



UNIVERSIDADE FEDERAL DE GOIÁS (UFG)

INSTITUTO DE INFORMÁTICA (INF)

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO

HEDENIR MONTEIRO PINHEIRO

**Reflexo Pupilar à Luz como
Biomarcador para Identificação de
Glaucoma: Avaliação Comparativa de
Redes Neurais e Métodos de
Aprendizado de Máquina**

Goiânia
2024



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Hedenir Monteiro Pinheiro

3. Título do trabalho

Reflexo Pupilar à Luz como Biomarcador para Identificação de Glaucoma: Avaliação Comparativa de Redes Neurais e Métodos de Aprendizado de Máquina

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

- a) consulta ao(à) autor(a) e ao(à) orientador(a);
 - b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
- O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Ronaldo Martins Da Costa, Professor do Magistério Superior**, em 23/09/2024, às 13:28, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Hedenir Monteiro Pinheiro, Discente**, em 25/09/2024, às 20:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4846574** e o código CRC **953000C0**.

HEDENIR MONTEIRO PINHEIRO

Reflexo Pupilar à Luz como Biomarcador para Identificação de Glaucoma: Avaliação Comparativa de Redes Neurais e Métodos de Aprendizado de Máquina

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Informática (INF), da Universidade Federal de Goiás (UFG), como requisito para obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Linha de Pesquisa: Sistemas Inteligentes e Aplicações

Orientador: Prof. Ronaldo Martins da Costa

Goiânia

2024

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Pinheiro, Hedenir Monteiro

Reflexo Pupilar à Luz como Biomarcador para Identificação de Glaucoma: Avaliação Comparativa de Redes Neurais e Métodos de Aprendizado de Máquina [manuscrito] / Hedenir Monteiro Pinheiro. - 2024.

CXXXIV, 134 f.: il.

Orientador: Prof. Ronaldo Martins da Costa.

Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Informática (INF), Programa de Pós-Graduação em Ciência da Computação, Goiânia, 2024.

Bibliografia. Apêndice.

Inclui abreviaturas, gráfico, tabelas, algoritmos, lista de figuras, lista de tabelas.

1. Glaucoma. 2. Classificação. 3. Triagem. 4. Reflexo Pupilar à Luz. 5. Diagnóstico Auxiliado por Computador. I. da Costa, Ronaldo Martins, orient. II. Título.

CDU 004



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE INFORMÁTICA

ATA DE DEFESA DE TESE

Ata Nº 24/2024 da sessão de Defesa de Tese de **Hedenir Monteiro Pinheiro** que confere o título de Doutor em **Ciência da Computação**, na área de concentração em **Ciência da Computação**.

Aos vinte e dois dias do mês de agosto de dois mil e vinte e quatro, a partir das catorze horas e trinta minutos, na sala 257 do Instituto de Informática, realizou-se a sessão pública de Defesa de Tese intitulada “**Reflexo Pupilar à Luz como Biomarcador para Identificação de Glaucoma: Avaliação Comparativa de Redes Neurais e Métodos de Aprendizado de Máquina**”. Os trabalhos foram instalados pelo Orientador, Professor Doutor Ronaldo Martins da Costa (INF/UFV) com a participação dos demais membros da Banca Examinadora: Professora Doutora Mônica Mitiko Soares Matsumoto (PPGEEC/ITA), membra titular externa; Professor Doutor Eduardo Nery Rossi Camilo (UNIFAN), membro titular externo; Professor Doutor João Paulo Papa (PPGCC/Unesp), membro titular externo, cujas participações ocorreram através de videoconferência; Professor Doutor Rommel Melgaço Barbosa (INF/UFV), membro titular interno. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido o candidato aprovado pelos seus membros. Proclamados os resultados pelo Professor Doutor Ronaldo Martins da Costa, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos vinte e dois dias do mês de agosto de dois mil e vinte e quatro.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **João Paulo Papa, Usuário Externo**, em 22/08/2024, às 17:42, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Nery Rossi Camilo, Usuário Externo**, em 22/08/2024, às 18:08, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ronaldo Martins Da Costa, Professor do Magistério Superior**, em 22/08/2024, às 18:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Monica Mitiko Soares Matsumoto, Usuário Externo**, em 22/08/2024, às 18:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rommel Melgaco Barbosa, Professor do Magistério Superior**, em 22/08/2024, às 18:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Hedenir Monteiro Pinheiro, Discente**, em 22/08/2024, às 23:59, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4755182** e o código CRC **6BE76497**.

Referência: Processo nº 23070.034346/2024-69

SEI nº 4755182

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Hedenir Monteiro Pinheiro

Graduou-se em Ciência da Computação pela PUC-GO - Pontifícia Universidade Católica de Goiás em 1999. Especializou-se em Banco de Dados pela UFG - Universidade Federal de Goiás em 2003. Obteve em 2016 o título de Mestre no Programa de Pós-Graduação em Ciência da Computação também pela UFG. Durante o Doutorado, participou do grupo de pesquisa em Reconhecimento de Padrões e Redes Neurais, foi bolsista da FAPEG. Atualmente é Pesquisador Tecnologista no Instituto Nacional de Metrologia Qualidade e Tecnologia - INMETRO, onde desenvolve soluções para metrologia legal.

Dedico esse trabalho a Deus, o qual é digno de toda honra e toda glória. À minha mãe, a quem tanto amo e admiro. À minha esposa tão preciosa. A meus dois garotos que enchem minha vida de alegrias...

Agradecimentos

Agradeço a Deus pelo dom da vida, por me guiar, conduzir e fortalecer.

À Universidade Federal de Goiás, ao Instituto de Informática, especialmente, ao Programa de Pós-Graduação em Ciência da Computação, e a seu excepcional corpo docente e administrativo, sempre notavelmente atenciosos e prestativos.

Expresso também minha gratidão à Fundação de Amparo à Pesquisa (FAPEG) pelo suporte financeiro.

Ao meu professor, orientador e amigo, Dr. Ronaldo Martins da Costa, expresso um agradecimento especial. Foi uma honra trabalhar ao seu lado. Suas palavras sábias guiaram-me como um farol, e seu otimismo ofereceu-me o porto seguro necessário durante essa jornada.

Agradeço aos colegas do grupo de pesquisa pelo constante apoio e mútua colaboração.

À minha família, parte essencial da minha vida e deste trabalho. Em especial à minha querida mãe, Maria Botelho, pelo encorajamento, pelas orações e por estar sempre pronta a me apoiar.

À minha amada esposa, Thalita Fernanda, pelo amor, carinho, suporte e companheirismo. Aos meus filhos, André Rocha e Daniel Max, os quais são minha motivação.

A vocês, meu muito obrigado.

Não há nenhum segredo para o sucesso. É o resultado de preparação, trabalho duro, e aprender com o fracasso.

Colin L. Powell,

Resumo

Pinheiro, Hedenir. **Reflexo Pupilar à Luz como Biomarcador para Identificação de Glaucoma: Avaliação Comparativa de Redes Neurais e Métodos de Aprendizado de Máquina**. Goiânia, 2024. 134p. Tese de Doutorado. Programa de Pós-graduação em Ciência da Computação, Instituto de Informática (INF), Universidade Federal de Goiás (UFG).

O estudo das células ganglionares da retina, sua característica de fotossensibilidade e a relação com processos físicos e cognitivos tem impulsionado pesquisas com o reflexo pupilar. Controladas pelo Sistema Nervoso Autônomo (SNA), a dilatação (midríase) e contração (miose) são reflexos involuntários. Variações no diâmetro pupilar podem significar alterações físicas ou cognitivas de um indivíduo. Por este motivo, o reflexo pupilar tem sido considerado um importante biomarcador para diversos tipos de diagnósticos. Este estudo buscou aprimorar a identificação automatizada do glaucoma usando dados do reflexo pupilar à luz. Nele foi realizada uma análise comparativa entre redes neurais e técnicas clássicas para segmentar o sinal pupilar. Além disso, foi avaliado o desempenho de diversos métodos de processamento de dados, como técnicas de filtragem, extração de características, balanceamento de amostras e seleção de características, analisando seus efeitos no processo de classificação. Os resultados apresentam uma acurácia de 73,90% na classificação geral do glaucoma, 98,10% para a classificação do glaucoma moderado e 98,73% para o glaucoma severo, e oferece percepções e diretrizes para triagem e diagnóstico do glaucoma por meio do sinal proveniente da resposta pupilar à luz.

Palavras-chave

Glaucoma, Classificação, Triagem, Reflexo Pupilar à Luz, Diagnóstico Auxiliado por Computador, Aprendizado de Máquina

Abstract

Pinheiro, Hedenir. **Pupillary Light Reflex as a Biomarker for Glaucoma Identification: Comparative Evaluation of Neural Networks and Machine Learning Methods**. Goiânia, 2024. 134p. PhD. Thesis . Programa de Pós-graduação em Ciência da Computação, Instituto de Informática (INF), Universidade Federal de Goiás (UFG).

The study of retinal ganglion cells, their photosensitivity characteristics, and their relationship with physical and cognitive processes has driven research on the pupillary reflex. Controlled by the Autonomic Nervous System (ANS), dilation (mydriasis) and contraction (miosis) are involuntary reflexes. Variations in pupil diameter may indicate physical or cognitive changes in an individual. For this reason, the pupillary reflex has been considered an important biomarker for various types of diagnoses. This study aimed to improve the automated identification of glaucoma using data from the pupillary light reflex. A comparative analysis between neural networks and classical techniques was performed to segment the pupillary signal. In addition, the performance of various data processing methods was evaluated, including filtering techniques, feature extraction, sample balancing, and feature selection, analyzing their effects on the classification process. The results show an accuracy of 73.90% in the overall classification of glaucoma, 98.10% for moderate glaucoma classification, and 98.73% for severe glaucoma, providing insights and guidelines for glaucoma screening and diagnosis through the signal derived from the pupillary light response.

Keywords

Glaucoma, Classification, Screening, Pupillary Light Reflex, Computer-Aided Diagnosis, Machine Learning

Sumário

Lista de Figuras	17	
Lista de Tabelas	19	
1 Introdução	20	
1.1	Objetivos	22
1.2	Motivação	23
1.3	Justificativa	23
1.4	Organização do Trabalho	24
2 Fundamentos Teóricos	25	
2.1	Reflexo Pupilar à Luz	25
2.2	Pupilometria e Coleta de Dados	26
2.3	Gravação	26
2.4	Segmentação	26
2.5	Filtragem	27
2.5.1	Frequência Nyquist	27
2.6	Extração de Características	27
2.7	Classificação	27
2.8	Avaliação de Desempenho	28
2.8.1	Métricas de Desempenho	28
2.8.1.1	Acurácia (<i>Accuracy</i>)	29
2.8.1.2	Precisão (<i>Precision</i>)	29
2.8.1.3	Sensibilidade ou <i>Recall</i> (True Positive Rate - TPR)	29
2.8.1.4	Especificidade (True Negative Rate - TNR)	29
2.8.1.5	F1	29
2.8.1.6	Erro Médio Absoluto	30
2.8.1.7	Raiz do Erro Quadrático Médio	30
2.8.2	Matriz de Confusão (MC)	30
2.8.3	Curva ROC	31
2.8.4	Diagrama de Caixa (Boxplot)	31
2.8.5	Gráfico Violino	32
2.9	Validação Cruzada (<i>Cross Validation</i>)	32
2.10	Interseção sobre União e Precisão Média	33
2.11	Avaliação Estatística	35
2.11.1	Teste de Hipótese (TH)	35
2.11.2	Processo do Teste de Hipótese	35
2.11.3	Nível de Significância	36

2.11.4	P-Valor	36
2.11.5	Teste Shapiro-Wilk	36
2.11.6	Teste t de Student (T-Test)	37
2.11.7	Teste de Wilcoxon	37
2.12	Aprendizado de máquina (<i>Machine Learning</i>)	37
2.13	Tipos de modelos de detecção de objetos	37
2.14	Transferência de aprendizado (<i>Transfer Learning</i>)	38
2.15	<i>Backbone</i>	39
2.16	Considerações Finais do Capítulo	39
3	Revisão Sistemática da Literatura	40
3.1	Metodologia da Revisão	40
3.1.1	Planejamento	40
3.1.2	Seleção	41
3.2	Resultados da Revisão Sistemática	42
3.2.1	Dispositivos usados para gravação - Pupilômetro	42
3.2.2	Protocolo de Gravação do PLR	44
3.2.2.1	Luminosidade do Ambiente de Captura	46
3.2.2.2	Comprimento de Onda	46
3.2.2.3	Duração do Estímulo	46
3.2.2.4	Intensidade	46
3.2.2.5	Período de Adaptação	47
3.2.2.6	Acomodação entre Estímulos	47
3.2.2.7	Reflexo Direto ou Consensual	47
3.2.2.8	Quantidade de Repetições da Estimulação	48
3.2.3	Pesquisas em Algoritmos para Segmentação	48
3.2.4	Pesquisas que Realizaram Classificação	51
3.3	Discussão	53
3.4	Considerações Finais do Capítulo	54
4	Desenvolvimento de um Repositório de Imagens da Pupila, Associado com uma Avaliação de Técnicas de Segmentação Pupilar	55
4.1	Contextualização	55
4.2	Trabalhos Relacionados	56
4.3	Materiais e Métodos	58
4.3.1	Base de dados	58
4.3.1.1	Dispositivo de Gravação	58
4.3.1.2	Eyes Project (Alcoolemia) - Base de dados de pessoas que consumiram álcool	59
4.3.1.3	Eyes Project (Diabetes) - Conjunto de dados de pessoas com Diabetes Mellitus Tipo II	61
4.3.1.4	Eyes Project (Glaucoma) - Conjunto de Dados de Pessoas com Glaucoma	63
4.3.1.5	Anotação Manual (<i>ground truth</i>)	64
4.3.2	Avaliação de Algoritmos para Detecção e Medição Pupilar	65
4.3.2.1	Metodologia de Comparação entre Redes Neurais	65
4.3.2.2	Metodologia de Comparação entre Redes Neurais e Algoritmos Clássicos	68

4.4	Resultados	71
4.4.1	Resultados da Comparação entre Redes Neurais	71
4.4.2	Resultado da Comparação entre Redes Neurais e Algoritmos Clássicos	72
4.5	Discussão	74
4.6	Considerações Finais do Capítulo	75
5	Avaliação de Técnicas de Aprendizado de Máquina	76
5.1	Contextualização	76
5.2	Materiais e Métodos	78
5.2.1	A Base de Dados de Exames Pupilar	79
5.2.2	Protocolo de Gravação dos Vídeos	80
5.2.3	Dispositivo de Gravação	82
5.2.4	Segmentação e Pré-processamento	82
5.2.5	Filtragem	84
5.2.6	Extração de Características	84
5.2.7	Balanceamento	86
5.2.8	Seleção de Características	87
5.2.9	Fatiamento do Sinal	88
5.2.10	Classificadores	88
5.2.11	Classificação	90
5.2.12	Validação	90
5.2.13	Avaliação dos Resultados	91
5.2.14	Avaliação Estatística	91
5.2.15	Avaliação pela Melhor Acurácia	92
5.2.16	Métricas de Avaliação	92
5.3	Resultados	93
5.3.1	Resultado da Análise Estatística do Sinal Completo	93
5.3.1.1	Análise Estatística das Técnicas de Filtragem	93
5.3.1.2	Análise Estatística da Eficácia da Extração de Características	94
5.3.1.3	Análise Estatística das Técnicas de Balanceamento	96
5.3.1.4	Análise Estatística da Seleção de Características	98
5.3.2	Análise pela Melhor Acurácia	99
5.3.2.1	Avaliação do Sinal Completo pela Melhor Acurácia	99
5.3.2.2	Avaliação do Sinal após o Fatiamento pela Melhor Acurácia	100
5.3.3	Implementando a Técnica de Classificação Mais Promissora	101
5.4	Discussão	103
5.4.1	Sobre a Filtragem	104
5.4.2	Extração de Características	104
5.4.3	Sobre o Balanceamento de Amostras	104
5.4.4	Seleção de Características	104
5.4.5	Sobre o Fatiamento do Sinal	105
5.4.6	Análise dos Classificadores	105
5.4.7	Recomendações Práticas	106
5.4.8	Limitações do Estudo	106
5.4.9	Limitações da Avaliação do PLR	106
5.5	Considerações Finais do Capítulo	107

6	Conclusões	108
6.1	Acerca da Revisão da Literatura	108
6.2	Acerca da Segmentação	108
6.3	Acerca da Classificação	109
6.4	Trabalhos Futuros	110
	Referências Bibliográficas	111
A	Instruções para Download do Conjunto de Dados	128
B	Algoritmo Customizado para Filtragem do Sinal Pupilar	129
C	Arquitetura das Redes Neurais Usadas como Classificadores	130
D	Contribuições Científicas	133

Lista de Figuras

1.1	Número anual de pesquisas que utilizaram o reflexo pupilar à luz para auxiliar diagnósticos, conforme identificado na revisão sistemática da literatura realizada neste estudo.	21
2.1	Representação gráfica da MC.	31
2.2	Representação gráfica do cálculo da IoU, da curva Precisão-Sensibilidade e do cálculo da Precisão Média (Average Precision - AP) dos modelos de detecção. Imagem adaptada de [Adam et al. 2022].	34
3.1	O processo de revisão sistemática.	41
4.1	Distribuição de 2.000 imagens anotadas e 15.000 imagens não anotadas no conjunto de dados sobre álcool de acordo com gênero e nível de consumo de álcool.	60
4.2	Representação do protocolo de aquisição das imagens para a base álcool.	60
4.3	Imagens representativas de infravermelho (IR) amostradas do conjunto de dados. (a) Pupila dilatada gravada com o aparelho PG. (b) Pupila contraída gravada com o aparelho PG. (c) Pupila dilatada gravada com o aparelho SC. (d) Pupila contraída gravada com o aparelho SC.	61
4.4	Distribuição de 1.400 imagens anotadas e 32.000 imagens não anotadas entre pacientes com Diabetes Mellitus tipo II, categorizadas por gênero e olho gravado.	62
4.5	Representação do protocolo de aquisição para conjunto de dados de diabetes.	62
4.6	Distribuição das 6.600 imagens anotadas manualmente de pacientes com glaucoma, categorizadas por estágio da doença, gênero e olho registrado.	64
4.7	Distribuição das 211.790 imagens não anotadas de pacientes com glaucoma categorizadas por estágio da doença, gênero e olho registrado.	65
4.8	Exemplo da anotação da região da pupila usando o programa Labellmg.	66
4.9	Exemplo de pupila localizada por redes neurais convolucionais, sua caixa delimitadora e a elipse correspondente, calculada a partir das coordenadas da caixa.	67
4.10	Curva de aprendizado da YOLOv7 ao longo das épocas de treinamento.	68
4.11	Fluxograma do processo para comparar e identificar o melhor algoritmo de medição do diâmetro pupilar em imagens infravermelho do olho.	71
5.1	Diagrama da avaliação proposta dividida em três etapas principais: processamento do sinal pupilar, classificação dos dados e análise dos resultados.	79

5.2	Ilustração do protocolo estabelecido, abrangendo a fase inicial de adaptação pupilar, o espectro de cores, a duração da estimulação cromática e os intervalos especificados para adaptação entre cada estímulo.	81
5.3	Exemplo da curva do sinal pupilar segmentado, após a remoção dos períodos redundantes.	83
5.4	Fluxograma da aplicação das técnicas de aprendizado de máquina.	94
5.5	Gráfico de violino ilustrando a distribuição normalizada (por <i>escore-Z</i>) dos valores das 20 principais características do seletor de características LassoCV.	102
5.6	Mapa de calor das 20 principais características selecionadas pelo seletor de características LassoCV (a) e pelo seletor RFE (b).	102
5.7	Gráfico <i>boxplot</i> apresentando a acurácia de cada um dos classificadores avaliados, para o agrupamento Controle vs Patológico, usando a abordagem mais eficaz.	103
5.8	Matriz de Confusão (a) e Curva ROC (b) exibindo o desempenho do classificador LDA para o agrupamento Controle vs. Patológico usando a abordagem mais eficaz.	103
A.1	(A) Link para acesso à plataforma; (B) Tela de acesso para parceiros e não parceiros; (C) Ícone para download do conjunto de dados.	128
C.1	A arquitetura do classificador Rede Neural Totalmente Conectada (do inglês, FCN).	130
C.2	A arquitetura do classificador de Rede Neural Convolutacional Unidimensional (do inglês, 1D-CNN).	131
C.3	A arquitetura da primeira cabeça do classificador de Rede Neural Transformer (TRANS).	132

Lista de Tabelas

3.1	Pupilômetros usados nas pesquisas encontradas na literatura.	44
3.2	Protocolo de gravação e estimulação.	48
3.3	Algoritmos usados para detectar e segmentar a pupila.	51
3.4	Pesquisas que utilizaram técnicas envolvendo IA na classificação.	52
4.1	Detalhes das Imagens e dos Equipamentos de Filmagem.	59
4.2	Características dos modelos.	66
4.3	Desempenho de modelos baseados em redes neurais na detecção da região da pupila.	72
4.4	Comparação entre modelos baseados em redes neurais convolucionais e algoritmos clássicos de detecção pupilar aplicados em 2.000 imagens.	73
5.1	Distribuição dos exames disponíveis na base de dados de vídeos por grupo, gênero e idade dos voluntários.	80
5.2	Principais parâmetros usados pelos classificadores no processo de classificação do sinal pupilar.	91
5.3	Análise estatística das metodologias de filtragem destacando a abordagem mais eficaz ou identificando os equivalentes para cada agrupamento.	95
5.4	Análise estatística das metodologias de extração de características, destacando a abordagem mais eficaz ou identificando os equivalentes para cada agrupamento.	96
5.5	Avaliação estatística das técnicas de balanceamento por sobre amostragem e subamostragem, destacando a abordagem mais eficaz ou identificando os equivalentes para cada agrupamento.	97
5.6	Análise da técnica de seleção estatística de características destacando a abordagem mais eficaz para cada agrupamento. (Parte 1).	98
5.6	Análise da técnica de seleção estatística de características destacando a abordagem mais eficaz para cada agrupamento. (Parte 2).	99
5.7	Melhores acurácias alcançadas por classificadores agrupados por interesse e abordagem, utilizando o sinal completo.	100
5.8	Melhores acurácias de classificação alcançadas para cada agrupamento por interesse e abordagem, utilizando um segmento de sinal associado a uma cor de estimulação.	100

Introdução

Há mais de 50 anos o exame da pupila tem sido útil a profissionais de saúde na avaliação do estado físico e cognitivo dos pacientes. Inicialmente o método de avaliação consistia em se apontar uma lanterna luminosa para a pupila e verificar possível comprometimento de consciência [Levatin, Prasloski e Collen 1973, Teasdale e Jennett 1974]. Os avanços da tecnologia permitiram o desenvolvimento de dispositivos capazes de medir com exatidão o diâmetro pupilar, isto é, realizar o processo de pupilometria. Atualmente, estudos mostram que essas avaliações mais sofisticadas e precisas da reação pupilar à luz podem ser usadas para auxiliar especialistas a estabelecer diversos tipos de diagnósticos.

Diagnóstico é definido pelo NCI (National Cancer Institute) como o processo de identificação de uma doença, condição ou lesão a partir de seus sinais e sintomas. O histórico de saúde, testes físicos e exames, tais como: exames de sangue, de imagem e biópsias, podem ser usados para apoiar a elaboração de um diagnóstico [NCI 2022].

