a maior quantidade de

amostras possível para a fase de treinamento, garantindo uma maior assertividade do sistema classificatório.

Tabela 9 - Distribuição das doenças para o treinamento no *dataset* inicial.

Condição Ocular	Número de Imagens	Condição Ocular	Número de Imagens
Normal/Saudável	401	Buraco macular (MHL)	11
Retinopatia diabética (DR)	376	Telangiectasia parafoveal (PT)	11
Opacidade da mídia (MH)	317	Retinite pigmentosa (RP)	6
Escavação disco óptico (ODC)	282	Vasos tortuosos oculares (TV)	6
Fundo ocular tesselado (TSLN)	186	Shunts Optociliares (ST)	5
Drusa ocular (DN)	138	Ruptura Coroidal Pós-Traumática (PTCR)	5
Miopia (MYA)	101	Edema macular cistóide (CME)	4
Degeneração macular pela idade (ARMD)	100	Mancha algodonosa (CWS)	3
Oclusão de ramo da veia retiniana (BRVO)	73	Dobras coroidais (CF)	3
Palidez do disco óptico (ODP)	65	Disco óptico inclinado (TD)	3
Edema do disco óptico (ODE)	58	Fibra nervosa mielinizada (MNF)	3
Cicatrizes de laser (LS)	47	Oclusão do ramo da artéria retiniana (BRAO)	2
Retinite (RS)	43	Oclusão da artéria retiniana central (CRAO)	2
Retinopatia Serosa Central (CSR)	37	Hemorragia pré-retiniana (PRH)	2
Coriorretinite (CRS)	32	Vasos retinianos colaterais (CL)	1
Oclusão da veia retiniana central (CRVO)	28	Coloboma (CB)	1
Alteração do epitélio pigmentar da retina (RPEC)	22	Descolamento epitelial hemorrágico (HPED)	1
Neuropatia óptica isquêmica anterior (AION)	17	Macroaneurisma retiniano (MCA)	1
Hialose asteroide (AH)	16	Placa retiniana (PLQ)	1
Cicatriz macular (MS)	15	Vasculite retiniana (VS)	1
Exsudação Ocular (EDN)	15	Hemorragia vítrea (VH)	1
Tração Retiniana (RT)	14	Maculopatia de fosseta de disco óptico (ODPM)	0
Membrana epirretiniana (ERM)	14	Retinopatia hemorrágica (HR)	0

Fonte: Adaptado de Sengar et al. (2023).

Ressalta-se que, após as devidas exclusões e manipulações do *dataset* (descritas no Capítulo 4), as imagens vinculadas à centralização no disco óptico (assim como anomalias exclusivas de tal porção) foram desconsideradas no processo de treinamento do presente trabalho.