Um sistema computacional de auxílio à detecção e diagnóstico (Computer-Aided Detection and Diagnosis - CAD) é uma classe de sistemas computacionais cujo objetivo é auxiliar a encontrar e diagnosticar doenças e fornecer uma “segunda opinião” na interpretação de exames [Suzuki 2012]. O objetivo de sistemas CAD é melhorar a taxa de acerto dos especialistas e reduzir o tempo necessário para a interpretação dos exames [Firmino et al. 2016]. Eles utilizam métodos de aprendizado de máquina para analisar dados de imagem ou dados dos pacientes e avaliar sua condição, auxiliando os médicos no processo de tomada de decisão [Chan, Hadjiiski e Samala 2020]. Neste sentido, a pupilometria automatizada é um sistema CAD para imagens pupilares que pode auxiliar a diagnosticar patologias, condições fisiológicas, estado cognitivo ou emocional de uma pessoa. Também permite avaliar o interesse, o esforço na tomada de decisões [Bafna e Hansen 2021], cansaço [Hu e Lodewijks 2020], fadiga [Chen et al. 2022], uso de drogas [Amodio et al. 2018] e as funções do sistema autônomo [Giovannangeli et al. 2024].

A medições pupilares são sem contato, facilmente acessíveis e objetivas, exigindo apenas uma mínima cooperação do participante examinado [Kelbsch et al. 2019]. Por isso, a pupilometria tem ganhado destaque no meio científico, resultando em um cres-

cimento do número de estudos que investigam a aplicabilidade do Reflexo Pupilar à Luz (Pupillary Light Reflex - PLR) para auxiliar diagnósticos, como ilustrado a Figura 1.1. O PLR é a reação de contração e dilatação involuntária que ajusta o diâmetro da pupila conforme às variações na luminosidade, para regular a quantidade de luz que entra nos olhos [Joshi e Gold 2020]. A constrição da pupila, regulada pelo sistema parassimpático, é controlada pelo núcleo de Edinger-Westphal. Por outro lado, a ativação do músculo dilatador da pupila, sob influência do sistema simpático, é desencadeada por estímulos dos núcleos hipotalâmicos [Metzger e Emborg 2019].

Um biomarcador é uma característica objetivamente mensurável que serve como indicador de processos biológicos normais, patogênicos ou respostas farmacológicas a uma intervenção terapêutica [Puntmann 2009]. Estudos recentes têm destacado a resposta pupilar à luz como um biomarcador promissor, associando-a a aproximadamente 50 condições distintas ligadas ao PLR, segundo Pinheiro e colaboradores [Pinheiro e Costa 2021]. Esse crescente reconhecimento sublinha o potencial diagnóstico e terapêutico da medição da resposta pupilar em contextos clínicos [Arévalo-López et al. 2023].

A maioria dos estudos, contudo, têm se limitado a fazer análise estatística de características pupilométricas [Pattan et al. 2023, Najjar et al. 2023]. A constante evolução das técnicas de Inteligência Artificial (IA) e o crescente aumento nas bases de dados disponíveis para testes pupilométricos estão criando oportunidades para que ferramentas de IA aprimorem essas análises. Elas podem gerar percepções adicionais que podem auxiliar médicos e especialistas a diagnosticar patologias e condições fisiológicas com maior precisão e eficácia.

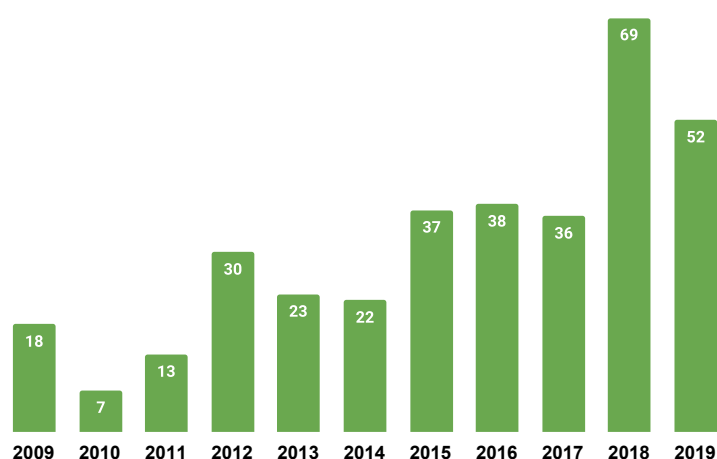


Figura 1.1: Número anual de pesquisas que utilizaram o reflexo pupilar à luz para auxiliar diagnósticos, conforme identificado na revisão sistemática da literatura realizada neste estudo.

Essa pesquisa se propõe a atuar nessa lacuna e explorar técnicas de aprendizado de máquina e redes neurais artificiais para aprimorar o diagnóstico médico através do reflexo pupilar. O estudo abrangeu várias etapas do exame pupilar, desde a captura de vídeos da pupila no ambiente clínico até a medição dos diâmetros pupilares, processamento do sinal incluindo filtragem e análise, balanceamento de amostras, seleção e extração de características, culminando na classificação e no suporte ao diagnóstico utilizando o sinal pupilar. Essa abordagem integrada busca otimizar o uso do reflexo pupilar como ferramenta diagnóstica, elevando sua precisão e eficácia.

1.1 Objetivos

Este trabalho tem por objetivo geral avaliar e adaptar técnicas computacionais e *frameworks* para apoiar o diagnóstico de doenças ou condições fisiológicas relacionadas às alterações no comportamento pupilar. A abordagem proposta baseia-se no uso de redes neurais artificiais convolucionais para segmentar a pupila em imagens de infravermelho dos olhos, extrair o sinal produzido pelas variações no diâmetro pupilar ao longo do tempo e, finalmente, classificar esse sinal. Essas técnicas serão integradas em uma solução de software denominada *Eyes Project*, cujo objetivo é proporcionar um ambiente robusto para o estudo do reflexo pupilar.

Os objetivos específicos deste trabalho são:

- Criar uma base de dados pública de imagens dos olhos, com e sem anotação da região pupilar, para fomentar o avanço das pesquisas relacionadas a alterações no reflexo pupilar à luz.
- Avaliar o desempenho de modelos computacionais na detecção da pupila e na medição precisa de seu diâmetro.
- Avaliar o impacto de técnicas de filtragem, balanceamento de amostras, seleção e extração de características na eficácia da classificação do glaucoma através do PLR.
- Avaliar e comparar modelos de inteligência artificial para a classificação do glaucoma a partir do sinal pupilar, extraído de vídeos de exames, identificando os classificadores mais precisos.
- Aprimorar a plataforma web de análise pupilar *Eyes Project*, integrando a ela as melhores técnicas e modelos avaliados.

1.2 Motivação

Diversos estudos destacam o comportamento da pupila frente à estimulação luminosa como um biomarcador eficaz para identificar sinais de doenças que impactam a visão, tais como glaucoma [Duque-Chica et al. 2018, Arévalo-López et al. 2023], Defeito Pupilar Aferente Relativo (DPAR) [Temel et al. 2019], ambliopia [Yetkin et al. 2019], Degeneração Macular Relacionada a Idade (DMRI) [Maynard, Zele e Feigl 2015]. Além disso, a pupila serve como indicativo de alterações neurológicas, incluindo aumento da pressão intracraniana [Chen et al. 2011], depressão [Bullock et al. 2019], Alzheimer [Park et al. 2019] e Parkinson [Bartosova et al. 2018]. A pupila também pode refletir estados fisiológicos influenciados pelo consumo de substâncias como drogas ou álcool [Rollins et al. 2014, Pinheiro et al. 2015].

Dada sua natureza não invasiva, praticidade e baixo custo, o exame pupilar apresenta vantagens em relação a outros métodos diagnósticos e tem o potencial de tornar-se um exame clínico importante. Embora, sua adoção na prática clínica ainda seja limitada [Bayraktar et al. 2023]. Dessa forma, este estudo busca aprimorar a pupilometria e avaliar ferramentas computacionais que auxiliem os profissionais de saúde e especialistas nas etapas do processo, desde a coleta e análise de imagens da pupila, passando pela extração e processamento do sinal pupilar, até a geração de indicativos diagnósticos.

1.3 Justificativa

A revisão sistemática desse trabalho e uma ampla gama de pesquisas destacam o potencial do Reflexo Pupilar à Luz (PLR) no auxílio ao diagnóstico de diversas doenças e condições fisiológicas. Apesar dessa evidência, sua implementação regular na prática clínica ainda é limitada. Existe demanda por estudos que aprofundem o entendimento do PLR, aumentem a precisão diagnóstica e tornem o exame mais prático, reduzindo o tempo necessário para a adaptação pupilar à escuridão, diminuindo a duração das filmagens e minimizando a dependência de equipamentos específicos, como o pupilômetro.

Neste contexto, esta pesquisa focou em explorar e demonstrar as técnicas mais eficazes para a medição do diâmetro pupilar, a filtragem, seleção, extração de características e classificação do sinal pupilar. O objetivo é melhorar significativamente o rastreamento e o diagnóstico de condições patológicas e fisiológicas através do PLR, tornando-o uma ferramenta diagnóstica mais eficiente e acessível.

1.4 Organização do Trabalho

Os demais capítulos deste estudo estão organizados da seguinte maneira:

Capítulo 2: Fundamentos Teóricos - Este capítulo fornece uma introdução aos conceitos fundamentais de pupilometria, discute métricas relevantes para a avaliação de algoritmos e aborda princípios das redes neurais artificiais.

Capítulo 3: Revisão Sistemática da Literatura - Nesta seção, foi realizada uma análise abrangente dos estudos anteriores que abordam a segmentação da pupila, a implementação de técnicas de redução de ruído e a aplicação de inteligência artificial na classificação pupilar. O objetivo é situar a pesquisa atual no contexto do campo de estudo, ressaltando tanto os avanços já realizados quanto as lacunas existentes.

Capítulo 4: Desenvolvimento do Banco de Dados de Imagens Anotadas e Metodologia de Avaliação - Este capítulo detalha a criação de um banco de dados contendo 10.000 imagens de pupilas anotadas manualmente e descreve a metodologia utilizada para conduzir um estudo comparativo destinado a avaliar o desempenho de diferentes algoritmos na localização e medição do diâmetro pupilar.

Capítulo 5: Avaliação de Técnicas de Aprendizado de Máquina - Aqui, as técnicas de aprendizado de máquina foram aplicadas e avaliadas com o intuito de explorar como elas podem aprimorar a precisão da classificação do sinal pupilar e auxiliar no diagnóstico de condições como o glaucoma.

Capítulo 6: Conclusões - A última seção apresenta uma discussão sobre os resultados alcançados pela pesquisa, oferece conclusões derivadas dos estudos realizados e propõe direções para futuros trabalhos na área.

Fundamentos Teóricos

Neste capítulo, será oferecida uma visão geral sobre conceitos relacionados ao (1) reflexo pupilar à luz. (2) As etapas necessárias para realização do diagnóstico a partir do sinal pupilar, quais sejam: utilização de um equipamento de filmagem (pupilômetro) para captação das imagens, segmentação da pupila, filtragem do sinal pupilar, extração de características, seleção de características e classificação. Apresenta conceitos acerca das (3) métricas de avaliação de desempenho e comparação entre os modelos. Por fim, descreve tópicos acerca das (4) Redes Neurais Artificiais.

Esta seção busca estabelecer uma base de conhecimento que suportará os desenvolvimentos e os resultados discutidos nos capítulos subsequentes do documento.

2.1 Reflexo Pupilar à Luz

O Reflexo Pupilar à Luz é uma reação involuntária da pupila frente a variação de luminosidade, essa resposta é promovida pela interação entre o sistema nervoso simpático e parassimpático que compõem o sistema nervoso autônomo [Purves et al. 2008].

As células bastonetes, cones e as células ganglionares retiniais intrinsecamente foto sensíveis (intrinsically photosensitive retinal ganglion cells - ipRGCs) são células fotorreceptoras que induzem o PLR. As células bastonetes são responsáveis pela visão no escuro (visão escotópica). As células cones são sensíveis a cor e responsáveis pela visão em ambiente claro (visão fotópica). Ambas estão envolvidas na visão em luz intermediária (visão mesópica) [Asakawa et al. 2019].

As células ipRGC recebem os estímulos das células bastonetes e cones, e também possuem fotopigmentos (melanopsina). As células bastonetes e cones são responsáveis pela constrição inicial relativa ao PLR (0,2 a 1,5 segundos) e as células ipRGCs são responsáveis por manter a constrição pupilar, e também são responsáveis pela resposta após a estimulação (post-illumination pupil response - PIPR) [Adhikari, Zele e Feigl 2015], que corresponde ao retorno gradual da pupila ao diâmetro inicial. Por fim, essas células também são responsáveis pela regulação do ritmo circadiano

[Mathôt 2018] que é o comportamento fisiológico, mental e comportamental a cada ciclo de 24 horas [NIH 2022].

2.2 Pupilometria e Coleta de Dados

Pupilometria é a medição do diâmetro pupilar e pode ser realizada de maneira estática ou dinâmica. A pupilometria estática consiste na medição pontual do diâmetro pupilar em um determinado momento, é normalmente feita administrando-se substâncias diretamente sobre a pupila dos pacientes para estimular sua constrição ou dilatação. Esse método, porém, está sujeito à permeabilidade da córnea de cada indivíduo [Bremner 2009].

A pupilometria dinâmica é conduzida através da medição contínua do diâmetro pupilar ao longo de um determinado período, sob condições de luminosidade variadas: baixa (escotópica), média (mesópica) ou alta (fotópica).

2.3 Gravação

A medição manual (visual) do comportamento pupilar tem se mostrado imprecisa e inadequada para análises que vão além da verificação de estado de coma ou se a pessoa está ou não viva. Desde a década de cinquenta, equipamentos tem sido desenvolvidos e aprimorados para medir com precisão o diâmetro pupilar [Lowenstein e Loewenfeld 1958]. Nessa fase é realizada a produção de um ou mais vídeos que registram o comportamento da pupila dos pacientes. A gravação é realizada com um equipamento conhecido como pupilômetro, o qual deve permitir a medição do diâmetro pupilar frente a estímulos com duração, cor e intensidade adequados ao tipo de condição ou patologia que se deseja examinar.

2.4 Segmentação

Fazer a segmentação em uma imagem significa classificar cada um de seus pixels como pertencente ou não à região de interesse (Region of Interest - ROI). Essa classificação permite localizar, extrair, destacar ou medir a área da imagem considerada relevante. Nessa pesquisa foram usados algoritmos clássicos como PuRe [Santini, Fuhl e Kasneci 2018] e o PuReST [Santini, Fuhl e Kasneci 2018] e as redes neurais Faster R-CNN [Ren et al. 2015] e a rede YOLO nas versões 4, 5 e 7 [Bochkovskiy, Wang e Liao 2020, Jocher et al. 2022, Wang, Bochkovskiy e Liao 2023] para realizar a segmentação das imagens e medir o diâmetro pupilar.

2.5 Filtragem

A filtragem do sinal pupilar consiste em se remover ruídos causados por piscadas, desvios no olhar, perdas na detecção da pupila e outros. Um exemplo de filtro de sinal é o filtro passa-baixa (low-pass) o qual permite que sinais com frequências mais baixas passem através dele, enquanto atenua (reduz) as frequências mais altas [Proakis 2007].

2.5.1 Frequência Nyquist

A frequência de Nyquist é um conceito da teoria de amostragem de sinais, que define a frequência mínima necessária para capturar todas as informações de um sinal contínuo sem perder dados ou introduzir distorções. Ela é igual à metade da taxa de amostragem de um sistema. Isso significa que, para uma taxa de amostragem f , a frequência de Nyquist é $\frac{f}{2}$ [Por, Kooten e Sarkovic 2019].

2.6 Extração de Características

É o processo de identificar e isolar informações relevantes de um conjunto de dados brutos, a fim de facilitar sua análise ou classificação [Khalid, Khalil e Nasreen 2014]. No contexto da pupilometria, é a fase onde se sintetiza regiões do sinal pupilar em atributos que descrevem comportamentos relevantes do exame. São exemplos de características extraídas do reflexo pupilar: diâmetro inicial, máxima contração (menor diâmetro), amplitude de contração (diferença entre o diâmetro inicial e a máxima contração), tempo para máxima contração, velocidade de contração, velocidade de re-dilatação e outras.

2.7 Classificação

É a fase onde se define a classe a qual uma entrada pertence. O aprendizado supervisionado é uma técnica de *Machine Learning* onde um algoritmo é treinado a partir de um conjunto de dados rotulados, ou seja, entradas cujas saídas (classes) corretas já são conhecidas. Esse processo envolve fornecer ao modelo um conjunto de pares de entrada e saída, permitindo que ele aprenda a mapear as características das entradas para as respectivas classes. A fase de treinamento é crucial, pois é durante esse estágio que o classificador, uma função matemática ou modelo de aprendizado, ajusta seus parâmetros internos para minimizar os erros na classificação das entradas.

No contexto desta pesquisa, o sinal pupilar pode pertencer a duas classes distintas: a classe normal (controle), representando indivíduos sem alterações patológicas ou

fisiológicas, ou a classe alvo, que engloba sinais associados a alguma patologia ou condição fisiológica alterada. O processo de classificação começa com a criação de um modelo supervisionado, onde o classificador é treinado usando um conjunto de dados rotulados, contendo sinais pupilares já classificados nas categorias normais ou alteradas. Durante essa fase, o modelo analisa padrões nos dados de entrada que o ajudem a identificar corretamente a classe de cada novo sinal apresentado.

Após o treinamento, o classificador é então utilizado para avaliar novas entradas, ou seja, sinais pupilares de indivíduos cujas condições fisiológicas ainda não foram identificadas. O classificador atribui a cada nova entrada uma classe com base no que aprendeu durante o treinamento, desempenhando o papel de rotular esses sinais conforme os padrões previamente observados.

Os classificadores empregados nesta pesquisa incluíram tanto métodos tradicionais quanto modelos baseados em redes neurais. Os classificadores tradicionais, como os algoritmos de regressão logística e máquinas de vetores de suporte (SVM), são amplamente utilizados devido à sua simplicidade e eficácia em diversos tipos de dados. Já os classificadores baseados em redes neurais, como as redes neurais totalmente conectadas (Fully Connected Neural Networks - FCN), destacam-se por sua capacidade de modelar relações complexas entre as variáveis e aprender automaticamente padrões a partir de grandes volumes de dados, o que os torna especialmente úteis para tarefas mais desafiadoras de classificação.

2.8 Avaliação de Desempenho

A avaliação do desempenho de um modelo de classificação requer que, após treinado, o modelo classifique amostras cujo rótulo de classe ele não conheça. Estas amostras são chamadas de amostras de teste. Elas devem possuir rótulos de classe conhecidos, mas que serão omitidos do modelo de classificação. O valor predito é comparado com o valor real conhecido, e a quantidade de erros e acertos do modelo de classificação é observada.

2.8.1 Métricas de Desempenho

Diversas métricas são empregadas para avaliar o desempenho de modelos de classificação e oferecer uma visão quantitativa sobre a eficiência do modelo em classificar as amostras. As categorizações de acertos e erros em modelos são identificadas como:

- Verdadeiros Positivos (True Positives - TP): O número de instâncias positivas corretamente classificadas pelo modelo.

- Falsos Negativos (False Negatives - FN): O número de instâncias positivas incorretamente classificadas como negativas.
- Falsos Positivos (False Positives - FP): O número de instâncias negativas incorretamente classificadas como positivas.
- Verdadeiros Negativos (True Negatives - TN): O número de instâncias negativas corretamente classificadas pelo modelo.

Os valores TP, TN, FP e FN são usados para calcular métricas de desempenho, tais como Acurácia, Sensibilidade e Especificidade e F1 [Fawcett 2006], as quais são calculadas conforme as seguintes equações:

2.8.1.1 Acurácia (*Accuracy*)

É a proporção de todas as previsões corretas (tanto positivas quanto negativas) em relação ao total de instâncias.

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN}$$

2.8.1.2 Precisão (*Precision*)

Traz a proporção de instâncias preditas como positivas que são realmente positivas.

$$\text{Precisão} = \frac{TP}{TP + FP}$$

2.8.1.3 Sensibilidade ou *Recall* (True Positive Rate - TPR)

É a proporção de instâncias positivas corretamente classificadas.

$$\text{Sensibilidade} = \frac{TP}{TP + FN}$$

2.8.1.4 Especificidade (True Negative Rate - TNR)

É a proporção de instâncias negativas corretamente classificadas.

$$\text{Especificidade} = \frac{TN}{TN + FP}$$

2.8.1.5 F1

A métrica F1 demonstra o equilíbrio entre a precisão e a sensibilidade, fornecendo uma visão holística do desempenho do modelo. Ela calcula o harmônico médio de

precisão e sensibilidade. Diferentemente da média aritmética, o médio harmônico penaliza extremos, assim a métrica F1 será baixa se qualquer uma das métricas (precisão ou sensibilidade) for baixa. A fórmula da métrica F1 é:

$$F1 = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$$

2.8.1.6 Erro Médio Absoluto

O Erro Médio Absoluto (Mean Absolute Error - MAE) é uma métrica que mede a média das diferenças absolutas entre os valores previstos e os valores reais. Ele é calculado como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2-1)$$

Onde y_i e \hat{y}_i são os valores previstos e n é o número total de observações. O MAE fornece uma medida direta da precisão das previsões, sem penalizar severamente grandes erros.

2.8.1.7 Raiz do Erro Quadrático Médio

A Raiz do Erro Quadrático Médio (Root Mean Square Error - RMSE) é uma métrica que mede a raiz quadrada da média dos erros ao quadrado entre os valores previstos e os valores reais. Sua fórmula é:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2-2)$$

O RMSE penaliza mais severamente os grandes erros, tornando-se útil para avaliar a precisão do modelo quando grandes discrepâncias são indesejadas. Ele fornece uma medida da magnitude média dos erros em unidades equivalentes às dos dados originais.

2.8.2 Matriz de Confusão (MC)

A matriz de confusão é uma ferramenta que permite visualizar o desempenho do algoritmo comparando os valores preditos com os valores reais [2.1](#).

Para um problema de classificação binária (duas classes: Positiva e Negativa), a matriz de confusão é uma tabela 2x2 com as seguintes entradas:

	Predição Positiva	Predição Negativa
Real Positiva	TP	FN
Real Negativa	FP	TN

Figura 2.1: Representação gráfica da MC.

2.8.3 Curva ROC

A curva ROC (Receiver Operating Characteristic) é uma ferramenta gráfica usada para avaliar o desempenho de um modelo de classificação binária, especialmente quando se deseja compreender a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (1 - especificidade) para diferentes limiares de decisão [Youden 1950]. A curva produzida pode ser mapeada em um valor escalar que mede a área abaixo da curva (AUC) e que representa o desempenho do classificador, como na Figura 5.8(b).

Quanto à interpretação da curva ROC, cada ponto na curva representa um par (Taxa de Falsos Positivos - FPR, Taxa de Verdadeiros Positivos - TPR). Uma curva ROC que se aproxima do canto superior esquerdo do gráfico indica um melhor desempenho do modelo, pois maximiza a sensibilidade enquanto minimiza a taxa de falsos positivos. Uma linha diagonal (de (0,0) a (1,1)) representa um classificador cujo desempenho não é melhor do que uma escolha aleatória.

A AUC é uma métrica que quantifica a área total sob a curva ROC. Ela varia de 0 a 1. Um AUC de 1 indica um modelo perfeito, enquanto um AUC maior que 0,5 e menor que 1 indica que o modelo tem algum poder discriminativo.

2.8.4 Diagrama de Caixa (Boxplot)

O diagrama de caixa é uma ferramenta gráfica usada para exibir a distribuição de um conjunto de dados baseado em cinco estatísticas: mínimo, primeiro quartil (Q1), mediana (Q2), terceiro quartil (Q3) e máximo. Ele é especialmente útil para identificar a dispersão dos dados, a presença de *outliers* e a simetria da distribuição, como na Figura 5.7.

O boxplot é composto pela caixa, os *whiskers* e os *outliers*. A caixa representa o intervalo interquartil (Interquartile Range - IQR), que é a diferença entre o terceiro quartil (Q3) e o primeiro quartil (Q1). A parte inferior da caixa é o primeiro quartil (Q1),

abaixo do qual 25% dos dados se encontram, e a parte superior é o terceiro quartil (Q3), abaixo do qual 75% dos dados se encontram. A linha no interior da caixa representa a mediana (Q2), que é o valor central dos dados. Os *whiskers* (fios de bigode) são linhas que se estendem a partir da caixa até 1,5 vezes o IQR a partir dos quartis, mostrando a extensão dos dados. Os valores que caem além dos *whiskers* são considerados *outliers* e são plotados individualmente como pontos.

A posição da mediana dentro da caixa fornece informações sobre a simetria da distribuição dos dados. Se a mediana está centralizada, a distribuição é aproximadamente simétrica; se está deslocada, indica assimetria. O tamanho da caixa (IQR) indica a variabilidade dos dados: uma caixa grande sugere alta variabilidade, enquanto uma caixa pequena indica baixa variabilidade. *Whiskers* longos podem indicar a presença de dados dispersos, quando curtos mostram que os dados estão mais concentrados próximos à mediana. *Outliers* identificados fora dos *whiskers* podem indicar variabilidade incomum ou erros nos dados.

2.8.5 Gráfico Violino

O gráfico violino é uma ferramenta gráfica utilizada para visualizar a distribuição de um conjunto de dados. Ele combina aspectos de um diagrama de caixa e uma estimativa de densidade, proporcionando uma visão mais detalhada da densidade dos dados ao longo de todo o seu intervalo. Essa representação permite que se observe a concentração dos dados, como na Figura 5.5.

Nesse gráfico em forma de violino, a linha central indica a mediana; as linhas horizontais marcam os quartis (primeiro e terceiro). A largura do violino em qualquer ponto indica a densidade relativa dos dados naquele ponto, sendo mais largo onde os dados são mais concentrados. Em algumas implementações, os *outliers* podem ser mostrados como pontos individuais fora da forma do violino.

A simetria da forma ao redor da linha central sugere a simetria da distribuição dos dados. Gráficos violinos são úteis para comparar distribuições de diferentes conjuntos de dados, permitindo uma visualização das diferenças em termos de dispersão e densidade.

2.9 Validação Cruzada (*Cross Validation*)

A validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados [Kohavi et al. 1995]. O conceito central da validação cruzada é particionar o conjunto de dados em k subconjuntos mutuamente exclusivos, e usar $k-1$ conjuntos para treinamento e o conjunto restante conjunto para testar o modelo.

Além da validação cruzada, duas outras técnicas podem ser usadas para particionar os dados. O *hold out*, que consiste em se dividir os dados em um conjunto para treino e outro para teste, frequentemente na proporção 80%/20% (80% para treino e 20% para teste). E o *leave-one-out*, que é um caso particular da validação cruzada, em que cada subconjunto possui apenas uma amostra. É usado quando o conjunto de dados é pequeno.

2.10 Interseção sobre União e Precisão Média

O índice de Jaccard (J) ou coeficiente de similaridade de Jaccard foi desenvolvido por Paul Jaccard [Jaccard 1901] em 1901. Ele é usado para medir a similaridade entre dois conjuntos de amostras (conjunto A e conjunto B), como definido na Equação 2-3.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2-3)$$

No contexto de detecção de objetos as posições cartesianas x, y , a largura da caixa w e a altura h definem uma caixa delimitadora, a Interseção sobre União (*Intersection Over Union* - IoU) permite verificar a sobreposição das duas caixas delimitadoras excedem um determinado limiar (ex. 0,75) e, portanto, seja considerada como uma detecção correta, conforme ilustrado na Figura 2.2.

$$IoU(A, B) = \frac{Area |A \cap B|}{Area |A \cup B|} \quad (2-4)$$

Sejam $A = [x, y, w, h]$ e $B = [\hat{x}, \hat{y}, \hat{w}, \hat{h}]$ as áreas da caixa delimitadora do objeto real e do objeto previsto, respectivamente, a IoU é a razão entre a área da intersecção da caixa A com a caixa B, pela união das áreas das duas caixas. Quanto maior a sobreposição entre as caixas, maior será o valor da IoU. Essa métrica varia entre 0 a 1, sendo zero quando as caixas não possuem nenhuma sobreposição e um quando as caixas possuem a máxima sobreposição, ou seja, têm a mesma localização e dimensões. Limiares de 0,5 e 0,75 costumam ser usados quando se avalia uma predição de detecção de objeto como correta ou errada.

As previsões em que a IoU supera determinado limiar são consideradas corretas. As que ficam abaixo desse limiar são consideradas incorretas. Mede-se a acurácia do detector de objetos pela quantidade de:

- **Verdadeiros Positivos (True Positive - TP):** Quando a previsão de um objeto na imagem coincide com o objeto real e supera o limiar de IoU estabelecido.
- **Falsos Negativos (False Negative - FN):** Quando o objeto está presente na imagem, mas não é detectado ou a detecção não é suficientemente precisa para superar o limiar de IoU.

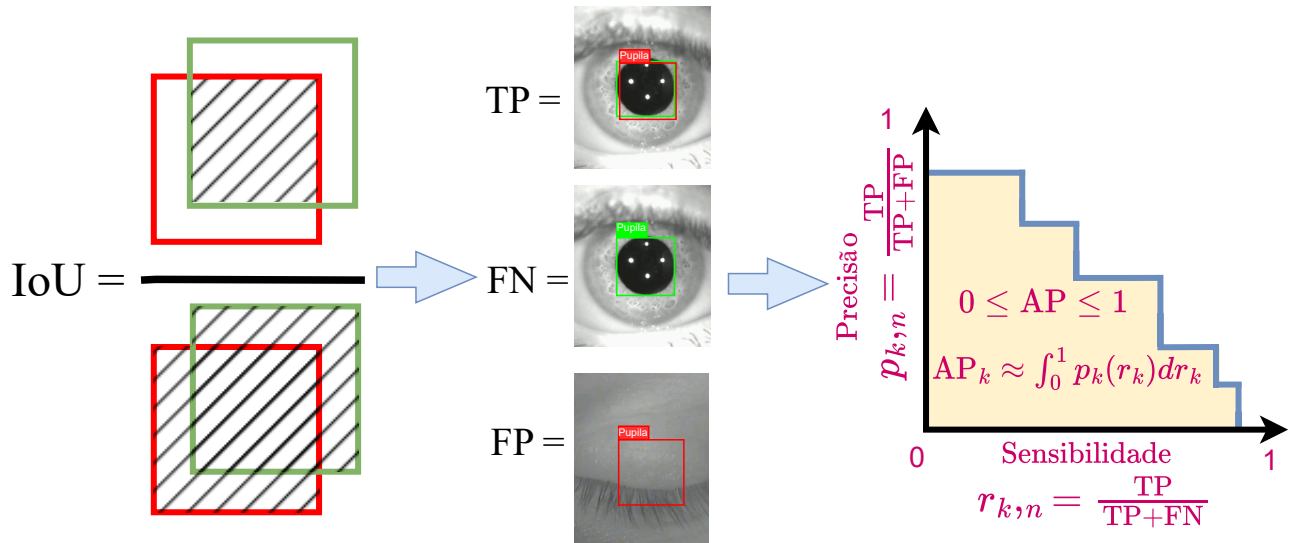


Figura 2.2: Representação gráfica do cálculo da IoU, da curva Precisão-Sensibilidade e do cálculo da Precisão Média (Average Precision - AP) dos modelos de detecção. Imagem adaptada de [Adam et al. 2022].

- **Falsos Positivos (False Positive - FP):** Quando um objeto é detectado, mas não há nenhum objeto real correspondente na imagem.

Tanto as classificações corretas quanto as incorretas são usadas para calcular a Precisão Média, que fornece uma métrica de desempenho para comparação de modelos de detecção de objetos [Padilla, Netto e Silva 2020]. Essa métrica, adotada na maioria das competições entre modelos [Padilla, Netto e Silva 2020, Everingham et al. 2010], combina as avaliações de precisão e sensibilidade, oferecendo uma medida única que resume o desempenho geral do modelo, calculada pela área sob a curva precisão-sensibilidade (*recall-precision*). A precisão média será a área sob a curva para as n imagens $\in \{1 \dots N\}$ da classe $k \in \{1 \dots K\}$

$$AP_k \approx \int_0^1 p_k(r_k) dr_k \quad (2-5)$$

A mAP é a média da AP de todas as classes de objetos, ela oferece uma métrica para o desempenho geral do modelo em tarefas de detecção de objetos, como visto na Equação 2-6, onde AP_i é a precisão média para a classe i e K o número total de classes. Neste estudo, AP e mAP são equivalentes, pois a pupila é a única categoria de objeto de interesse nas imagens usadas para detecção.

$$mAP = \frac{1}{K} \sum_{i=1}^K AP_i \quad (2-6)$$

2.11 Avaliação Estatística

Neste trabalho, foi realizada análise estatística a partir dos resultados de cada um dos classificadores para determinar para cada agrupamento, se os resultados usando uma abordagem específica são estatisticamente superiores, inferiores ou equivalentes às outras abordagens, por meio do teste de hipótese.

2.11.1 Teste de Hipótese (TH)

O teste de hipótese é uma técnica estatística que permite tomar decisões com base em dados amostrais. Ele envolve formular duas hipóteses opostas: a hipótese nula (H_0) e a hipótese alternativa (H_1) [Montgomery e Runger 2010].

Hipótese Nula (H_0): É a suposição inicial que não há efeito ou diferença. Por exemplo, “não há diferença significativa entre as abordagens de classificação”.

Hipótese Alternativa (H_1): É a suposição contrária que há um efeito ou diferença. Por exemplo, “há uma diferença significativa entre as abordagens de classificação”.

Fisher chamou os testes de hipóteses de “testes de significância” e sugeriu a probabilidade de 0,05 como o nível de corte para rejeitar a hipótese nula. Ele aplicou essa probabilidade a uma distribuição normal utilizando um teste bicaudal. “Estes testes podem ser chamados testes de significância, e quando estes testes estão disponíveis é possível descobrir se uma segunda amostra é ou não é significativamente diferente da primeira [amostra].” [Fisher 1944]

2.11.2 Processo do Teste de Hipótese

O processo de construção do TH pode ser exemplificado nos seguintes passos: primeiramente, definem-se as hipóteses H_0 e H_1 . Em seguida, escolhe-se o nível de significância (α), que determina a probabilidade de rejeitar a hipótese nula quando ela é verdadeira, comumente sendo $\alpha = 0,05$ (5%). Utilizam-se os dados amostrais para calcular uma estatística de teste que segue uma distribuição conhecida (como a distribuição t ou a distribuição normal). Calcula-se então o p-valor, o qual é a probabilidade de se obter o valor observado, assumindo que a hipótese nula é verdadeira. Por fim, compara-se o p-valor com o nível de significância: se o p-valor $\leq \alpha$, rejeita-se H_0 ; caso contrário, se o p-valor $> \alpha$, não se rejeita H_0 .

2.11.3 Nível de Significância

O nível de significância, denotado por α , é o limiar usado para decidir rejeitar ou não a hipótese nula [Lehmann, Romano e Casella 1986]. É a probabilidade de cometer um erro do tipo I, ou seja, rejeitar H_0 quando ela é verdadeira. Os valores comuns de α são 0,05 (5%), 0,01 (1%) e 0,10 (10%).

Quando $\alpha = 0,05$ há uma chance de 5% de rejeitar a hipótese nula erroneamente. Quando $\alpha = 0,01$ há uma chance de 1% de rejeitar a hipótese nula erroneamente. Quando $\alpha = 0,10$ há uma chance de 10% de rejeitar a hipótese nula erroneamente.

2.11.4 P-Valor

O p-valor é uma medida que auxilia na decisão sobre a hipótese nula [Wasserstein e Lazar 2016]. Ele representa a probabilidade de observar os dados que se tem, assumindo que a hipótese nula seja verdadeira. Um p-valor baixo indica que os dados observados são improváveis sob a hipótese nula, sugerindo que H_0 pode não ser verdadeira.

Se o P-valor $\leq \alpha$ rejeita-se a hipótese nula, pois há evidências suficientes contra H_0 . Se o P-valor $> \alpha$ não se rejeita a hipótese nula, pois não há evidências suficientes contra H_0 . Exemplificação com $\alpha = 0,05$ Se tendo sido definido $\alpha = 0,05$ e tendo sido obtido um p-valor de 0,03, rejeita-se a hipótese nula porque $0,03 < 0,05$. Isso sugere que os resultados observados são estatisticamente significativos e não são devidos ao acaso.

2.11.5 Teste Shapiro-Wilk

O teste Shapiro-Wilk é um teste de normalidade usado para verificar se uma amostra segue uma distribuição normal [Shapiro e Wilk 1965]. Nesse teste, inicialmente, formulam-se as hipóteses, onde a hipótese nula (H_0) afirma que os dados seguem uma distribuição normal e a hipótese alternativa (H_1) afirma que os dados não seguem uma distribuição normal. O teste de Shapiro-Wilk calcula uma estatística W, que mede quão bem os dados se ajustam a uma distribuição normal. O p-valor é então calculado a partir dessa estatística W. Um p-valor baixo (geralmente $\leq 0,05$) indica que se rejeita a hipótese nula, ou seja, os dados não seguem uma distribuição normal. A decisão com base no teste de Shapiro-Wilk, nesse trabalho, foi a seguinte: se os dados forem normais ($p > 0,05$), usa-se o teste t de Student; se os dados não forem normais ($p \leq 0,05$), usa-se o teste de Wilcoxon.

2.11.6 Teste t de Student (T-Test)

O teste t de Student é usado para comparar as médias de dois grupos quando os dados são normalmente distribuídos [Student 1908]. Inicialmente, formulam-se as hipóteses, onde a hipótese nula (H_0) afirma que não há diferença significativa entre as médias dos dois grupos, e a hipótese alternativa (H_1) afirma que há uma diferença significativa entre as médias dos dois grupos. A estatística t é calculada com base nas médias dos grupos, variâncias e tamanho das amostras. O p-valor é calculado a partir da distribuição t. Se o p-valor for menor ou igual ao nível de significância, rejeita-se a hipótese nula.

2.11.7 Teste de Wilcoxon

O teste de Wilcoxon é um teste não paramétrico usado para comparar dois grupos quando os dados não são normalmente distribuídos [Wilcoxon 1992, Conover 1999]. Inicialmente, formulam-se as hipóteses, onde a hipótese nula (H_0) afirma que não há diferença significativa entre as distribuições dos dois grupos, e a hipótese alternativa (H_1) afirma que há uma diferença significativa entre as distribuições dos dois grupos. A estatística de Wilcoxon é calculada com base nas posições dos dados em vez dos valores em si. O p-valor é calculado a partir da distribuição da estatística de Wilcoxon. Se o p-valor for menor ou igual ao nível de significância, rejeita-se a hipótese nula.

2.12 Aprendizado de máquina (*Machine Learning*)

Área do conhecimento que estuda como os sistemas de computador podem aproveitar a experiência (geralmente dados) para melhorar seu desempenho em tarefas específicas. Ele combina ideias de estatísticas, mineração de dados e otimização. Muitas vezes, é usado como meio de implementação de soluções de inteligência artificial [Zhang et al. 2021].

2.13 Tipos de modelos de detecção de objetos

Os modelos de detecção de objetos baseados em redes neurais podem ser classificados em dois tipos: modelos de etapa única e de duas etapas [Zhao et al. 2019]. Os detectores de duas etapas inicialmente identificam regiões de interesse na imagem e, em seguida, classificam e localizam objetos dentro dessas regiões. Em contraste, os modelos de etapa única detectam objetos diretamente, sem a necessidade de uma etapa preliminar de proposição de regiões.

1. Abordagens de duas etapas (*two-stage*) - Nestes modelos, a primeira etapa envolve a proposição de regiões de interesse, seguida por uma segunda etapa que classifica essas regiões e ajusta os limites dos objetos (*bounding boxes*). Algumas das principais redes incluem RCNN [Girshick et al. 2014], VGG net (2014) [Simonyan e Zisserman 2014], Fast R-CNN [Girshick 2015], Faster R-CNN (2015) [Ren et al. 2015], Mask R-CNN (2017) [He et al. 2017], Redes Piramidais/FPN (2017) [Lin et al. 2017], G-RCNN (2021) [Pramanik et al. 2021].
2. Abordagens de etapa única (*one-stage*) - Neles, a detecção e classificação de objetos ocorrem simultaneamente em uma única passagem pela rede. Eles processam a imagem inteira dividida em uma grade e fazem previsões de localização dos objetos e probabilidades de classe para cada célula da grade. Entre as principais estão YOLO (2016) [Redmon et al. 2016], SSD (2016) [Liu et al. 2016], RetinaNet (2017) [Lin et al. 2017], YOLOv3 (2018) [Redmon e Farhadi 2018], YOLOv4 (2020) [Bochkovskiy, Wang e Liao 2020], YOLOv5 (2020) [Jocher et al. 2022], e YOLOv7 (2022) [Wang, Bochkovskiy e Liao 2023].

2.14 Transferência de aprendizado (*Transfer Learning*)

No contexto de redes neurais profundas, a transferência de aprendizado é uma técnica que aproveita as características previamente aprendidas por uma rede em um determinado problema para solucionar outro problema, similar ao primeiro. Essa abordagem aprimora o desempenho do modelo em um domínio ao transferir informações de um domínio relacionado [Weiss, Khoshgoftaar e Wang 2016].

Uma das vantagens da transferência de aprendizado é reduzir o tempo computacional necessário para treinar a rede, pois, ao invés de se treinar o novo modelo do início, faz-se uso da informação já aprendida em um modelo pré-treinado. Outra vantagem é estender o conhecimento já adquirido do modelo anterior, e por fim melhorar o desempenho da rede quando o conjunto de treinamento é pequeno. Em outras palavras, treinar os pesos de uma rede neural profunda pode requerer bastante de tempo e exigir abundância de dados. Esses requisitos poderiam tornar inviável o uso dessas redes em imagens médicas, por exemplo, onde o número de imagens pode ser limitado. Entretanto, graças à transferência de aprendizado esse problema pode ser contornado.

2.15 Backbone

O *backbone* no contexto de redes neurais convolucionais (CNN) refere-se à parte da rede pré-treinada responsável pela extração de características da imagem. É quem processa a imagem de entrada para produzir um conjunto de características então utilizadas pelas partes subsequentes do modelo para a detecção de objetos. Existem muitas arquiteturas de CNN populares que podemos ser usadas como backbone em redes neurais. Algumas delas são: as VGGs (Visual Geometry Group)[[Simonyan e Zisserman 2014](#)] que incluem as redes convolucionais VGG-16 e VGG-19 com 16 e 19 camadas; a Inception (GoogleNet) [[Szegedy et al. 2015](#)]; e as ResNets (Redes Neurais Residuais) [[He et al. 2016](#)] que incluem as ResNet-50 e ResNet-101.

A definição do *backbone* impacta na precisão da detecção de objetos e no equilíbrio entre precisão e eficiência computacional. Redes mais profundas e complexas podem oferecer melhor precisão, mas também exigem mais recursos computacionais.

2.16 Considerações Finais do Capítulo

Esse capítulo apresentou, conceito sobre o reflexo pupilar e definições sobre as tarefas necessárias para realização da classificação dos exames, tais como: gravação, segmentação, filtragem, seleção de características, e classificação. Traz também uma discussão sobre as principais métricas usadas na avaliação de desempenho e conceitos sobre redes neurais, *backbone*, transferência de aprendizado e modelos de detecção de um e dois estágios.

Revisão Sistemática da Literatura

Este capítulo oferece uma revisão detalhada de estudos que se propuseram a realizar filtragem, segmentação ou classificação do sinal pupilar por meio de técnicas de IA, ou que buscaram auxiliar diagnóstico considerando a estimulação luminosa da pupila, proporcionando uma análise sobre os progressos realizados no emprego do reflexo pupilar à luz.

3.1 Metodologia da Revisão

A metodologia usada nessa revisão foi proposta por Kitchenham [Kitchenham 2004] que considera o planejamento, a condução e a produção de relatórios com os resultados da revisão, as principais etapas de construção de uma Revisão Sistemática da Literatura (RSL).

3.1.1 Planejamento

Na fase de planejamento, artigos previamente classificados como relevantes foram usados para realizar buscas exploratórias e apoiar a definição dos critérios de pesquisa. Nessa etapa definiu-se que:

- A busca na literatura seria feita nas bases de dados ACM Library¹; IEEE Explorer²; Science Direct³ e PubMed⁴.
- Seriam consideradas dentro do escopo dessa revisão, as pesquisas realizadas entre 2009 a 2019.
- Para serem considerados aptos a constarem nessa RSL os estudos deveriam ser originais, realizados em humanos e utilizariam a pupilometria automatizada para ava-

¹Acesso em: <https://dl.acm.org/search/advanced>

²Acesso em: <https://ieeexplore.ieee.org/search/advanced>

³Acesso em: <https://www.sciencedirect.com/search/advanced>

⁴Acesso em: <https://www.ncbi.nlm.nih.gov/pubmed/advanced>

liar o PLR. Também foram incluídos estudos que buscaram aprimorar dispositivos de gravação, segmentação das imagens e classificação.

- Estudos incompletos, não escritos em língua inglesa, indisponíveis na internet ou que não foram revisados por pares foram considerados inaptos para esta RSL.
- As palavras-chave utilizadas nas buscas foram: “pupil”, “pupillometry”, “pupillometer”, “pupillary”, “pupillometric”, “pupillography”; concatenados com o operador “OR”, e as palavras-chave “teacher” e “school” concatenadas com o operador “NOT”. A *string* de busca resultante foi: “*pupil OR pupillometry OR pupillometer OR pupillary OR pupillometric OR pupillography NOT teacher NOT school*”. Essa *string* foi aplicada nos metadados título, resumo e palavras-chave do autor.

3.1.2 Seleção

As buscas primárias foram realizadas em março de 2020 e retornaram 5547 referências, sendo 452 do repositório ACM Library, 1025 do IEEE, 2112 do PubMed e 1958 do ScienceDirect, Fig. 3.1, que resultaram em 5547 referências e que após a remoção das 363 referências duplicadas resultaram em 5184 referências para leitura do título e resumo, as quais foram avaliadas por avaliadores independentes (H.P e R.C).

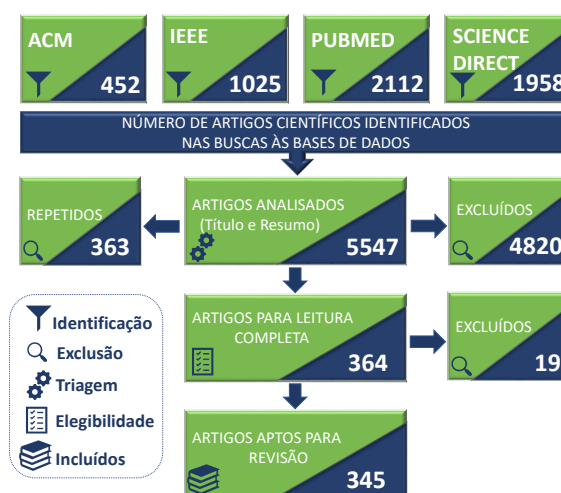


Figura 3.1: O processo de revisão sistemática.

Os artigos foram então avaliados pelos seguintes critérios de exclusão: (I) utilizar estimulação luminosa da pupila (exclusivamente ou em conjunto com outros estímulos) para auxiliar em diagnósticos, (II) estabelecer valores de referência para o comportamento pupilar, (III) aprimorar aspectos técnicos da pupilometria (e. g. desenvolver dispositivo de gravação, desenvolver algoritmos de detecção e segmentação pupilar, desenvolver técnicas para detecção de piscadas ou remoção de artefatos). A aplicação desses critérios produziu a exclusão de 4820 artigos e resultou em 364 artigos para leitura completa.

A leitura completa e aplicação dos critérios de exclusão resultando em mais 19 artigos desconsiderados e por fim selecionou 345 artigos para extração de informações. Dessas referências, 174 utilizaram o reflexo pupilar à luz para auxiliar diagnósticos e 171 estabeleceram valores de referência para o comportamento pupilar, ou desenvolveram dispositivos, ou protocolo de gravação, trabalharam na detecção e segmentação pupilar ou fizeram revisões acerca desses assuntos.

Para compreender detalhadamente os avanços no uso do reflexo pupilar à luz, essa revisão, buscou responder as seguintes questões do ponto de vista computacional sobre as técnicas aplicadas:

- Q 1: Quais os dispositivos utilizados para aquisição das imagens na realização da pupilometria?
- Q 2: Qual o protocolo aplicado na realização do exame?
- Q 3: Quais os métodos utilizados para mediar o diâmetro pupilar (detecção e segmentação pupilar)?
- Q 4: Quais as técnicas utilizadas para tratar ruídos e outros artefatos como piscadas?
- Q 5: Para apoio ao diagnóstico de quais patologias a pupilometria tem sido aplicada?
- Q 6: Quais técnicas de Inteligência Artificial (IA) têm sido aplicadas e quais os resultados obtidos?

3.2 Resultados da Revisão Sistemática

3.2.1 Dispositivos usados para gravação - Pupilômetro

Existe uma variedade de dispositivos usados para realizar a pupilometria, os quais tem sido e aprimorados para medir com precisão o diâmetro pupilar, e para permitir que se configure as condições de medição e o protocolo de gravação necessários às pesquisas.

Dentre os diversos modelos de pupilômetros comerciais, os mais usados em auxílio a diagnósticos foram os pupilômetros da NeurOptics (20% dos artigos), em especial o modelo NeurOptics NPi-200 (NeurOptics Inc., Irvine CA USA) aplicado em 20 estudos. O RAPDx (Konan Medical USA, Irvine, California, USA) aparece logo em seguida, tendo sido o equipamento escolhido em 15 pesquisas (13%), o ViewPoint EyeTracker (Arrington Research, Scottsdale, AZ) foi aplicado em 14 pesquisas (12%) e o Compact Integrated Pupillograph (CIP; AMTech, Dossenheim, Germany) por 12 (10%). Estes quatro pupilômetros comerciais juntos representam mais da metade de todas as

pesquisas que optaram por dispositivos comerciais (52%), a seguir eles serão brevemente descritos.

O NeurOptics NPi-200 trabalha com imagens infravermelho. A estimulação pupilar é realizada com luz branca aplicada com duração e intensidade configuráveis. Ele foi bastante empregado na medição da reatividade pupilar em pacientes críticos com lesões cerebrais avançadas, pois disponibiliza o cálculo automático do NPi (índice neurológico pupilar). Esse índice é um valor escalar que varia de 0 a 5, valores maiores que 3 indicam comportamento normal, valores inferiores sugerem um prognóstico de lesão cerebral. Esse índice é calculado comparando-se a amplitude e a velocidade do reflexo pupilar com valores de referência [Chen et al. 2011]. Nos 20 artigos em que este dispositivo foi utilizado, a pupilometria dinâmica foi aplicada para identificar lesões cerebrais, obtenção de valores de referência, prognóstico pós-acidente vascular cerebral, asma, delírio, neurite ótica, síndrome de Horner e identificação de consumo de drogas. Dentre os artigos que utilizaram este dispositivo, apenas Yung Ju Yoo *et al.* [Yoo, Yang e Hwang 2017] apresenta os valores de sensibilidade e especificidade na identificação da síndrome de Horner, 84,6% e 92,3% respectivamente.

O dispositivo RAPDx utiliza tecnologia de imagem infravermelha para análise ocular, com opções de estimulação pupilar usando luz branca ou policromática. As filmagens e a estimulação podem ser realizadas de maneira independente para cada olho, o que aumenta sua versatilidade. Além disso, o dispositivo oferece conectividade com redes de computadores e serviços em nuvem, expandindo suas possibilidades de uso. Ele foi empregado em 15 artigos para investigar condições como glaucoma, deficiência pupilar aferente relativa (DPAR), ambliopia, degeneração macular relacionada à idade e para a obtenção de valores de referência. Entre os estudos analisados, o destaque vai para o trabalho de Chang *et al.* [Chang et al. 2013], que reportou sensibilidade e especificidade de 80% na detecção do glaucoma.

O dispositivo ViewPoint EyeTrackers, utilizado para rastreamento ocular (*eye-tracking*), consiste em duas câmeras integradas a um suporte que acomoda a face do usuário para captura de imagens. Este sistema não oferece isolamento da iluminação externa nem dispõe de um sistema de iluminação ou estimulação da pupila. Ele foi empregado em 14 artigos visando identificar condições como glaucoma, diabetes, neurite ótica, hipertensão intracraniana idiopática, sono ou fadiga, e apneia obstrutiva do sono, além de auxiliar na obtenção de valores de referência. Jason C Park *et al.* [Park, Moss e McAnany 2016] reportaram uma sensibilidade e especificidade de 85% na detecção de hipertensão intracraniana idiopática. Contudo, os outros estudos não divulgaram índices de acurácia, sensibilidade ou especificidade.

O dispositivo Compact Integrated Pupillograph, utilizado em 12 artigos, apresenta características semelhantes ao ViewPoint EyeTrackers, como a ausência de isola-

Tabela 3.1: Pupilômetros usados nas pesquisas encontradas na literatura.

Pupilômetro	Qtd	Perc (%)
NeuroOptics NPi-200 (NeuroOptics Inc., Irvine, CA, USA)	20	17%
RAPDx (Konan Medical USA, Irvine, CA, USA)	15	13%
ViewPoint EyeTrackers (Arrington Research, Scottsdale, AZ, USA)	14	12%
Compact Integrated Pupillograph (CIP; AMTech, Dossenheim, Germany)	10	8%
MonPack One Software (Metrovision, Perenchies, France)	8	7%
NeuroOptics NPi-100 (NeuroOptics Inc., Irvine, CA, USA)	7	6%
EyeLink 1000 eye tracker (SR Research, Ottawa, ON, Canada)	7	6%
NeuroLight AlgiScan (iDMed, Marseille, France)	6	5%
NeuroOptics DP-2000 binocular pupillometer (NeuroOptics Inc., Irvine, CA, USA)	6	5%
IrisCorder Dual C10641 (Hamamatsu Photonics, Hamamatsu, Japan)	4	4%
NeuroOptics PLR-2000 (NeuroOptics, Irvine, CA, USA)	4	4%
TOBII eye tracker (Tobii AB, Danderyd, Sweden)	4	4%
ETL-100H Pupilometry Lab (ISCAN Inc., Woburn, MA, USA)	3	3%
GC660 (Allied Vision Technologies, Stadtroda, Germany)	3	3%
RK-7261 (ISCAN, Inc., Woburn, MA, USA)	2	2%

Quantidade de artigos que utilizaram o equipamento (Qtd). Percentual de artigos que usaram o dispositivo (Perc).

mento da iluminação externa. No entanto, ele inclui um sistema de iluminação e estimulação pupilar. Este dispositivo também consiste em um sistema de câmeras acoplado a um suporte onde o usuário posiciona o rosto para captar imagens dos olhos. Ele foi empregado em pesquisas voltadas para a identificação de condições como abstinência alcoólica, Alzheimer, enxaqueca, frequência cardíaca, hemianopsia, retinite pigmentosa, sono e fadiga. Notavelmente, nenhum dos estudos citados reportou os índices de acurácia, sensibilidade ou especificidade em seus resultados.

Os quatro dispositivos comerciais descritos anteriormente representam os mais empregados nas pesquisas ao longo dos últimos 10 anos. Além destes, existem outros pupilômetros disponíveis no mercado. A predominância do uso de dispositivos comerciais nas pesquisas deve-se à sua praticidade, precisão e à variedade de recursos que oferecem. No entanto, devido a fatores como custo ou necessidade de configurações mais flexíveis, alguns estudos optaram por desenvolver seus próprios dispositivos customizados. Os dispositivos comerciais identificados, incluindo os quatro já mencionados, são catalogados na Tabela 3.1, que elenca quinze tipos diferentes de dispositivos comerciais, ordenados do mais ao menos utilizado, conforme analisado em 115 artigos.

3.2.2 Protocolo de Gravação do PLR

No contexto da pupilometria com estimulação luminosa, o protocolo de gravação estabelece os métodos pelos quais as células fotorreceptoras serão estimuladas. Este protocolo define vários parâmetros importantes para o exame, incluindo o comprimento de onda da luz utilizada (*waveform*), a intensidade do estímulo luminoso, a duração de cada estímulo, a frequência dos exames, os períodos de adaptação à luminosidade tanto

inicial quanto entre os estímulos subsequentes. Além disso, o protocolo determina o tipo de reflexo pupilar a ser observado: reflexo direto, no qual o olho estimulado é também o olho filmado, ou reflexo consensual, no qual um olho é estimulado enquanto o outro é observado e gravado.

As células bastonetes, cones e as células ganglionares retiniais intrinsecamente foto sensíveis (intrinsically photosensitive retinal ganglion cells - ipRGCs) são células fotorreceptoras que induzem o PLR. As células bastonetes são responsáveis pela visão no escuro (visão escotópica). As células cone são sensíveis à cor e são responsáveis pela visão em ambiente claro (visão fotópica). Ambas estão envolvidas na visão em luz intermediária (visão mesópica) [Asakawa et al. 2019].

A especificidade de cada tipo de célula fotorreceptora influencia diretamente suas respostas e comportamentos em relação à cor (comprimento de onda da luz), intensidade, ângulo e duração da estimulação, conforme discutido por Lei *et al.* [Lei et al. 2014]. Portanto, é necessário que os dispositivos utilizados para a pupilometria sejam configuráveis, permitindo a adequação desses parâmetros conforme as necessidades específicas de cada pesquisa ou diagnóstico.

Uma análise detalhada foi conduzida na tentativa de identificar valores de referência que indicassem a configuração mais adequada ou mesmo o melhor protocolo de gravação para a pupilometria. O processo envolve uma série de variáveis, e as informações disponíveis nos artigos científicos são, muitas vezes, inconclusivas. Não foi possível identificar dois estudos que avaliassem a mesma patologia utilizando protocolos idênticos de gravação e aplicasse as mesmas técnicas de processamento computacional de imagens, evidenciando a diversidade e complexidade nas abordagens de pesquisa nessa área.

Existe a possibilidade de que o protocolo de gravação interfira na eficiência do método, mas não é possível realizar esta afirmação baseado nos artigos existentes. Por exemplo, Fan *et al.* [Fan et al. 2009] e Georgina *et al.* [Lynch, James e VanDam 2018] realizaram pesquisas para identificação do autismo e obtiveram acurácia de 92.5% e 72.2% respectivamente. Os protocolos de gravação adotados foram diferentes, e os pacientes examinados por Fan *et al.* eram crianças enquanto Georgina *et al.* examinou adolescentes, isso impossibilitou determinar o quanto a diferença nos protocolos de gravação influenciou nas acurácias obtidas e o quanto a diferença nos grupos estudados interferiu no resultado.

Em outros estudos que investigaram diferentes patologias, nota-se que tanto as acurácias apresentadas quanto os protocolos variam significativamente. Portanto, a definição de um protocolo adequado deve levar em consideração as características específicas da patologia e a classe de indivíduos sendo analisados. Isso destaca a necessidade de uma abordagem personalizada para otimizar os resultados da pupilometria. A seguir, serão apresentados os principais componentes dos protocolos de gravação.

3.2.2.1 Luminosidade do Ambiente de Captura

Devido à característica foto sensível das células bastonetes e cone, a luminosidade do ambiente onde é realizada a captura das imagens pode alterar as medições pupilares. Ong *et al.* [Ong, Hutch e Smirnakis 2019] afirmam que para produzir resultados confiáveis os examinadores devem padronizar essa luminosidade.

3.2.2.2 Comprimento de Onda

O comprimento de onda define a cor da luz que será utilizada na estimulação. Dentre elas, a luz branca foi a mais utilizada para estimular a pupila, nas pesquisas avaliadas por essa RSL. Em seguida, as mais frequentes são as luzes monocromática Azul e Vermelha. A estimulação com luz branca foi a preferida por pesquisas anteriores a 2012 e também quando os estudos estão relacionados a lesões cerebrais. A estimulação com luz monocromática, entretanto, tem-se tornado a tendência em pesquisas mais recentes, especialmente as que buscam avaliar a resposta pós-iluminação.

Paul Richter *et al.* [Richter *et al.* 2017] destacam que a luz azul provoca reflexos pupilares mais intensos em comparação com outras cores. Park *et al.* [Park *et al.* 2011] complementam essa observação, recomendando o uso de estímulos na cor azul para acessar a resposta pupilar e sugerindo a utilização de luz vermelha de baixa intensidade para avaliar a reatividade das células cone. Além disso, eles concluem que a resposta das células melanopsina é mais eficazmente acessada através da aplicação de luz monocromática azul em alta intensidade.

3.2.2.3 Duração do Estímulo

O estímulo luminoso costuma ser aplicado em pulsos, com durações que variam de alguns milissegundos até vários segundos, ou pode ser realizado por meio de cintilações (*flickers*) em frequências específicas. Park *et al.* [Park *et al.* 2011, Park e McAnany 2015] indicam que é possível isolar as contribuições de bastonetes, cones e células melanopsina utilizando estímulos de 1 segundo. Em relação à resposta pós-estimulação, Shaobo Lei *et al.* [Lei *et al.* 2014] observam que a duração do estímulo necessário para medir a atividade pós-estimulação (PIPR) não precisa exceder 400 ms.

3.2.2.4 Intensidade

A intensidade da luminosidade aplicada à pupila é frequentemente medida em lux (lx), lúmen (lm) ou candela (cd). O lúmen quantifica a quantidade de luz emitida por uma fonte, enquanto o lux se refere à incidência de luz sobre uma superfície, sendo que 1 lúmen ilumina 1 metro quadrado com a equivalência de 1 lux. A candela, por sua vez, é

definida pela intensidade luminosa e descreve a distribuição uniforme de luz que irradia de um ponto em direção a um local específico.

Park *et al.* [Park *et al.* 2011] destacam que a atividade dos bastonetes é melhor investigada com estímulos de baixa intensidade, enquanto as células cone e melanopsina requerem flashes de alta intensidade após uma adaptação ao escuro. Shaobo Lei *et al.* [Lei *et al.* 2014] observam que o PIPR (Post-Illumination Pupil Response) aumenta monotonicamente com a intensidade do estímulo, variando de 0,1 até 40,0 cd/m^2

3.2.2.5 Período de Adaptação

Diz respeito ao tempo que se aguarda para que as células, bastonetes e cones se adaptem à iluminação ambiente. Refere-se à adaptação inicial, antes da medição pupilar ou à adaptação entre estimulações. Park *et al.* [Park *et al.* 2011] sugere ser necessário 10 minutos de adaptação inicial ao escuro, antes de se realizar testes sobre o reflexo pupilar. Bin Wang *et al.* [Wang *et al.* 2015] sugerem 20 minutos de adaptação ao escuro. Ken Asakawa *et al.* [Asakawa *et al.* 2019] por sua vez, afirmam que a iluminação natural é suficiente para captar a resposta cone com 5 minutos de adaptação ao ambiente claro. Afirmam também que a resposta bastonetes pode ser obtida com pelo menos 10 minutos de adaptação ao escuro.

3.2.2.6 Acomodação entre Estímulos

As células bastonetes e cone requerem um período de adaptação para se ajustar à iluminação ambiente. Em estudos que não focam na influência acumulada da iluminação, é essencial incorporar um intervalo de acomodação entre os períodos de estimulação, permitindo que o diâmetro pupilar retorne ao seu estado inicial antes do próximo estímulo. Este intervalo evita que a resposta ao estímulo subsequente seja afetada pelo anterior, conforme destacado em pesquisas anteriores [Becktepe *et al.* 2019, Najjar *et al.* 2018, Ishibashi *et al.* 2017].

3.2.2.7 Reflexo Direto ou Consensual

O reflexo pupilar à luz pode ser induzido diretamente no olho que será avaliado, nesse caso o reflexo utilizado é o direto. Ou pode ser induzido pelo outro olho (visto que a contração pupilar ocorre de maneira sincronizada) e nesse caso o reflexo utilizado é o consensual. Sindri Traustason *et al.* [Traustason *et al.* 2016] afirma que não há diferença na resposta direta ou consensual com iluminação azul. Já com iluminação vermelha, afirma haver uma resposta ligeiramente maior durante a iluminação consensual, quando comparada com a direta.

Tabela 3.2: Protocolo de gravação e estimulação.

Aspecto	Tipos	Moda
Luminosidade	Escotópica, Mesópica, Fotópica	Escotópica
Cor do Estímulo	Branca, Azul, Vermelha, Verde, Amarela	Branca
Reflexo	Direto, Consensual, Ambos	Consensual
Duração do Estímulo	10 ms a 2 min	1 s
Intensidade	1 a 1000 lux	100 lux
Repetição das Gravações	1 a 40 vezes	1 vez
Acomodação Inicial	2 s a 40 min	10 min
Acomodação entre Medições	100 ms a 15 min	30 s

A coluna moda traz o tipo de protocolo utilizado com mais frequência nas pesquisas.

3.2.2.8 Quantidade de Repetições da Estimulação

A maioria das pesquisas aplica uma única estimulação à pupila; no entanto, algumas patologias podem ser melhor identificadas observando-se o tempo necessário para que as células retornem ao estado original após a estimulação, assim como suas reações a estímulos subsequentes. Por exemplo, o estudo de Daniel Bittner *et al.* [Bittner et al. 2014] repetiu a estimulação pupilar 40 vezes para investigar as alterações no reflexo à luz da pupila em pessoas com a doença de Alzheimer.

A Tabela 3.2 sintetiza os protocolos de gravação e estimulação encontrados em pesquisas que utilizam o PLR para auxiliar diagnósticos. A coluna aspecto mostra diferentes maneiras de se realizar a pupilometria. A coluna tipo mostra as possibilidades encontradas, por exemplo, quanto ao aspecto *Wavelength* encontrou-se pesquisas que utilizaram estimulação nas cores Branca, Azul, Vermelha, Verde e Amarela. A coluna moda informa o valor quantitativamente mais frequentemente nas pesquisas, por exemplo, a cor de estimulação mais frequente foi a branca.

A análise dos protocolos mostrou diversas maneiras de se fazer a pupilometria. A definição de cada aspecto do protocolo deve ser associada ao comportamento das células fotorreceptoras que serão avaliadas.

3.2.3 Pesquisas em Algoritmos para Segmentação

Para a realização eficaz da pupilometria computadorizada, é essencial a detecção e a segmentação precisas da região da pupila. A segmentação em imagens é realizada através da delimitação precisa das bordas da pupila e da extração da região de interesse. Esta seção reúne artigos que apresentam os principais algoritmos desenvolvidos para a detecção e segmentação da pupila.

A detecção pupilar em condições do mundo real enfrenta desafios significativos, como variações nas condições de iluminação, mudanças na distância e no ângulo de gravação, reflexos que criam pontos brilhantes, e a obstrução visual causada pelo uso

de óculos ou lentes de contato. Diante dessas dificuldades, diversas pesquisas têm focado no desenvolvimento de métodos mais eficazes para a detecção pupilar. Algoritmos como StarBust, Swirski, SET, ExCuSe, ElSe, PupilNet, PuRe e PuResT, dentre outros, foram projetados para aprimorar a precisão da medição pupilar, buscando torná-la robusta mesmo em ambientes não controlados.

O algoritmo StarBust proposto por Winfield *et al.* [Li, Winfield e Parkhurst 2005], localiza inicialmente os reflexos da imagem e os remove. Então usa uma técnica iterativa baseada em características para localizar as bordas da imagem. Então uma elipse é ajustada ao subconjunto de pontos de borda usando o paradigma *Random Sample Consensus* - RANSAC [Fischler e Bolles 1981]. Os melhores parâmetros de ajuste são usados para inicializar um modelo de pesquisa local para os parâmetros de elipse que maximizam o encaixe dos dados da imagem.

O algoritmo Swirski *et al.* [Świrski, Bulling e Dodgson 2012] estima os contornos pupilares primeiramente aplicando *Haar-like features* para achar a região da pupila, depois segmenta a região da pupila agrupando o histograma com o *k-means* e aplicando o detector de borda Canny. Por fim utiliza o RANSAC para estimar o contorno pupilar.

O algoritmo SET proposto por Javadi *et al.* [Javadi *et al.* 2015] assume que a pupila é circular. Binariza a imagem e aplica a segmentação dos píxeis relacionados em grupos. As segmentações que contém menos que uma determinada quantidade de píxeis são descartadas. Para os segmentos restantes realiza o encaixe de elipses. A elipse que for mais circular é selecionada como o centro da pupila.

O algoritmo ExCuSe proposto por Fuhl *et al.* [Fuhl *et al.* 2015] é baseado em detecção de bordas e encaixes de elipses. Utiliza o detector de bordas Canny [Canny 1986] refinado com operações morfológicas. Apresenta as bordas às elipses para encontrar a melhor borda. Um segundo algoritmo detecta a posição estimada da pupila por meio de uma função de projeção integral angular e refina essa projeção até a convergência.

O algoritmo ElSe também proposto por Fuhl *et al.* [Fuhl *et al.* 2016] continua utilizando Canny e operações morfológicas para detectar as bordas, mas seleciona a melhor borda usando heurísticas tais como forma e encapsulamento. As bordas são então usadas para encaixar elipses e retornam o centro da pupila e seu contorno.

Os autores do PupilNet [Fuhl *et al.* 2016] e sua respectiva melhoria PupilNet V2.0 [Fuhl *et al.* 2017] propõem um método de detecção pupilar em imagens IR composto por um pipeline de duas CNNs. A primeira identifica de maneira não tão precisa a posição da pupila e as sub-regiões de interesse, a segunda refina a posição pupilar usando as sub-regiões e a posição inicial estimada.

George, e Aurobinda propõem o ESCaF [George e Routray 2018] um algoritmo para detecção do centro da pupila que utiliza tanto as bordas quanto as informações de intensidade, juntamente com a abordagem de filtragem de candidatos para identificar a

melhor pupila. Afirmam obter desempenho melhor que o do algoritmo ELSe.

O PuRe [Santini, Fuhl e Kasneci 2018] também utiliza Canny como detector para seleccionar segmentos de bordas curvas do contorno da pupila. Possui acurácia superior ao PupilNet baseado em Redes Neurais Convolucionais (CNN do inglês Convolutional Neural network). Perde para o Vera Olms em acurácia, mas segundo os autores tem menor tempo de processamento.

Setiawan *et al.* [Setiawan, Wibirama e Setiawan 2018] propõe uma metodologia usando *Hough circle transform* para localizar a pupila em imagens com oclusão. Afirmam ter melhor desempenho quando a pupila está entre 70 e 90% coberta.

Vera-Olmos e Malpica [Vera-Olmos e Malpica 2017] propõem um método também baseado em CNN que segundo os autores ultrapassam a acurácia de pesquisas anteriores em 19%.

Han *et al.* [Han *et al.* 2019] propuseram o uso indireto de CNN baseado em CNN U-Net para primeiramente segmentar a região da pupila e posteriormente encontrar seu centro de massa.

O PuReST [Santini, Fuhl e Kasneci 2018] é um aperfeiçoamento do PuRe, melhora seu desempenho em 5.44%. Ele utiliza informações do quadro corrente e do quadro anterior para melhorar seu desempenho, sendo classificado como *Pupil Tracker*.

Dogancan Temel *et al.* [Temel *et al.* 2019] utilizou o algoritmo Alex-Net [Krizhevsky, Sutskever e Hinton 2012], que foi treinado com a ImageNet para classificar objetos genéricos em 1.000 classes diferentes. Este modelo foi transformado de reconhecimento de objeto em um detector de pupila utilizando as camadas convolucionais da rede que geram representações visuais. O algoritmo adaptado obteve acurácia de 90.60%.

Também baseado em imagens IR obtidas em dispositivos montados na cabeça, Miron *et al.* [Miron *et al.* 2019] propõem uma arquitetura baseada em CNN sem treinamento prévio que consiste em 12 camadas convolucionais e duas camadas *max pooling* com 108.450 parâmetros de predição. Afirmam que a abordagem alcança precisão superior ao ExCuSe, ElSe, ESCaF e PupilNet V2.0 nas bases de dados ExCuSe.

Zhiqiang Zhao *et al.* [Zhao, Zhang e Zheng 2019] usa o algoritmo de filtragem bilateral para remover ruídos, executa equalização de histograma na imagem. Após a detecção do olho com o classificador Ada Boost aplica a transformada circular de Hough para identificação do centro da pupila. Com este algoritmo obteve acurácia de 97.90%.

Para imagens obtidas com iluminação visível (*visible light - VL*), Kitazumi e Nakazawa [Kitazumi e Nakazawa 2018] propõem o uso de CNN para detecção do centro e segmentação pupilar em imagens que não requerem iluminação IR. Afirmam alcançar acurácia na maioria das vezes superior a apresentada por Gou *et al.* [Gou *et al.* 2017].

Jun Ho Choi *et al.* [Choi *et al.* 2019] propuseram um algoritmo usando modelos

Tabela 3.3: Algoritmos usados para detectar e segmentar a pupila.

Id.	Algoritmo	Referência	Técnica	Base de Dados	Iluminação	Deteção	Acur (%) ¹
1	StarBust	[Li, Winfield e Parkhurst 2005]	RANSAC	LPW	IV	Pupila	16,65 ²
2	Swirski	[Świrski, Bulling e Dodgson 2012]	Haar Wavelets, Kmeans, RANSAC	LPW	IV	Pupila	49,76 ²
3	SET	[Javadi et al. 2015]	Convex Hull, Ellipse Fitting	LPW	IV	Pupila	29,42 ¹
4	ExCuSe	[Fuhl et al. 2015]	Histogram, Angular Integral Projection, Ellipse Fitting	LPW	IV	Pupila	47,79 ¹
5	ELSe	[Fuhl et al. 2016]	Canny, Ellipse Fitting	LPW	IV	Pupila	68,92 ¹
6	ESCaF	[George e Routray 2018]	Canny, Ellipse Fitting	LPW	IV	Pupila	73,23
7	PuRe	[Santini, Fuhl e Kasneci 2018]	Canny, Morphological Operations, K-Cosine Chain	ExCuSe, ElSe	IV	Pupila	76,71 ²
8	PuReST	[Santini, Fuhl e Kasneci 2018]	Morphological Operations, Mask of Previous Pupila, Canny	Swirski, ExCuSe, ElSe, LPW, PupilNet	IV	Pupila	87,62
9	Setiawan	[Setiawan, Wibirama e Setiawan 2018]	Circular Hough Transform	CASIA	IV	Pupila	80,00 ³
10	PupilNet	[Fuhl et al. 2016]	CNN	ExCuSe	IV	Pupila	65,88
11	PupilNet 2.0	[Fuhl et al. 2017]	CNN	Else, Excuse, PupilNet	IV	Pupila	79,60
12	Vera-Olmos	[Vera-Olmos e Malpica 2017]	CNN	ExCuSe, ElSe	IV	Pupila	82,17
13	Han	[Han et al. 2019]	UNet-CNN	ExCuSe, ElSe	IV	Pupila	83,00
14	RAPDNet	[Temel et al. 2019]	CNN, Circular Hough	ElSe	IV	Pupila	90,60
15	Miron	[Miron et al. 2019]	CNN	ExCuSe	IV	Pupila	92,91
16	Kitazumi	[Kitazumi e Nakazawa 2018]	CNN	UBIRIS.v2, GI4E	LV	Olho/Pupila	80,10 98,62
17	Choi	[Choi et al. 2019]	CNN	BioID, GI4E	LV	Olho/Pupila	93,30 99,60
18	Zhao	[Zhao, Zhang e Zheng 2019]	Ada Boost algorithm, Canny, Circular Hough	BioID, FERET	LV	Olho/Pupila	96,40 97,90

Identificador (Id.), Infravermelho próximo (IV), Luz Visível (LV), Acurácia (Acur), ¹ Capacidade de encontrar o centro da pupila nas imagens da base de dados, com tolerância de cinco píxeis. ² Dados obtidos de [George e Routray 2018]. ³ Com 70% de oclusão da pupila. Está destacada em negrito a melhor acurácia por abordagem e base de dados.

CNN que primeiramente localizam os pontos de referência nos rostos e após a remoção de artefatos como óculos fazem a localização da pupila. Em seus experimentos obtiveram acurácia de 99.60% na localização da pupila.

A Tabela 3.3 mostra alguns dos métodos computacionais usados para localizar e medir o diâmetro pupilar. Dentre os algoritmos que não usaram redes neurais artificiais, o PuReST, segundo os autores, é o algoritmo que apresenta o melhor desempenho. Dentre os baseados em redes neurais convolucionais, se destaca o método proposto por Miron *et al.*.

3.2.4 Pesquisas que Realizaram Classificação

Nessa seção serão mencionadas pesquisas que utilizaram o PLR como ferramenta para distinguir entre indivíduos saudáveis e patológicos, ou ainda que fizeram distinção entre pessoas em condição fisiológica normal de pessoas em condição alterada, isto é, sob a influência de drogas, medicamentos, estresse, desconforto, privação de sono, dentre outros fatores. Este processo de diferenciação, conhecido como classificação, geralmente se apoia no uso de métodos estatísticos ou técnicas de inteligência artificial para ser efetivamente realizado.

Tabela 3.4: Pesquisas que utilizaram técnicas envolvendo IA na classificação.

Referência	Propósito	Ano	Técnica	Acur (%)
[Czajka 2015]	<i>Liveness</i> ¹	2015	SVM	100,00
[Fan et al. 2009]	Autismo	2009	Perceptron algorithm	92,50
[Amodio et al. 2018]	Alcoolemia (use)	2018	SVM	90,48
[Yano, Zimmer e Ling 2015]	Biometria	2015	SVM	88,73
[Nowak et al. 2016]	Alzheimer	2016	Random Florest	87,60 saudável, 53,60 doente
[González et al. 2017]	Idade/Sobrepeso	2017	K-means	C1: 33,30 C2: 87,50, C3: 25,00, C4: 77,70
[Pinheiro et al. 2015]	Alcool (uso)	2015	SVM	85,19
[López et al. 2017]	Diabetes	2017	SVR	37,30

¹*Liveness detection* é uma área de pesquisa voltada a distinguir informações biométricas reais de possíveis fraudes. Acurácia (Acur).

As pesquisas que fazem análise estatística comparam características do PLR entre grupos de pessoas. Um grupo de controle (com pessoas sem patologia ou alterações) e um grupo alvo (composto por pessoas com alguma patologia ou condição fisiológica alterada). Esse último grupo pode ser subdividido em função da severidade da patologia, ou condição fisiológica. A comparação estatística busca identificar diferença significativa no comportamento das características pupilares entre os grupos.

As pesquisas que utilizaram técnicas de IA empregam na maioria das vezes aprendizado de máquina supervisionado. Essa abordagem usa entradas rotuladas na fase de treinamento para encontrar padrões e automaticamente classificar novas entradas. Elas buscam adaptar sua curva de classificação (ou pesos quando se tratando de redes neurais) de tal forma que novas entradas, caso sigam o padrão das amostras utilizadas no treinamento, sejam corretamente classificadas.

A Tabela 3.4 listas pesquisas que usaram técnicas de IA em seus estudos para classificação e auxílio ao diagnóstico baseado em estimulação luminosa da pupila. Desse apanhado, vê-se que o SVM (Support Vector Machine) foi a técnica mais utilizada. As Redes Neurais Profundas tem sido utilizadas na detecção pupilar, entretanto, observa-se que não foram encontradas pesquisas que empregassem diretamente as Redes Neurais na classificação de patologias ou no auxílio ao diagnóstico por meio da pupila. Portanto, percebe-se que há possibilidades a serem exploradas por meio da utilização de redes neurais para classificar o comportamento pupilar.

3.3 Discussão

Analisando-se os dados coletados, verificou-se que o pupilômetro mais utilizado nas pesquisas foi o NeurOptics NPi-200 (NeurOptics Inc., Irvine, CA, USA) (respondendo Q1). Esse dispositivo tem sido usado principalmente em pesquisas que investigam lesões cerebrais e disfunção autonômica. Diversas pesquisas, entretanto, pelo custo ou por peculiaridades do próprio estudo não utilizaram pupilômetros comerciais, preferiram desenvolver dispositivos próprios.

A Tabela 3.2 mostra os protocolos de estimulação mais utilizados (respondendo Q2). Destaca-se que o protocolo mais frequente por vezes coincide com o sugerido em pesquisas que buscaram definir um protocolo prático e adequado para a pupilometria. Por exemplo, o período de adaptação inicial de 10 minutos e a duração do estímulo de 1 segundo são os mais usados nas pesquisas, e estão conforme a recomendação de Asakawa *et al.* [Asakawa *et al.* 2019].

A luminosidade escotópica é a mais comum entre os estudos e apresenta benefícios, tais como, diminuição dos reflexos sobre a imagem, padronização da luminosidade e potencialização da amplitude de contração, pois em ambiente escuro a pupila estará em sua máxima dilatação. A estimulação com luz branca é a mais frequente, e se mostra eficaz quando o objetivo da pesquisa está relacionado à lesão cerebral e a disfunção autonômica, ela costuma ser realizada com intensidade de até 100 lux. Pesquisas voltadas à definição do protocolo [Park *et al.* 2011, Richter *et al.* 2017], sugerem que o azul é a mais eficaz para analisar as funções bastonetes e a vermelha a mais eficaz para analisar as funções cones.

A utilização do reflexo consensual é o mais usual, pois permite que as filmagens produzam imagens com menos reflexos. A pupila gravada não recebe o estímulo luminoso diretamente e, por isso, não precisa tratar os reflexos luminosos que seriam gerados pela estimulação.

Quanto aos métodos de detecção e segmentação pupilar, a Tabela 3.3 mostra os principais métodos computacionais utilizados para localizar e medir o diâmetro da pupila (respondendo Q3). Destaca-se nesse apanhado o algoritmo PuReST que apresenta o melhor desempenho dentre os algoritmos que não usaram redes neurais artificiais e o algoritmo proposto por Miron *et al.* que utiliza uma abordagem baseada em Redes Neurais Convolucionais.

Para fazer a detecção de piscadas, a maioria das pesquisas aplica filtro de mediana com janela variando de 200ms a 1s. Sendo essa a técnica mais utilizada para tratar piscadas e artefatos (respondendo Q4). Ela se baseia no fato das reações pupilares ocorrerem gradualmente ao longo do tempo, dessa forma, mudanças abruptas no diâmetro pupilar são consideradas ruídos.

A revisão realizada mostrou que o Reflexo Pupilar à Luz foi utilizado em pesquisas para auxiliar 54 tipos de diagnósticos ou condição fisiológicas sendo as pesquisas que estudam o glaucoma as mais frequentes, seguida pelos estudos da diabetes, lesão cerebral, defeito pupilar aferente relativo, cansaço e fadiga, (o que responde à questão Q5).

As características do reflexo pupilar mais utilizadas foram: o diâmetro máximo da pupila em repouso, a latência calculada para o início da contração e a velocidade da contração. Essas características pupilares foram consideradas em mais da metade dos estudos analisados.

Sobre as técnicas de inteligência artificial que costumam ser utilizadas (respondendo Q6), percebe-se que há muito a ser explorado nessa área, pois, a maioria das pesquisas voltadas ao diagnóstico se restringiram a fazer análise estatística do comportamento pupilar (94%), apenas uma fração delas utiliza técnicas de IA (6%). Quando há dados suficientes para treinamento, técnicas de IA tem se mostrado eficientes e podem ser utilizados no auxílio a diagnóstico. As Redes Neurais Profundas tem sido utilizadas na detecção pupilar, entretanto, não foram encontradas nessa RSL pesquisas que as empregasse diretamente na classificação de patologias ou no auxílio ao diagnóstico.

3.4 Considerações Finais do Capítulo

Neste capítulo, apresentou-se um panorama abrangente sobre pesquisas que se dedicaram à filtragem, segmentação e classificação do sinal pupilar, além daquelas que se propuseram a facilitar a obtenção de diagnóstico por meio da estimulação luminosa da pupila. A análise desses estudos possibilitou a formulação de recomendações para futuras investigações e evidenciou a necessidade de se explorar a aplicação de técnicas de aprendizado de máquina e redes neurais artificiais como suporte para a elaboração de diagnósticos por meio do PLR.

Desenvolvimento de um Repositório de Imagens da Pupila, Associado com uma Avaliação de Técnicas de Segmentação Pupilar

Esse capítulo mostra como foi criado um repositório com imagens manualmente anotadas da pupila que proporcionaram o treinamento de redes neurais para encontrar e medir o diâmetro pupilar em imagens infravermelho oriundas da pupilometria. Ele e detalha também como foi realizada a avaliação comparativa entre redes neurais convolucionais e outras técnicas aplicadas na segmentação pupilar.

4.1 Contextualização

Medir, quantificar e qualificar informações relacionadas ao olho humano tem despertado o interesse de diversos campos do conhecimento, tais como marketing, computação gráfica, psicologia, medicina, a indústria automobilística, dentre outras. Anteriormente, o exame pupilar se limitava a identificar, com o auxílio de uma lanterna, o estado de consciência de um paciente [Levatin, Prasloski e Collen 1973, Teasdale e Jennett 1974]. No entanto, devido aos avanços tecnológicos e aos algoritmos atualmente disponíveis, é possível realizar o exame pupilar de forma não invasiva, com maior riqueza de telhas e precisão.

Os avanços possibilitados por essas inovações tem aberto caminho para várias aplicações práticas. Por exemplo, rastrear o foco de atenção de uma pessoa monitorando a direção do seu olhar, o que permite ao setor de marketing desenvolver campanhas e posicionar produtos ou anúncios exatamente onde os clientes estão olhando [Wedel, Pieters et al. 2008]. A indústria de jogos digitais também pode aprimorar a interação do usuário com a tela ao estudar o comportamento visual dos jogadores [Corcoran et al. 2012]. No setor automotivo, é possível obter informações sobre o nível de atenção e sonolência do motorista gravando seus movimentos oculares [Liu et al. 2010, Sahayadhas, Sundaraj e Murugappan 2012, Al-Madani et al. 2021].

Essas possibilidades melhoram significativamente as experiências, a segurança e a eficiência em vários domínios.

Esses avanços foram possíveis principalmente graças ao aprimoramento dos algoritmos utilizados. É bem conhecido que o desenvolvimento de algoritmos mais precisos requer uma vasta quantidade de dados. Para atender a essa necessidade, existem diversos conjuntos de dados disponíveis publicamente para a chamada "fase de treinamento do algoritmo". No entanto, é importante notar que a maioria desses conjuntos de dados contém apenas informações de indivíduos típicos e saudáveis [Pinheiro e Costa 2021].

Este trabalho apresenta a construção e disponibilização pública de um repositório de imagens do olho humano de pacientes diagnosticados com glaucoma em estágios inicial, moderado ou severo, pacientes diagnosticados com diabetes mellitus tipo II, e indivíduos que consumiram álcool sob supervisão, com medições realizadas usando um etilômetro em três níveis (0,0 mg/l, 0,3 mg/l e 0,5 mg/l).

Além do desenvolvimento do conjunto de dados, este estudo comparou algoritmos bem conhecidos usados para segmentação pupilar e rastreamento ocular, tais como StarBurst [Li, Winfield e Parkhurst 2005], Swirski2D [Świrski, Bulling e Dodgson 2012], ElSe [Fuhl et al. 2016], PuRe [Santini, Fuhl e Kasneci 2018] e PuReSt [Santini, Fuhl e Kasneci 2018], juntamente com modelos de redes neurais convolucionais Faster R-CNN [Ren et al. 2015], YOLOv4 [Bochkovskiy, Wang e Liao 2020], YOLOv5 [Jocher et al. 2022] e YOLOv7 [Wang, Bochkovskiy e Liao 2023]. O conjunto de dados produzido nesse trabalho foi utilizado para realizar este comparativo. Os resultados demonstraram que a rede YOLOv7 é a mais precisa na medição do diâmetro pupilar, superando os algoritmos clássicos não baseados em IA especializados na detecção e medição pupilar.

4.2 Trabalhos Relacionados

Imagens no espectro de luz infravermelho próximo são usadas em várias aplicações. Por não serem visíveis, elas não interferem na dinâmica de contração e dilatação da pupila. Além disso, não ofuscam nem incomodam a pessoa sendo gravada. Essas características fazem com que esse tipo de imagem seja frequentemente utilizado em exames e pesquisas que envolvam pupilometria e são as utilizadas nesse trabalho.

Em relação aos algoritmos de detecção de propósito geral, pode-se dizer que, antes de 2014, abordagens clássicas como a proposta pioneira de detecção de objetos apresentada por Viola e Jones [Viola e Jones 2001] eram empregadas. O descritor de características Histograma de Gradientes Orientados (HOG), proposto por Dalal [Dalal e Triggs 2005], e o detector de Modelo de Partes Deformáveis (DPM), introduzido por Felzenszwalb *et al.* [Felzenszwalb, McAllester e Ramanan 2008], que intro-

duziu o conceito da caixa delimitadora mínima, alcançaram bons resultados em tarefas de detecção de forma geral.

Algoritmos especializados na detecção da pupila, como ElSe [Fuhl et al. 2016], ExCuSe [Fuhl et al. 2015], PuRe [Santini, Fuhl e Kasneci 2018], e o PuReST [Santini, Fuhl e Kasneci 2018], também alcançaram bom desempenho na localização da região da pupila, mesmo em imagens desafiadoras para segmentação.

Desde 2014, abordagens inovadoras baseadas em redes neurais profundas começaram a emergir e alcançar bons resultados. Essas abordagens são geralmente categorizadas em dois tipos: redes detectoras de objetos de duas etapas e de etapa única.

O estudo de Mandal *et al.* [Mandal et al. 2021] apresenta o uso da rede Unet para localizar a pupila e medir seu diâmetro por meio de segmentação semântica. Eles também sugerem o uso da métrica *Dice Score Coefficient* (DSC) para comparar abordagens baseadas em redes neurais com técnicas clássicas não baseadas em redes neurais. Nosso trabalho por sua vez avalia as redes Faster R-CNN e as redes da série YOLO aplicadas à tarefa de localizar e medir o diâmetro pupilar.

A pesquisa de Cimbalnik *et al.* [Cimbalnik et al. 2022] apresenta um banco de dados contendo informações de EEG intracraniano (iEEG) e da pupilometria realizada usando um dispositivo de rastreamento ocular, juntamente com dados das vozes dos pacientes, todos eles com epilepsia ou depressão. O trabalho de Proença *et al.* [Proença et al. 2009] fornece o banco de dados de imagens infravermelho próximo (UBIRIS v2) para reconhecimento biométrico com a região da íris segmentada e uma análise comparativa entre bancos de dados públicos focados em reconhecimento biométrico. Nosso estudo se diferencia desses dois por usar imagens da pupila de pacientes com glaucoma, diabetes e indivíduos alcoolizados.

O estudo de Fuhl *et al.* [Fuhl, Rosenstiel e Kasneci 2019] apresenta uma arquitetura de CNN que realiza segmentação dupla da pálpebra e da pupila. Eles empregam um conjunto de dados público contendo 16.200 imagens com pálpebras e pupilas anotadas manualmente para treinamento em rede. Por fim, fornecem um repositório composto por 500 mil imagens segmentadas utilizando a rede proposta em seu trabalho.

Nosso estudo se diferencia desses pelo desenvolvimento de um repositório específico, pela análise comparativa detalhada do desempenho entre diferentes redes neurais destinadas à detecção de objetos (utilizando aprendizagem por transferência, diversos conjuntos de dados e uma variedade de *backbones*), e por incluir nessa análise comparativa técnicas tradicionais (não baseadas em redes neurais) empregadas em localizar e medir o diâmetro pupilar.

Para isso, criou-se um banco de dados de imagens pupilares para apoiar e aprimorar a avaliação da segmentação pupilar e algoritmos de rastreamento ocular. Esse banco de dados é composto por 10.000 imagens anotadas e 258.790 imagens não anotadas

de indivíduos com diagnóstico de doenças como glaucoma e diabetes, além de indivíduos sob efeito de álcool.

Essas três categorias de pessoas foram escolhidas para compor o conjunto de dados por abrangerem tanto patologias – diabetes mellitus tipo II e glaucoma – quanto condições fisiológicas representadas pela intoxicação alcoólica, as quais possuem potencial para serem identificadas pela análise da resposta pupilar à luz, e que por isso realçam a importância do reflexo pupilar à luz como marcador clínico.

O repositório produzido foi disponibilizado publicamente para fornecer à comunidade científica recursos que favoreçam o desenvolvimento de algoritmos e técnicas relacionadas às imagens infravermelho dos olhos.

4.3 Materiais e Métodos

Na introdução desta seção, será detalhada a composição do repositório e as características das imagens incluídas. Além disso, serão fornecidas informações sobre os dispositivos utilizados para gravação, o protocolo de aquisição das imagens, detalhes sobre os voluntários participantes e o procedimento de anotação manual das imagens. Em seguida, será explicado o processo empregado para comparar os diversos algoritmos de detecção e medição do diâmetro pupilar. Os conjuntos de dados utilizados foram fundamentais para o treinamento, teste e validação das redes neurais. Por último, orientações sobre como acessar o repositório completo estão disponíveis no Apêndice A.

4.3.1 Base de dados

4.3.1.1 Dispositivo de Gravação

Foram usados dois tipos de pupilômetros para registrar os movimentos pupilares, aplicar o protocolo de gravação e isolar os olhos dos voluntários da luz ambiente. As imagens que possuem um agrupamento de reflexos na forma de ‘ferradura’ em seu centro foram capturadas usando um pupilômetro baseado em Arduino e equipado com uma câmera de vigilância. Este dispositivo, denominado ‘aparelho SC’, foi utilizado exclusivamente para o conjunto de dados de álcool. O quantitativo de imagens obtidas com esse equipamento é de 1.334 imagens anotadas (13,4% do conjunto de dados anotado) e 11.500 imagens não anotadas (4,4% do conjunto de dados não anotado),

As imagens com quatro reflexões centrais foram capturadas utilizando o pupilômetro proposto por Pinheiro H. *et al.* [Pinheiro *et al.* 2015] e aperfeiçoado por Silva *et al.* [Silva *et al.* 2017] para viabilizar a pupilometria cromática; doravante denominado ‘aparelho PG’. Este dispositivo possui duas câmeras Point Grey Firefly MTV022 para

gravar imagens do olho desejado. Quatro LEDs infravermelhos próximos iluminam a pupila durante a gravação e não interferem na dinâmica da pupila. Para estimular a pupila com luz visível, o aparelho possui 3 LEDs no espectro de cores RGB que ficam ligados por um determinado período de registro para provocar a contração pupilar.

O quantitativo de imagens capturadas com esse dispositivo foi de 8.666 imagens anotadas (86,6% do conjunto de dados anotado) e 245.290 imagens não anotadas (95,6% do conjunto de dados não anotado). Todas as imagens são derivadas de vídeos capturados a uma taxa de quadros de 30 quadros por segundo. Esse dispositivo foi usado nos conjuntos de dados de álcool, diabetes e glaucoma. As características dos dispositivos de gravação, detalhes das imagens e conjuntos de dados que utilizam cada equipamento estão resumidos na Tabela 4.1.

Tabela 4.1: Detalhes das Imagens e dos Equipamentos de Filmagem.

Equipamento SC			
Câmera: Surveillance Câmera	Iluminação: 10 Mini IR LEDs	Estimulação: 1 White LED	Usado na Base: Alcool
Equipamento PG			
Câmera: Point Grey Firefly MTV022	Iluminação: 4 IR LEDs	Estimulação: 3 RGB LEDs	Usado na Base: Alcool, Diabetes, Glaucoma
Detalhes das Imagens de Ambos os Dispositivos			
Largura: 640 píxeis	Resolução Horizontal: 96 dpi	Vídeo Taxa de quadros: 30 fps	Bit Profundidade: 24 bits
Altura: 480 pixels	Resolução Vertical: 96 dpi	Dimensões: 640 x 480	Formato: JPEG

4.3.1.2 Eyes Project (Alcoolemia) - Base de dados de pessoas que consumiram álcool

Este conjunto de dados compreende 17.000 imagens, sendo 2.000 anotadas manualmente e 15.000 imagens não anotadas; contém imagens de pupilas contraídas e dilatadas obtidas sem aplicação de substâncias. Em vez disso, estímulos luminosos foram empregados ao olho oposto ao gravado, utilizando o reflexo consensual [Campbell William W. Barohn 2021]. As imagens são RGB, possuem resolução de 640 x 480 píxeis, foram obtidas de vídeos com taxa de gravação de 30 quadros por segundo.

A figura 4.1 apresenta a distribuição de imagens e indivíduos por classe e gênero, anotadas manualmente e não anotadas.

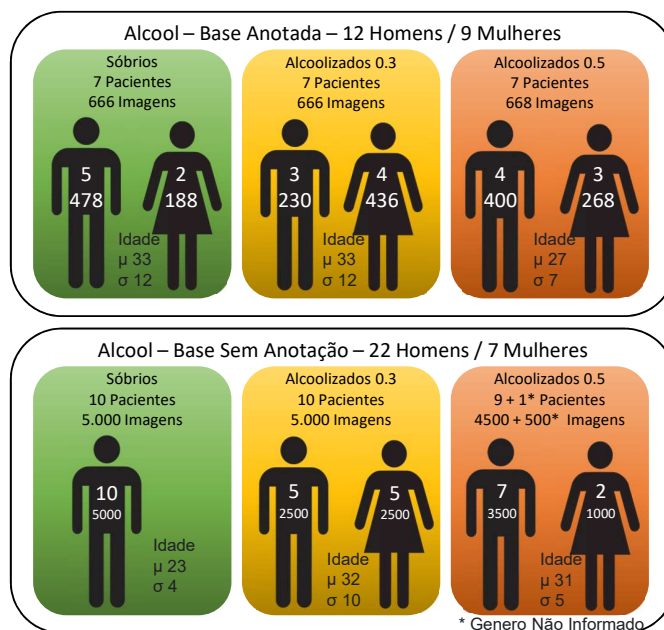


Figura 4.1: Distribuição de 2.000 imagens anotadas e 15.000 imagens não anotadas no conjunto de dados sobre álcool de acordo com gênero e nível de consumo de álcool.

O conjunto de dados inclui imagens de indivíduos sóbrios sem ingestão de álcool, indivíduos com 0,3 mg/l e aqueles com 0,5 mg/l. Um etilômetro fabricado pela Mercury foi utilizado para medir o nível de álcool no sangue da pessoa.

O processo de captura da imagem começou após 2 minutos de adaptação ao escuro (não gravados). A etapa inicial de gravação compreendeu um período de aquisição de 30 segundos sob completa escuridão. Seguiram-se 30 segundos de exposição à luz branca com iluminância média de 1.550 lux. O dispositivo foi calibrado para este nível de emissão de luz, e a iluminância real foi quantificada usando o luxímetro digital Mastech PM6612. Posteriormente, o procedimento foi repetido, com outro ciclo de 30 segundos no escuro e 30 segundos sob estimulação luminosa. O protocolo de aquisição está descrito na Figura 4.2.

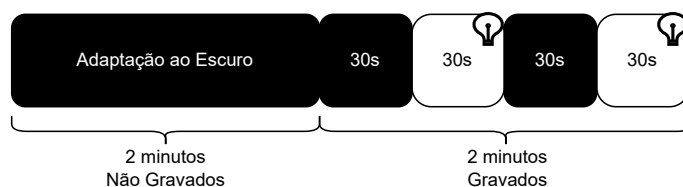


Figura 4.2: Representação do protocolo de aquisição das imagens para a base álcool.

Cada imagem apresenta uma barra informativa verde na parte inferior, exibindo detalhes como nome do projeto, Identificador do Exame (Id), Tipo de Estímulo (com símbolo “.” indicando ausência de estímulo e a letra W significando estímulos de cor

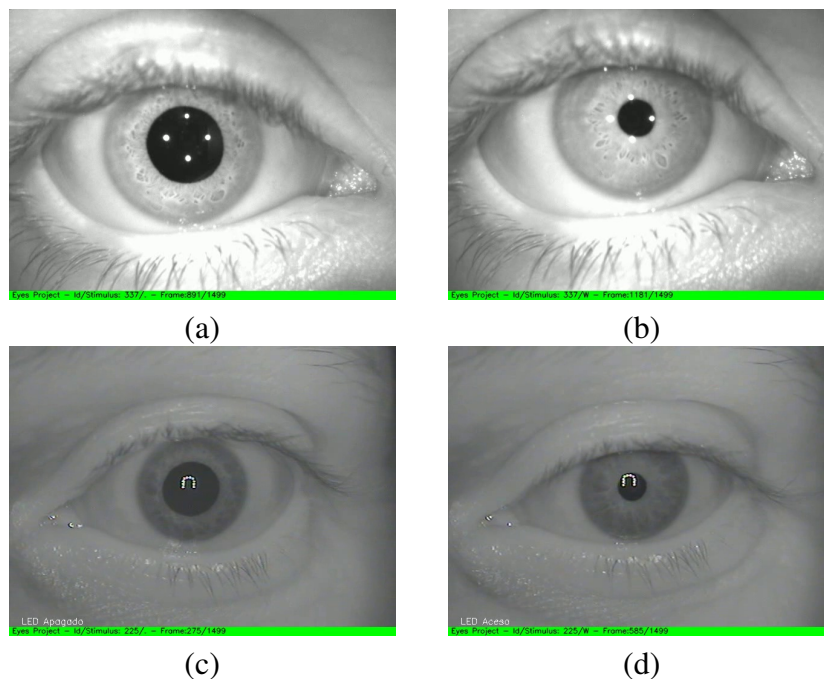


Figura 4.3: Imagens representativas de infravermelho (IR) amostradas do conjunto de dados. (a) Pupila dilatada gravada com o aparelho PG. (b) Pupila contraída gravada com o aparelho PG. (c) Pupila dilatada gravada com o aparelho SC. (d) Pupila contraída gravada com o aparelho SC.

branca), o número de quadros específico e a contagem total de quadros do vídeo. A figura 4.3 mostra alguns exemplos de imagens de conjuntos de dados e, por exemplo, na imagem (a), o projeto é identificado como *Eyes Project*, o identificador do exame é 337, o tipo de estímulo é “.” (sem estímulo), o número do quadro é 891 e o número total de quadros no vídeo é 1499.

4.3.1.3 *Eyes Project (Diabetes)* - Conjunto de dados de pessoas com Diabetes Mellitus Tipo II

O conjunto de dados de indivíduos com diabetes mellitus tipo II compreende 33.400 imagens de 24 pacientes, 1.400 anotadas manualmente e 32.000 sem anotações. O conjunto de dados contém imagens de pupilas contraídas e dilatadas, obtidas sem o uso de substâncias dilatadoras, utilizando apenas adaptação ao escuro e um estímulos luminosos em um dos olhos. A média de idade dos pacientes durante a fase de treinamento e teste do algoritmo é de 33 anos, com desvio padrão de 12 anos. A média de idade é de 55 anos para as imagens não anotadas, com desvio padrão de 7 anos.

A Figura 4.4 retrata a distribuição dos pacientes e das imagens por gênero e olho gravado. Neste conjunto de dados o protocolo seguido compreende um período inicial de adaptação ao escuro de 10 minutos, seguindo recomendado pela revisão de Pinheiro H. *et al.* [Pinheiro e Costa 2021]. A metodologia incluiu quatro estimulações cromáticas de

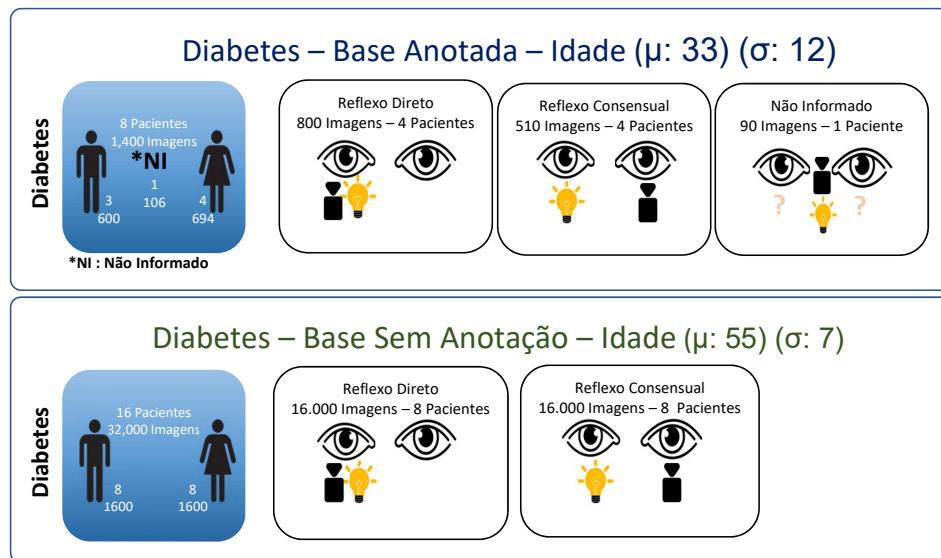


Figura 4.4: Distribuição de 1.400 imagens anotadas e 32.000 imagens não anotadas entre pacientes com Diabetes Mellitus tipo II, categorizadas por gênero e olho gravado.

LED: começando com comprimento de onda de 623 nm (vermelho), passando para 466 nm (azul), depois 517 nm (verde) e culminando com luz branca.

A seleção de cores para estimular a resposta da pupila foi orientada pelos resultados apresentados no estudo de Rukmini *et al.* [Rukmini, Milea e Gooley 2019] que recomenda o uso de luz vermelha e azul e pelo trabalho de Crippa *et al.* [Crippa, Domellöf e Kawasaki 2018] que sugeriu o uso de luz verde. Ambos os estudos destacaram a importância da pupilometria cromática na avaliação da saúde dos fotorreceptores no nervo óptico e na retina para diversas doenças. Cada estímulo foi caracterizado por duração e intensidade de 1 segundo e 250 cd/m², respectivamente, conforme as diretrizes de Park [Park *et al.* 2011].

O intervalo de adaptação entre os estímulos foi fixado em 59 segundos, seguindo as recomendações de Gracitelli *et al.* [Gracitelli *et al.* 2014]. A figura 4.5 representa o protocolo e a dinâmica dos estímulos. O tipo de estímulo apresentado na barra informativa verde na parte inferior das imagens deste conjunto de dados é indicado por “.” para significar a ausência de estímulo. As letras “R”, “B”, “G” e “W” correspondem, respectivamente, a estímulos nas cores vermelha, azul, verde e branca.

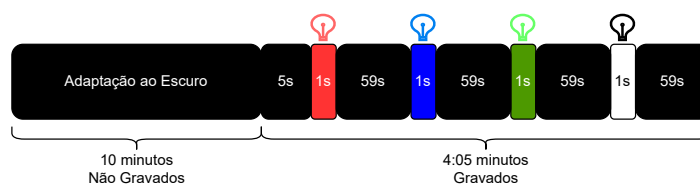


Figura 4.5: Representação do protocolo de aquisição para conjunto de dados de diabetes.

4.3.1.4 *Eyes Project (Glaucoma) - Conjunto de Dados de Pessoas com Glaucoma*

O conjunto de dados de pacientes com glaucoma inclui 218.390 imagens, das quais 6.600 são anotadas manualmente e 211.790 não possuem anotações. Assim como as imagens de pacientes com diabetes, as imagens com pupilas contraídas e dilatadas foram obtidas sem a aplicação de substâncias, utilizando apenas estímulo luminoso.

Este subconjunto de dados é composto por imagens de 10 pacientes sem glaucoma, com idade média de 42 anos e desvio padrão de 13 anos. Imagens de 30 pacientes com glaucoma em estágio inicial, com idade média de 61 anos e desvio padrão de 13 anos. Imagens de 30 pacientes com glaucoma em estágio moderado, com idade média de 58 anos e desvio padrão de 14 anos. E imagens de 19 pacientes com glaucoma em estágio grave, com idade média de 56 anos e desvio padrão de 11 anos.

A Figura 4.6 mostra a distribuição dos pacientes por gênero, o tipo de estímulo, tipo de reflexo, olho registrado e estágio da doença no subconjunto de dados anotados do glaucoma. A Figura 4.7 exibe a distribuição de pacientes no subconjunto de dados do glaucoma não anotado, categorizado em estágios da doença (sem glaucoma, glaucoma em estágio inicial, moderado e severo) distribuídos por gênero, idade média, olho registrado e reflexo direto ou consensual utilizado.

Um método para classificar o dano funcional no glaucoma deve ser objetivo, reprodutível e fácil de usar, fornecendo informações sobre as características dos defeitos no campo visual e sendo consistente com os dados de dano estrutural e capaz de monitorar pequenas mudanças na perda funcional ao longo do tempo [Jr e Vessani 2009]. A categorização do glaucoma nesse estudo foi feita seguindo os critérios de Hodapp, Parish e Anderson [Hodapp 1993]. O sistema de classificação HPA considera dois critérios: a extensão geral do dano no campo visual, o número de pontos defeituosos e a proximidade do(s) defeito(s) [Jr e Vessani 2009].

Para os grupos de pacientes com diabetes mellitus tipo II e glaucoma, foram realizados testes adicionais para avaliar a condição do paciente: Medição de acuidade visual, refração, teste de visão de cores, teste de sensibilidade ao contraste, teste de campo visual, pressão intraocular, biomicroscopia (catarata), fundoscopia (retinopatia diabética e glaucoma), retinografia com angiografia, varredura macular Cirrus HD-OCT, glicemia de jejum, glicemia pós-prandial e teste oral de tolerância à glicose.

A medição do diâmetro pupilar do grupo de pacientes com glaucoma utilizou o mesmo equipamento e o mesmo protocolo aplicado aos pacientes com diabetes tipo II, conforme apresentado na Figura 4.5.

As imagens e demais informações deste conjunto de dados foram analisadas pelo Comitê de Ética em Pesquisa (CEP) e submetidas à Plataforma Brasil, recebendo o código de registro CAAE 23723213.0.0000.5083.

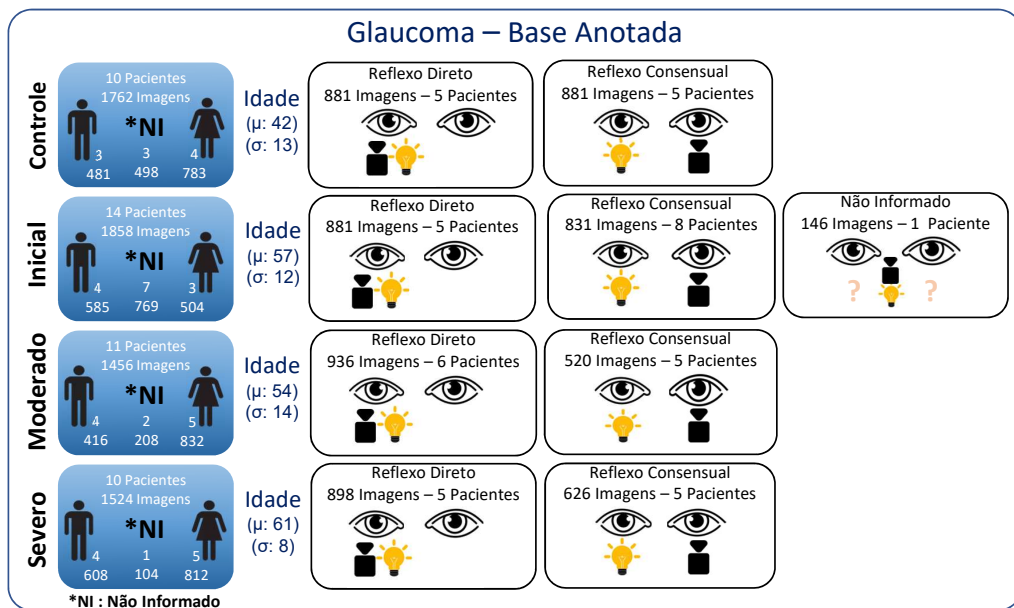


Figura 4.6: Distribuição das 6.600 imagens anotadas manualmente de pacientes com glaucoma, categorizadas por estágio da doença, gênero e olho registrado.

Todo o conjunto de imagens e anotações para o álcool, diabetes e glaucoma fazem parte da construção contínua da plataforma de software *Eyes Project*, destinada a auxiliar pesquisadores a investigar patologias e condições fisiológicas relacionadas ao reflexo pupilar à luz.

4.3.1.5 Anotação Manual (*ground truth*)

A anotação das pupilas foi realizada com o auxílio do software LabelImg [Tkachenko et al. 2020-2022], integrado ao Label Studio, como ilustrado na Figura 4.8. As anotações foram salvas em arquivos XML, estruturados conforme o padrão PascalVoc e com o mesmo nome da imagem correspondente. Foram realizadas anotações em situações diversas, inclusive em casos de pálpebras parcialmente fechadas, explorando a capacidade dos modelos modernos de redes neurais em detectar pupilas sob condições de oclusão parcial.

Em contrapartida, optou-se por deixar vazios os arquivos de anotação relativos a imagens nas quais a pupila se encontra totalmente oculta, evidenciando a dificuldade de determinar a localização da pupila nessas circunstâncias. A presença dessas imagens no banco de dados cumpre um papel estratégico, beneficiando modelos como o YOLOv4, que se valem de exemplos negativos — isto é, imagens sem nenhum objeto da classe especificada — durante o treinamento e a validação. Por outro lado, modelos como o Faster R-CNN tendem a desconsiderar imagens sem anotações das coordenadas das caixas delimitadoras. O banco de dados inclui 129 imagens com essas características, oferecendo

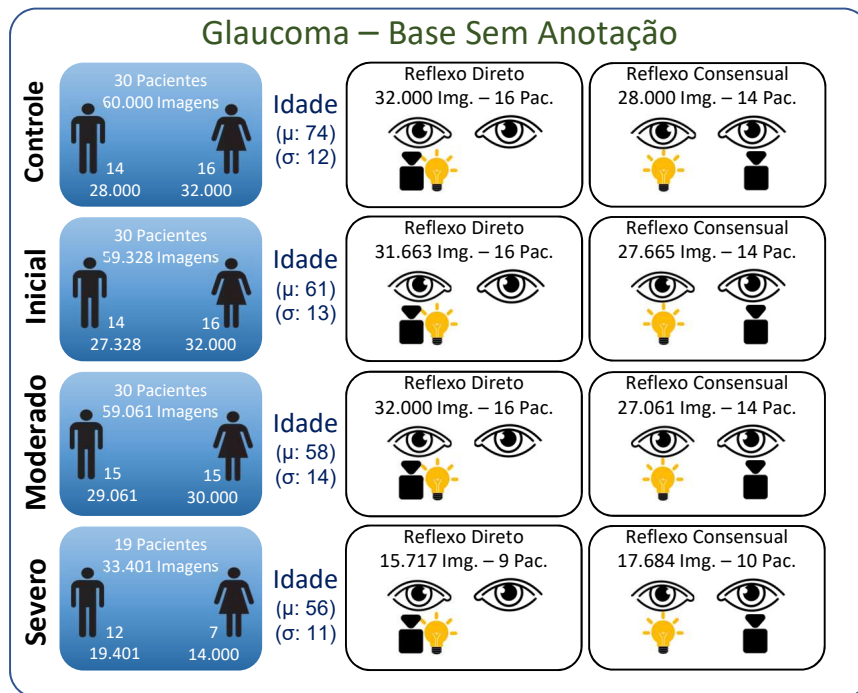


Figura 4.7: Distribuição das 211.790 imagens não anotadas de pacientes com glaucoma categorizadas por estágio da doença, gênero e olho registrado.

amostras para treinamento e validação de algoritmos em uma variedade de contextos.

4.3.2 Avaliação de Algoritmos para Detecção e Medição Pupilar

Nesta subseção, será discutido como comparar diferentes algoritmos usados para detectar a região da pupila em uma imagem. Inicialmente será explicando a Precisão Média, uma métrica bem conhecida para comparação de redes neurais usadas como detectores de imagens. Também será apresentada uma métrica proposta para comparar essas redes neurais com algoritmos clássicos que não retornam a caixa delimitadora da pupila, mas fornecem o centro pupilar e o diâmetro em píxeis.

4.3.2.1 Metodologia de Comparação entre Redes Neurais

Esta seção descreve a metodologia usada para comparar a precisão das redes neurais na identificação de caixas delimitadoras de alunos, empregando a métrica de precisão média como referência em relação às anotações manuais.

Para realizar esta análise, utilizou-se um conjunto de dados de 10.000 imagens anotadas manualmente, sendo 60% (6.000) utilizadas para treinamento, 20% (2.000) para validação e os 20% restantes (2.000) para teste da rede neural convolucional. O resumo

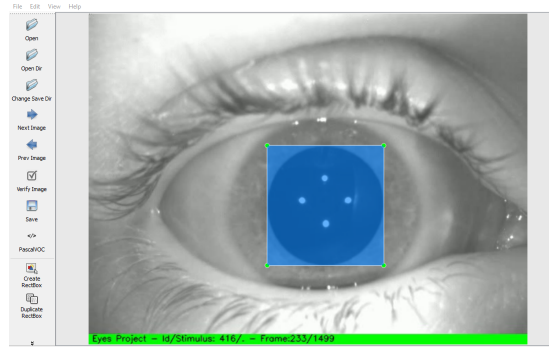


Figura 4.8: Exemplo da anotação da região da pupila usando o programa LabelImg.

dos modelos utilizados, juntamente com o tempo necessário para treinar e testar cada modelo, está sumarizado na Tabela 4.2.

Tabela 4.2: Características dos modelos.

Modelo (Head)	Backbone	Pesos (MB)	Imagens de Treinamento	Imagens de Validação	Imagens de Teste	Tempo Treino	Tempo Teste	QPS Treino	QPS Teste
Faster R-CNN	Resnet-C4-50	129	6.000	2.000	2.000	00:06:24	00:06:34	15,63	5,08
Faster R-CNN	Resnet-FPN-50	162	6.000	2.000	2.000	00:05:37	00:03:23	17,80	9,85
Faster R-CNN	Resnet-FPN-101	236	6.000	2.000	2.000	00:07:02	00:03:49	14,22	8,73
YOLOv4	CSPDarknet53	244	6.000	2.000	2.000	08:08:34	00:02:20	0,20	14,29
YOLOv5	CSPNet	169	6.000	2.000	2.000	09:19:08	00:01:59	0,18	16,81
YOLOv7	E-ELAN	554	6.000	2.000	2.000	03:25:47	00:02:32	0,49	13,16

Os tempos aproximados de teste e treinamento foram calculados usando uma GPU Tesla V100-PCIE-32GB. A coluna Quadros Por Segundo (QPS) Treino, representa a relação entre o tempo de treinamento e as 6.000 imagens usadas. A coluna QPS Teste representa o tempo de teste em relação às 2.000 imagens usadas.

Inicialmente foi testado o método de detecção de rede neural convolucional de dois estágios, a Faster R-CNN proposta por Ren *et al.* [Ren *et al.* 2015] em 2015, disponível na plataforma Detectron2 [Wu *et al.* 2019]. Ele usa uma arquitetura pré-treinada por 37 épocas no conjunto de dados de imagens COCO [Veit *et al.* 2016], contendo 330.000 imagens com 1,5 milhão de instâncias de objetos distribuídas em 80 categorias. No entanto, a pupila não está entre as categorias deste conjunto de dados. Portanto, para utilizar a rede para detectar a pupila, foi necessário retreiná-la com imagens da pupila. A técnica de aprendizagem por transferência foi utilizada para obter melhores resultados. Esta abordagem consiste em aproveitar informações de treinamentos anteriores em outro conjunto de dados.

A Faster R-CNN, e os subsequentes redes que serão avaliadas, foram retreinados por 2.000 épocas, usando 6.000 imagens para treinamento, 2.000 imagens para validação e 2.000 para testes, todas elas do conjunto de dados anotado manualmente. A Figura 4.9 apresenta um exemplo de localização da pupila, onde é ilustrada a caixa delimitadora mínima que inclui classe, pontuação de confiança e a elipse gerada a partir das coordenadas da caixa.

A rede Faster R-CNN também foi testada com o *backbone* C4 de 50 camadas que é o modelo original do artigo Faster R-CNN, ele utiliza uma única camada convolucional

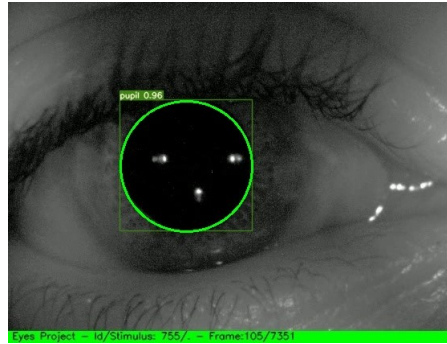


Figura 4.9: Exemplo de pupila localizada por redes neurais convolucionais, sua caixa delimitadora e a elipse correspondente, calculada a partir das coordenadas da caixa.

(geralmente a quarta camada, daí o nome C4) para a extração de características e todas as operações de extração de características são centralizadas nela.

Posteriormente a rede Faster R-CNN foi avaliada com *backbone* ResNet de 50 e 101 camadas usando o extrator de características *Feature Pyramid Network* (FPN). ResNet é abreviação de Residual Network (Rede Residual), ele se tornou popular devido à sua capacidade de permitir redes muito profundas utilizando conexões residuais, as quais ajudam a mitigar o problema de desaparecimento de gradiente, permitindo que a informação do gradiente flua através das camadas sem ser atenuado. No contexto da Faster R-CNN, utilizar uma ResNet como backbone significa aproveitar essa arquitetura profunda para uma extração de características mais rica. As ResNets são tipicamente identificadas pelo seu número de camadas, como ResNet-50, ResNet-101, etc., indicando a profundidade da rede.

Ao comparar C4 com ResNet, a principal diferença está na profundidade e complexidade da extração de características. As ResNets são mais profundas e mantêm com conexões residuais, o que pode resultar em melhor desempenho na detecção de objetos. Contudo, com aumento no custo computacional.

Posteriormente, foi testada a rede YOLOv4 (You Only Look Once). Que é um modelo de detecção de objetos de estágio único baseado em redes neurais convolucionais, proposto inicialmente por Joseph Redmon em 2016 [Redmon et al. 2016] e aprimorado na versão 4 por Alexey Bochkovskiy *et al.* [Bochkovskiy, Wang e Liao 2020]. O *backbone* padrão usado nesta rede é o CSPDarknet53 (derivado da arquitetura DenseNet). Esta rede neural, da mesma forma que Faster R-CNN, foi inicialmente treinada no conjunto de dados de imagens COCO e retreinada com imagens de pupilas.

O próximo modelo testado foi a YOLOv5, desenvolvida por Jocher *et al.* [Jocher et al. 2022]. Ela consiste em três partes: (1) *Backbone*: CSPDarknet, (2) *Neck*: PANet e (3) *Head*: camadas Yolo. Os dados entram no extrator de recursos CSPDarknet, passam pelo PANet para fusão de recursos e, finalmente, chegam às camadas de saída YOLO com resultados de detecção. A YOLOv5 inclui modelos diferentes, como YO-

LOv5s (parâmetros de 7,2 milhões), YOLOv5m (parâmetros de 21,2 milhões), YOLOv5l (parâmetros de 46,5 milhões) e YOLOv5x (parâmetros de 86,7 milhões). A configuração escolhida para o estudo foi o YOLOv5x, que possui um arquivo de pesos pré-treinado de 169 MB. Embora o modelo YOLOv5x tenha maior número de parâmetros em comparação com outros modelos, ele foi escolhido por ser apontado pelos autores como o mais preciso¹.

Por fim foi testado o modelo YOLOv7, oficialmente apresentada em julho de 2022 por Chien-Yao Wang, Alexey Bochkovskiy e Hong-Yuan Mark Liao. Segundo os autores, esta iteração da série YOLO incorpora melhorias arquitetônicas que aumentam a precisão da detecção de objetos com o mesmo custo computacional das versões anteriores [Wang, Bochkovskiy e Liao 2023]. A Figura 4.10 ilustra a curva de treinamento usando da YOLOv7, nela é possível observar que à medida que as épocas de treinamento progredem, o erro de localização da pupila diminui. Observa-se também que, apesar do modelo estar configurado para treinar por 2.000 épocas, o mecanismo de parada antecipada (*Early Stopping*) reduziu a duração real do treinamento do modelo YOLOv7 para 533 épocas. Os resultados em termos de AP da aplicação dos detectores mencionados estão resumidos na Tabela 4.3.

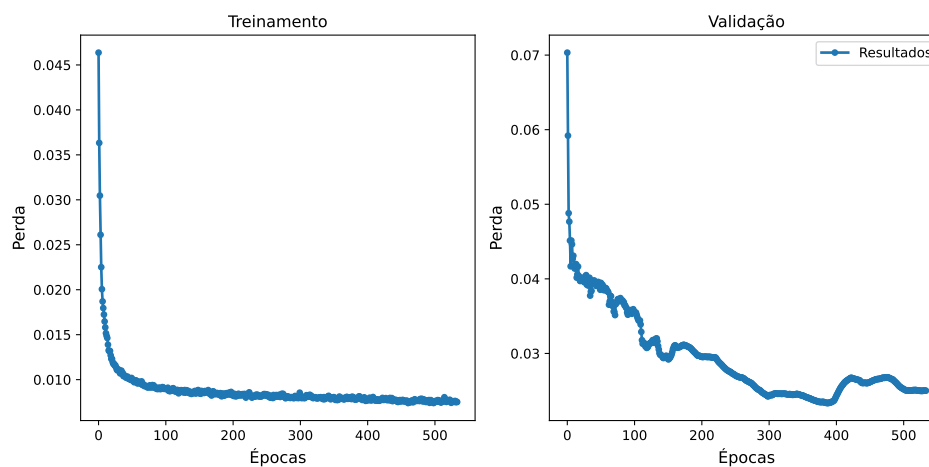


Figura 4.10: Curva de aprendizado da YOLOv7 ao longo das épocas de treinamento.

4.3.2.2 Metodologia de Comparação entre Redes Neurais e Algoritmos Clássicos

Nesta análise subsequente, será explicada a metodologia do comparativo entre algoritmos clássicos para medir o diâmetro da pupila (Pure, Purest, ExCuSe, ElSe, StarBust, Swirski2D, fornecidos por Zandi *et al.* [Zandi *et al.* 2021]) e modelos baseados em redes neurais, incluindo a Faster R-CNN (em três configurações distintas de backbone

¹Acesso em: <https://github.com/ultralytics/yolov5/releases>

e número de camadas) e YOLO (abrangendo as versões 4, 5 e 7). Esta avaliação busca evidenciar as distinções de desempenho e a aplicabilidade das abordagens de redes neurais no contexto da tarefa de detecção pupilar.

Para avaliar as técnicas tradicionais, que se limitam a determinar o diâmetro e o centro da pupila, em comparação com as metodologias baseadas em redes neurais modernas, que estabelecem as coordenadas da caixa delimitadora, este estudo adota a seguinte abordagem: Inicialmente, o diâmetro pupilar é estimado com base na maior dimensão da caixa delimitadora, seja altura ou largura, a fim de evitar que variações, como desvios do olhar, que podem reduzir a largura da caixa, ou pálpebras semiabertas, que diminuem a altura, comprometam a precisão da medição do diâmetro pupilar.

A seguir, a análise se concentra na comparação da similaridade entre o diâmetro da pupila medido manualmente em píxeis (D_{real}) e o diâmetro estimado pelos modelos (D_{estimado}). Quando a pupila não existia ou quando não foi encontrada, atribuí-se o valor -1 (menos um) ao diâmetro (real ou estimado).

A metodologia para calcular essa Similaridade é descrita na Equação 4-1. A Similaridade Média é descrita na Equação 4-2

$$f(D_{\text{real}}, D_{\text{estimado}}) = \begin{cases} 0 & \text{se } (D_{\text{real}} \neq -1 \wedge D_{\text{estimado}} = -1) \vee \\ (D_{\text{real}} = -1 \wedge D_{\text{estimado}} \neq -1) \\ \max\left(0, 1 - \frac{|D_{\text{real}} - D_{\text{estimado}}|}{D_{\text{real}}}\right) & \text{caso contrário} \end{cases} \quad (4-1)$$

$$\text{Similaridade Média (SM)} = \frac{1}{n} \sum_{i=1}^n f(D_{\text{real},i}, D_{\text{estimado},i}) \quad (4-2)$$

Onde:

- D_{real} representa o diâmetro real.
- D_{estimado} representa o diâmetro predito.
- A função $f(D_{\text{real}}, D_{\text{estimado}})$ retorna 0 se a condição em que um dos valores é -1 e o outro não for verdadeira.
- Caso contrário, calcula a diferença relativa entre D_{real} e D_{estimado} , garantindo que o resultado não seja negativo.

A Equação 4-1 permite calcular a diferença absoluta entre o diâmetro real e o estimado, em relação ao valor real. Para se quantificar a magnitude do erro na estimativa. É importante notar que a direção do erro, seja por superestimação ou subestimação, é considerada irrelevante. O cálculo finaliza subtraindo a magnitude do erro de 1 (um), fornecendo assim uma medida quantitativa da congruência entre os diâmetros estimado e real.

Esta subtração gera um índice de similaridade entre os diâmetros reais e os estimados, variando de 0 (zero) a 1 (um). A similaridade se aproxima de 1 à medida que o diâmetro estimado se alinha ao diâmetro real, medido manualmente. Ele será 1 quando ambos os valores forem iguais ou quando a pupila estiver ausente da imagem e nenhuma previsão for feita. Por outro lado, a similaridade será zero quando: (1) a pupila existir (D_{real} diferente de -1), mas não for detectada (D_{estimado} igual a -1), resultando em um falso negativo; (2) a pupila de fato não existir (D_{estimado} igual a -1), mas um diâmetro for estimado para ela (D_{estimado} diferente de -1), configurando um falso positivo; (3) ou quando a diferença entre o diâmetro real e o estimado em módulo subtraído da unidade produzir um valor negativo.

Após desenvolver um método para quantificar a similaridade entre medições, esse procedimento foi aplicado a cada uma das 2.000 imagens de teste do conjunto de dados. Nesse contexto, as previsões dos modelos foram avaliadas comparando-as com os diâmetros obtidos a partir de caixas delimitadoras anotadas manualmente.

Para avaliar o desempenho dos modelos de redes neurais em comparação com algoritmos clássicos na tarefa de segmentação da pupila, foram calculadas as seguintes métricas: (1) a Similaridade Média (SM), conforme descrito anteriormente; (2) o Erro Médio Absoluto (Mean Absolute Error - MAE); (3) a Raiz do Erro Quadrático Médio (Root Mean Square Error - RMSE); e (4) a acurácia de cada modelo para dois limiares de similaridade, 90% e 95%. Para que uma predição fosse considerada correta na primeira análise (limiar de 90%), a similaridade precisava ser, no mínimo, de 90%; na segunda análise (limiar de 95%), a similaridade deveria ser, no mínimo, de 95%.

Os resultados estão resumidos na Tabela 4.4. Essa tabela também apresenta a quantidade de falsos negativos (FN) — casos em que o modelo não detecta uma pupila presente na imagem, e a quantidade de falsos positivos (FP) — onde o modelo atribui erroneamente uma medição pupilar a imagens que não contêm pupila. Por fim, a coluna ‘Imprecisões’ quantificada o percentual de imprecisão de cada modelo, considerando o total de falsos negativos somados aos falsos positivos, divididos pelo número total de imagens avaliadas.

A Figura 4.11 apresenta o fluxograma dos procedimentos adotados para avaliar o desempenho de redes neurais na medição do diâmetro pupilar a partir de imagens infravermelhas do olho. Além disso, ilustra a extensão dessa comparação para incluir algoritmos clássicos de detecção pupilar que não empregam inteligência artificial. O objetivo principal é determinar o algoritmo mais eficiente para essa tarefa específica.

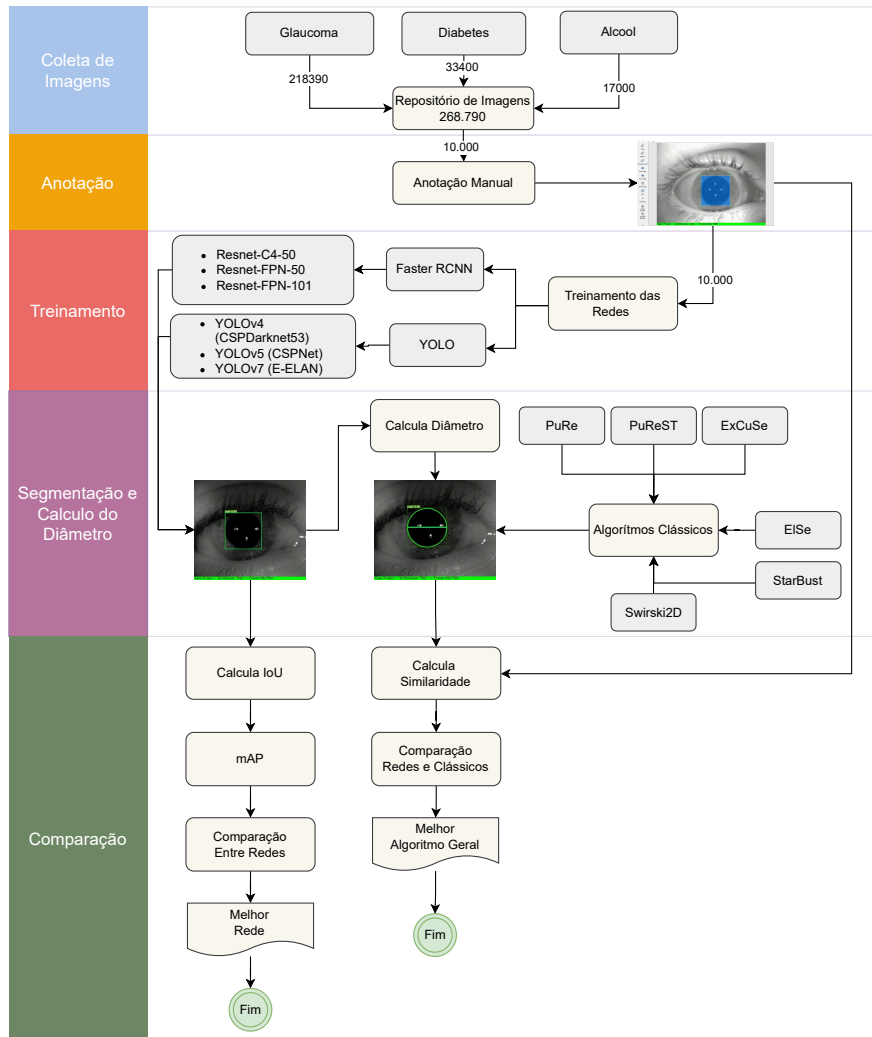


Figura 4.11: Fluxograma do processo para comparar e identificar o melhor algoritmo de medição do diâmetro pupilar em imagens infravermelho do olho.

4.4 Resultados

Nesta seção, apresenta-se os resultados e a análise da avaliação dos modelos de redes neurais para detecção de pupilas, incluindo uma avaliação comparativa de modelos contemporâneos de IA com os algoritmos clássicos tradicionalmente empregados na segmentação de pupilas.

4.4.1 Resultados da Comparação entre Redes Neurais

Após a realização do treinamento, teste e validação das redes, utilizando-se 6.000, 2.000 e 2.000 imagens respectivamente, observou-se que os modelos baseados na Faster R-CNN demandavam menos tempo para o treinamento. Especificamente, a configuração com Resnet-FPN-50 que foi treinada em cerca de 5 minutos com 6.000 imagens,

alcançando uma taxa de 17,8 quadros por segundo (FPS) no treinamento. Por outro lado, os modelos da família YOLO, típicos detectores de fase única, demonstraram uma velocidade superior durante a fase de inferências; a YOLOv5, por exemplo, processou 2.000 imagens em apenas dois minutos, atingindo uma taxa de 16,81 quadros por segundo (FPS) nas inferências. Os detalhes desses desempenhos estão sumarizados na Tabela 4.2.

Na avaliação da acurácia na detecção da região da pupila, como apresentado na Tabela 4.3, o YOLOv7 destacou-se entre todos os modelos testados, alcançando a maior precisão em cada limiar de Intersecção sobre União (IoU) e registrando a melhor Precisão Média (AP) de 94,14%.

Tabela 4.3: Desempenho de modelos baseados em redes neurais na detecção da região da pupila.

AP	Abordagens de Um Estágio			Abordagens de Dois Estágios		
	YOLOv4	YOLOv5	YOLOv7	Faster R-CNN C4 50	Faster R-CNN FPN 50	Faster R-CNN FPN 101
AP (@0,50) ¹	99,48%	99,43%	99,63%	99,22%	99,22%	99,06%
AP (@0,55)	99,48%	99,43%	99,58%	99,11%	99,22%	99,06%
AP (@0,60)	99,43%	99,43%	99,58%	99,06%	99,16%	99,01%
AP (@0,65)	99,32%	99,43%	99,53%	99,06%	99,06%	98,96%
AP (@0,70)	99,16%	99,32%	99,42%	99,00%	99,01%	98,96%
AP (@0,75)	98,95%	99,06%	99,27%	98,70%	98,85%	98,85%
AP (@0,80)	98,48%	98,90%	99,06%	97,56%	98,19%	98,74%
AP (@0,85)	97,50%	98,27%	98,52%	94,86%	95,90%	97,96%
AP (@0,90)	91,84%	93,62%	94,23%	81,42%	80,48%	88,71%
AP (@0,95)	42,50%	54,17%	52,58%	7,10%	6,99%	26,13%
AP ²	92,61%	94,11%	94,14% ³	87,51%	87,61%	89,60%

¹ AP@0,50 refere-se à Precisão Média (AP) calculada usando um limiar de Interseção sobre União (IoU) de 0,50. Da mesma forma, AP@0,55 indica o AP calculado com um limiar de IoU de 0,55. Este padrão continua para outros limiares de IoU especificados.

² Média de AP calculada com IoU começando de 0,5 até 0,95 e passo de 0,05.

³ Maior AP médio alcançado por modelos de redes neurais.

4.4.2 Resultado da Comparação entre Redes Neurais e Algoritmos Clássicos

Nesta sessão, serão apresentados os resultados obtidos através da aplicação da metodologia proposta, baseada em similaridade e descrita na Equação 4-1. Esta abordagem foi desenvolvida para superar as limitações dos algoritmos clássicos, que não conseguem fornecer as dimensões exatas da caixa delimitadora para a área pupilar.

Essa metodologia é preferível em comparação ao cálculo da precisão média (mAP) a partir das caixas delimitadoras porque se concentra em uma única medida crítica: o diâmetro pupilar. Enquanto a precisão média derivada das caixas delimitadoras leva em conta as variações entre a altura e a largura da caixa, a abordagem proposta se concentra especificamente no diâmetro, o qual é o alvo principal das medições. Esta técnica também

possibilita uma comparação direta entre modelos baseados em redes neurais e algoritmos clássicos que não utilizam inteligência artificial, avaliando-os conforme os diâmetros pupilares medidos.

Tabela 4.4: Comparação entre modelos baseados em redes neurais convolucionais e algoritmos clássicos de detecção pupilar aplicados em 2.000 imagens.

Modelo	Abord.	SM	MAE	RMSE	Acurácia @0,90	Acurácia @0,95	FN	FP	Imprecisões
PuRe	clássica	86,81%	14,86	34,52	80,00%	66,45%	0	85	4,25%
PuReST	clássica	83,89%	22,91	53,30	77,40%	56,25%	0	85	4,25%
ExCuSe	clássica	57,90%	62,47	95,32	50,85%	44,45%	724	3	36,35%
ElSe	clássica	62,83%	54,88	85,45	49,35%	39,20%	555	8	28,15%
StarBust	clássica	67,93%	48,57	78,03	38,10%	26,65%	50	53	5,15%
Swirski2D	clássica	78,48%	25,73	41,62	56,35%	40,90%	1	84	4,25%
Faster R-CNN50-C4	IA	95,72%	5,47	11,27	93,00%	73,90%	15	0	0,75%
Faster R-CNN50-FPN	IA	96,18%	4,89	10,78	96,05%	80,00%	15	0	0,75%
Faster R-CNN101-FPN	IA	95,36%	6,01	13,10	92,20%	69,65%	18	0	0,90%
YOLOv4	IA	97,51%	3,20	9,87	98,15%	92,90%	10	0	0,50%
YOLOv5	IA	97,97%	2,45	8,52	98,70%	95,55%	11	0	0,55%
YOLOv7	IA	98,20%	2,25	7,00	99,25%	96,90%	6	0	0,30%

*Os melhores desempenhos por abordagem estão destacados em negrito. Abordagem (Abord.), Inteligência Artificial (IA), Quantidade de Falsos Negativos (FN), Quantidade de Falsos Positivos (FP). A Similaridade Média (SM) foi calculada aplicando-se a equação 4-1. O Erro Médio Absoluto (Mean Absolute Error - MAE) aplicando-se a equação 2-1. A Raiz do Erro Quadrático Médio (Root Mean Square Error - RMSE) aplicando-se a equação 2-2. As acurácias com limiares de @90 e @95 foram calculadas considerando como corretas as predições com SM mínimo de 90% e 95% respectivamente. O percentual de imprecisões foi calculado como a soma dos falsos positivos e falsos negativos, dividida pelo total de imagens de teste e multiplicada por 100. Todos os índices foram calculados sobre as 2000 imagens de teste.

Os resultados detalhados na Tabela 4.4 mostram que todos os modelos baseados em IA superaram o PuRe, o qual é o algoritmo de melhor desempenho dentre os algoritmos não baseados em inteligência artificial. Também mostra que o YOLOv7 foi a abordagem com a maior taxa Similaridade Média (98,19%), menor Erro Absoluto Médio (2,25 píxeis), menor Raiz do Erro Quadrático Médio (7,00 píxeis), maior Acurácia com limiar de 0,90 (99,25%), maior Acurácia com limiar de 0,95 (96,90%) e o menor percentual de Imprecisões (0,30%). Deve-se notar que os algoritmos clássicos foram utilizados para medir o diâmetro da pupila usando seus parâmetros padrão, conforme descrito na literatura. Para aprimorar o desempenho de cada um desses algoritmos, seria necessário testá-los com as possíveis combinações desses parâmetros.

4.5 Discussão

Existem outras bases de dados com imagens oculares disponíveis para experimentos; contudo, este trabalho disponibiliza para download uma base de dados contendo indivíduos que consumiram diferentes níveis de álcool, pacientes com diabetes mellitus tipo II e pacientes com glaucoma. Pesquisadores podem utilizar esta base de dados para desenvolver métodos precisos para aplicações médicas.

Além disso, este estudo utilizou o conjunto de dados para o treinamento, teste e avaliação tanto de algoritmos de detecção de objetos gerais quanto de algoritmos especificamente adaptados para a detecção de pupilas. Os resultados indicam que o modelo Faster R-CNN tem a menor duração de treinamento, com Resnet-FPN-50 liderando em velocidade de treinamento. Por outro lado, os modelos da série YOLO se destacaram em velocidade durante a etapa de inferência, com YOLOv5 registrando a taxa de inferência mais rápida no conjunto de testes.

A análise comparativa da Precisão Média (AP) na Tabela 4.3, que avalia a correlação entre o diâmetro da pupila previsto e as anotações manuais, identificou o modelo YOLOv7 como o mais eficaz na detecção pupilar. Este modelo alcançou a mais alta precisão na localização das caixas delimitadoras das pupilas, registrando uma Precisão Média de 94,14%, o que evidencia sua acurácia próxima à das anotações humanas.

Adicionalmente, a análise na Tabela 4.4 revela que o modelo YOLOv7 se destaca entre todos os modelos (clássicos e baseados em IA) em termos de similaridade média, atingindo 98,19%. Esse resultado foi corroborado tanto pelo cálculo do erro médio absoluto quanto pela raiz do erro médio quadrático. A similaridade média foi determinada comparando-se os diâmetros das caixas delimitadoras previstas com os das caixas anotadas manualmente. Esses achados ressaltam a precisão e a confiabilidade do YOLOv7 na medição dos diâmetros pupilares.

Os experimentos demonstram que os modelos de IA são mais eficazes que os métodos não baseados em IA e que o re-treinamento desses modelos com 10.000 imagens para treinamento, validação e teste é adequado para alcançar alta precisão na detecção de pupilas.

Indicam também que, uma vez re-treinadas com imagens de pupilas, as redes neurais de detecção de objetos de propósito geral superam algoritmos especializados não baseados em redes neurais. Assim, em ambientes nos quais não haja limitação de *hardware*, as redes neurais são a melhor opção para tarefas que exijam a localização e a medição precisa de pupilas.

4.6 Considerações Finais do Capítulo

Esse capítulo apresentou a metodologia de criação de um banco de imagens anotadas para treinamento de redes neurais para encontrar e medir o diâmetro pupilar nas imagens oriundas da pupilometria. Ele também mostra como foi realizada a avaliação comparativa entre redes neurais e algoritmos clássicos aplicados à segmentação pupilar. Essa comparação destacou a superioridade das redes neurais, especialmente a YOLOv7, na detecção precisa da pupila e na medição de seu diâmetro.

Avaliação de Técnicas de Aprendizado de Máquina

Este capítulo apresenta a aplicação de múltiplas técnicas de aprendizado de máquina, acompanhadas de uma análise comparativa entre elas. As quais incluem processos como filtragem, balanceamento de amostras, extração e seleção das melhores características, visando melhorar a precisão na classificação aplicados em oferecer suporte ao diagnóstico do glaucoma.

5.1 Contextualização

O glaucoma engloba um grupo de neuropatias ópticas caracterizadas pela degeneração progressiva das células ganglionares da retina (RGCs) e seus axônios, resultando em perda do campo visual [Gracitelli et al. 2015, Weinreb e Khaw 2004, Weinreb, Aung e Medeiros 2014, Tatham et al. 2014]. A perda da função visual devido ao glaucoma é geralmente um processo irreversível e, se não tratado adequadamente, pode progredir para deficiência visual e até cegueira [Gracitelli et al. 2015, Weinreb e Khaw 2004, Weinreb, Aung e Medeiros 2014, Tatham et al. 2014]. Além disso, o tratamento do glaucoma torna-se mais complexo, caro e desafiador quanto mais avançada é a doença. Portanto, a detecção e diagnóstico precoces do glaucoma são essenciais para prevenir sua progressão e minimizar seus danos [Weinreb e Khaw 2004, Tatham et al. 2014]. Os testes mais frequentemente usados para diagnosticar glaucoma incluem:

- **Tonometria:** Procedimento padrão para medir a Pressão Intraocular (PIO), sendo um indicador importante na avaliação do risco de glaucoma, especialmente quando elevada [Eddy, Sanders e Eddy 1983]. A tonometria tem a desvantagem de ser desconfortável para alguns pacientes devido à necessidade de aplicar colírios anestésicos e ao contato do tonômetro com a córnea.
- **Campimetria:** Ferramenta diagnóstica que mapeia o campo visual do paciente, essencial para identificar regiões de perda visual potencialmente causadas pelo glau-

coma [Wu e Medeiros 2018]. A desvantagem deste teste é sua subjetividade, pois depende da resposta do paciente à percepção de estímulos visuais. Além disso, sua eficácia é limitada em crianças e idosos e só detecta danos significativos ao nervo óptico, visto que até 40% das células ganglionares da retina podem ser perdidas antes que quaisquer mudanças no campo visual possam ser detectadas [Yaqub 2012].

- **Fundoscopia (Oftalmoscopia):** Este exame permite uma inspeção detalhada da retina, da mácula, do disco ótico e dos vasos sanguíneos do fundo do olho [Hoyt, Frisen e Newman 1973]. Pode ser realizado com um oftalmoscópio direto ou indireto, retinografia ou OCT [Bussel, Wollstein e Schuman 2014] (Optical Coherence Tomography - Tomografia de Coerência Óptica), considerada a técnica de referência para a detecção estrutural do glaucoma. No entanto, esses exames requerem equipamento especializado e manejo por um oftalmologista, o que pode não estar disponível ou acessível para uma parte da população.

Os pesquisadores [Chang et al. 2013, Najjar et al. 2018, Duque-Chica et al. 2018, Sarezky et al. 2014, Sarezky et al. 2016, Park et al. 2019, Martucci et al. 2014, Tatham et al. 2014, Lawlor et al. 2017, Charalel, Lin e Singh 2014, Pillai et al. 2019, Tatham et al. 2014, Gracitelli et al. 2015, Kankipati, Girkin e Gamlin 2011, Pradhan et al. 2017, Rukmini et al. 2015, Carle et al. 2014, Lee et al. 2019] mostraram que a resposta pupilar à luz em indivíduos com glaucoma difere da resposta pupilar de pessoas saudáveis e, portanto, pode ser usada como um indicador dessa patologia. A pupilometria cromática apresenta-se como uma alternativa adicional na detecção do glaucoma, expandindo as opções de exames disponíveis para diagnóstico dessa condição [Arévalo-López et al. 2023, Bayraktar et al. 2023]. Consiste em avaliar os movimentos de contração e dilatação da pupila quando exposta a estímulos luminosos com luz visível de diferentes cores por determinados períodos.

A vantagem de se analisar o reflexo pupilar à luz é que pode ser realizado de maneira prática e não invasiva. No entanto, as taxas de sucesso na detecção do glaucoma primário de ângulo aberto em estágio inicial ainda não se tornaram altas o suficiente para inserir a pupilometria cromática na prática clínica.

Algoritmos de aprendizado de máquina podem reconhecer padrões por meio da análise de dados, por isso, têm sido amplamente utilizados para apoiar aplicações médicas em geral para diagnosticar e rastrear diversas patologias. Nesse contexto, eles também podem extrair informações de dados obtidos por meio da pupilometria cromática. Entretanto, ainda há necessidade de mais pesquisas que aprimorem a eficácia de algoritmos que usam a pupilometria e a análise do reflexo pupilar à luz para triar ou diagnosticar o glaucoma, especialmente na detecção do glaucoma de ângulo primário em seu estágio inicial.

Vários trabalhos publicados nos últimos cinco anos fazem apenas análises e comparações estatísticas para mostrar a capacidade do PLR de discriminar entre indivíduos saudáveis e glaucomatosos [Kankipati, Girkin e Gamlin 2011, Pradhan et al. 2017, Rukmini et al. 2015, Carle et al. 2014, Lee et al. 2019, Chang et al. 2013, Gracitelli et al. 2015, Adhikari et al. 2016, Pattan et al. 2023, Najjar et al. 2018, Najjar et al. 2023].

Os estudos [Wu et al. 2019, Gaddipati e Sivaswamy 2021, Talaat et al. 2019, An et al. 2018] aplicaram técnicas de aprendizado de máquina para diagnosticar o glaucoma, porém usando informações extraídas da segmentação de imagens de fundo de olho.

O trabalho de Quan *et al.* [Quan et al. 2023] realizou uma prova de conceito usando o PLR para detectar o glaucoma. Sua análise foi limitada, no entanto, a testar estatisticamente a significância das características usadas no teste.

Esse trabalho difere dos demais, pois traz uma avaliação de técnicas de aprendizado de máquina em dados da pupilometria cromática, que retratam o reflexo pupilar à luz dos voluntários (PLR). Nesse estudo estão sendo investigados métodos de identificação do glaucoma usando técnicas para (1) filtrar e remover os ruídos do sinal pupilar, (2) extrair características relevantes, (3) fazer o balanceamento das amostras, (4) realizar seleção das melhores características e (5) buscou-se definir a escolha apropriada de classificadores que possam aprimorar a análise da reação pupilar à luz no processo de triagem e diagnóstico do glaucoma.

5.2 Materiais e Métodos

O conjunto de dados empregado para avaliar as várias técnicas de aprendizado de máquina e determinar sua eficácia em auxiliar o diagnóstico do glaucoma compreende exames de indivíduos sem glaucoma (grupo controle) assim como de pacientes em diferentes graus de glaucoma. Os cenários de diagnósticos foram categorizados em classificações binárias e multi-classe.

Na abordagem de classificação binária, os dados foram agrupados nos seguintes grupos de interesse:

1. Controle vs. Patológico: Esta classificação foca em distinguir o grupo controle do grupo patológico sem considerar a severidade da patologia.
2. Controle vs. Estágio Inicial: nesse agrupamento, o objetivo é diferenciar especificamente o grupo controle do grupo de pacientes em estágio inicial de glaucoma.
3. Controle vs. Estágio Moderado: Nesta categoria, visamos identificar diferenças entre o grupo controle e aqueles com pacientes com glaucoma moderado.
4. Controle vs. Estágio Severo: Esta classificação é dedicada a contrastar o grupo controle de indivíduos no estágio de glaucoma severo.

Os agrupamentos foram projetados para refinar o entendimento e a detecção do glaucoma em várias etapas de sua progressão.

Na classificação multi-classe, os dados foram categorizados para discriminar simultaneamente as classes:

5. Controle vs. Inicial vs. Moderado vs. Severo: Esta classificação busca diferenciar todos os rótulos possíveis na base de dados de pacientes em uma única análise. Ela determina se um paciente é saudável ou patológico e, caso seja patológico, identifica o grau específico da patologia.
6. Inicial vs. Moderado vs. Severo: Esta abordagem é projetada para cenários onde já está estabelecido que o paciente tem glaucoma, mas o grau da patologia ainda precisa ser definido.

Após o estabelecimento desses grupos de interesse de diagnóstico, implementaram-se as técnicas, incluindo filtragem, extração de características, balanceamento de dados e seleção de características. Em seguida, avaliou-se a eficácia de vários classificadores. A Figura 5.1 apresenta simplificada o fluxo de avaliação das técnicas, que serão melhor detalhadas nas próximas seções, para determinar a técnica mais eficaz para cada um dos grupos de interesse.

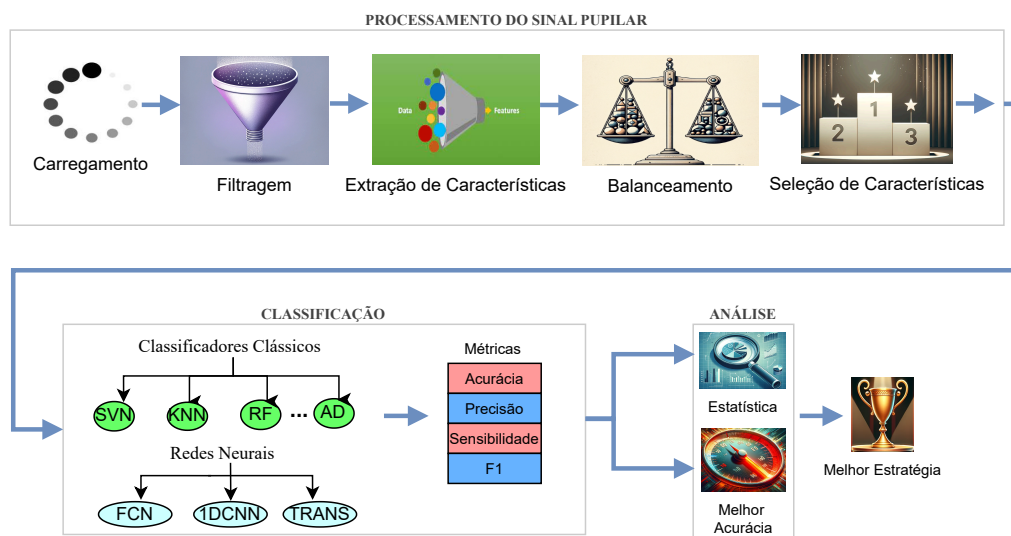


Figura 5.1: Diagrama da avaliação proposta dividida em três etapas principais: processamento do sinal pupilar, classificação dos dados e análise dos resultados.

5.2.1 A Base de Dados de Exames Pupilar

O conjunto de dados elaborado para esta pesquisa é composto por 512 vídeos, cada um com 4 minutos e 5 segundos de duração, que registram o reflexo pupilar em

resposta à luz de 250 voluntários. Dentre os participantes, 113 são saudáveis e 137 possuem algum grau de glaucoma. Alguns voluntários contribuíram com mais de uma gravação, as quais foram realizadas seguindo ao protocolo descrito na Seção 5.2.2. Os critérios de elegibilidade para os voluntários incluíam ter acuidade visual melhor que 20/100 em ambos os olhos, não ter passado por cirurgia ocular nos últimos três meses e ser maior de 18 anos. Os participantes também foram solicitados a abster-se de consumo de cafeína pelo menos uma hora antes da gravação.

Este estudo recebeu aprovação do conselho de ética do Hospital de Urgências de Goiânia (HUGO), sob o número de aconselhamento técnico 5.990.785, registrado na plataforma Brasil.

Durante o exame, os participantes se sentaram confortavelmente em uma sala completamente escura. Foram equipados com o pupilômetro e instruídos a focar o olhar em um ponto minimamente luminoso dentro do equipamento usado como referência visual, sem provocar estimulação nas pupilas. Os participantes do grupo glaucomatoso foram diagnosticados com glaucoma de ângulo aberto, a forma mais comum do glaucoma e alvo dessa análise. A distribuição por idade e gênero dos voluntários está detalhada na Tabela 5.1, que também traz na coluna PDB o percentual de desbalanceamento entre cada um dos grupos Glaucomatosos e o grupo Controle, calculados conforme a Equação 5-1

$$PDB(\%) = \left(\frac{\text{Vídeos Glaucomatoso} - \text{Vídeos Controle}}{\text{Vídeos Controle}} \right) \times 100 \quad (5-1)$$

Tabela 5.1: Distribuição dos exames disponíveis na base de dados de vídeos por, grupo, gênero e idade dos voluntários.

DISTRIBUIÇÃO DOS EXAMES					
Grupos	N	Vídeos	PDB (%)	Idade ($\mu \pm \sigma$)	Masculino
Controle	113	243	-	42 \pm 13	49 (43%)
Glaucomatoso	Inicial	108	10,7	57 \pm 12	62 (57%)
	Moderado	21	84,4	54 \pm 14	10 (48%)
	Severo	8	94,2	61 \pm 8	4 (50%)
	Sub. Total	137	270	-11,1	57 \pm 14
Total	250	512	-	53 \pm 8	125 (50%)

Número de Voluntários (N). Percentual de Desbalanceamento (PDB) em relação ao grupo Controle.

5.2.2 Protocolo de Gravação dos Vídeos

O protocolo de gravação detalha o processo de gravação dos vídeos para análise das reações pupilares à luz. Especifica o tempo de adaptação inicial ao ambiente escuro; a duração, intensidade e cor de cada estímulo luminoso; a quantidade de estímulos

aplicados, o intervalo de adaptação entre estímulos, qual dos olhos será estimulado e qual será gravado.

O protocolo escolhido inclui um período de adaptação ao escuro inicial de 10 minutos, seguindo as diretrizes compiladas por Pinheiro H. *et al.* [Pinheiro e Costa 2021]. Foram aplicados quatro períodos de estimulação usando LEDs cromáticos: o primeiro estímulo com comprimento de onda de 623 nm (vermelho), o segundo com luz no comprimento de onda de 466 nm (azul), o terceiro aplicado com luz no comprimento de onda de 517 nm (verde) e o último realizado com luz branca.

A seleção das cores a serem usadas para estimular a resposta pupilar seguiu o recomendado por Rukmini *et al.* [Rukmini, Milea e Gooley 2019] que indica o uso de luz vermelha e azul, e o recomendado por Crippa [Crippa, Domellöf e Kawasaki 2018] que sugere o uso de luz verde. Ambos os estudos destacam a importância da pupilometria cromática na avaliação da saúde dos fotorreceptores no nervo óptico e da retina para vários tipos de doenças.

A duração, intensidade e o intervalo de adaptação entre cada estímulo foram definidos em 1 segundo, com uma intensidade de 250 cd/m², conforme estabelecido por Park [Park *et al.* 2011]. O período de adaptação entre estímulos foi definido em 59 segundos, consoante as recomendações de Gracitelli *et al.* [Gracitelli *et al.* 2014].

Após o período de adaptação inicial, iniciou-se a gravação do vídeo. Cinco segundos após o início da gravação, foi iniciado o primeiro estímulo. Cada ciclo de repouso e estimulação pupilar constitui um trecho do sinal, resultando em quatro fases distintas de gravação, conforme ilustrado na Figura 5.2. A duração total de cada gravação foi de 4 minutos e 5 segundos, capturada a uma taxa de 30 quadros por segundo, totalizando 7.350 quadros.

Quanto ao olho submetido à estimulação, ela pode ser aplicada no mesmo olho que está sendo gravado (avaliação de reflexo direto) ou ao outro olho (avaliação de reflexo consensual). Neste estudo, ambos os reflexos foram utilizados nas gravações.

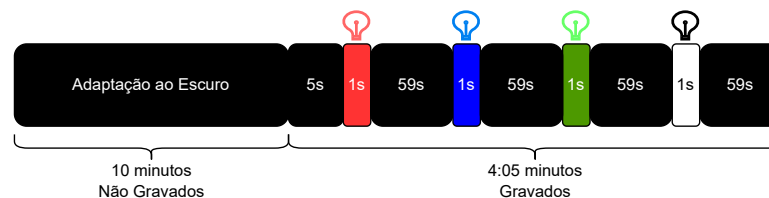


Figura 5.2: Ilustração do protocolo estabelecido, abrangendo a fase inicial de adaptação pupilar, o espectro de cores, a duração da estimulação cromática e os intervalos especificados para adaptação entre cada estímulo.

5.2.3 Dispositivo de Gravação

O pupilômetro utilizado neste estudo foi desenvolvido inicialmente por Pinheiro H. *et al.* [Pinheiro et al. 2015] e posteriormente, aprimorado por Silva *et al.* [Silva et al. 2017] para realizar a pupilometria cromática. Essa técnica permite o uso de estímulos luminosos em várias cores, não se restringindo apenas à luz branca.

Este dispositivo foi escolhido por atender a requisitos como permitir a estimulação cromática usando LEDs RGB, possibilitar que a duração e a intensidade da estimulação sejam personalizáveis, impedir que a luminosidade externa interfira no reflexo pupilar durante o período de gravação, fornecer iluminação infravermelha para gravação, que é invisível ao olho humano e, portanto, não impacta a estimulação pupilar; oferecer a flexibilidade de gravar e estimular qualquer um dos olhos, para capturar o reflexo direto ou o consensual.

5.2.4 Segmentação e Pré-processamento

Em processamento de imagem digital e visão computacional, a segmentação da imagem é a sua divisão em múltiplos segmentos ou regiões de interesse da imagem (Region of Interest - ROI). O objetivo principal da segmentação é simplificar a representação da imagem, tornando-a mais significativa e fácil de analisar [Stockman e Shapiro 2001].

Para a análise do Reflexo Pupilar à Luz, é essencial extrair o sinal pupilar dos exames, localizando a pupila e medindo o seu diâmetro em cada quadro (*frame*) do período de gravação. Isso se deve ao fato de o indicativo da patologia não estar presente diretamente em uma imagem, como ocorrem, por exemplo, nos exames de raio-x. O indicativo da patologia está no comportamento da pupila, variando seu diâmetro ao longo dos diversos quadros do período de gravação.

Para fazer a medição do diâmetro pupilar quadro a quadro (segmentação) e a composição do sinal pupilar, foi utilizada a rede neural convolucional detectora de objetos YOLOv7 [Wang, Bochkovskiy e Liao 2023], retreinada com um conjunto de 10.000 imagens da pupila manualmente anotadas.

Após o processo de retreinamento, a rede neural exibiu aprimorada proficiência na detecção, e em demarcar a caixa delimitadora da pupila em cada quadro do vídeo. As dimensões da caixa delimitadora possibilitaram o cálculo do diâmetro em píxeis, pois a altura da caixa representa o diâmetro vertical e a largura o diâmetro horizontal. Baseado na abordagem recomendada por Zandi *et al.* [Zandi et al. 2021], o maior dos dois diâmetros foi designado como a medição definitiva do diâmetro da pupila. Dessa forma foi possível medir o diâmetro pupilar em cada quadro e compor o sinal que traz o comportamento dinâmico e as flutuações de diâmetro da pupila ao longo do tempo. A Figura 5.3 exibe um exemplo de sinal pupilar de um voluntário.

Há várias fases envolvidas na condução de um exame pupilar. Primeiro, o vídeo do voluntário é gravado. Em seguida, o vídeo é analisado para localizar a pupila e medir seu diâmetro em cada quadro. Depois, uma curva ou sinal é criado para representar o comportamento pupilar. Feito isso, o reflexo pupilar pode ser avaliado. O processo de avaliar o reflexo envolve pré-processamento, filtragem, extração de características, seleção de características, balanceamento e classificação de sinal.

Nesta pesquisa, a fase de pré-processamento envolveu a exclusão de 24 segundos da etapa pós-estimulação e re-dilatação em cada segmento, para minimizar possíveis redundâncias no sinal.

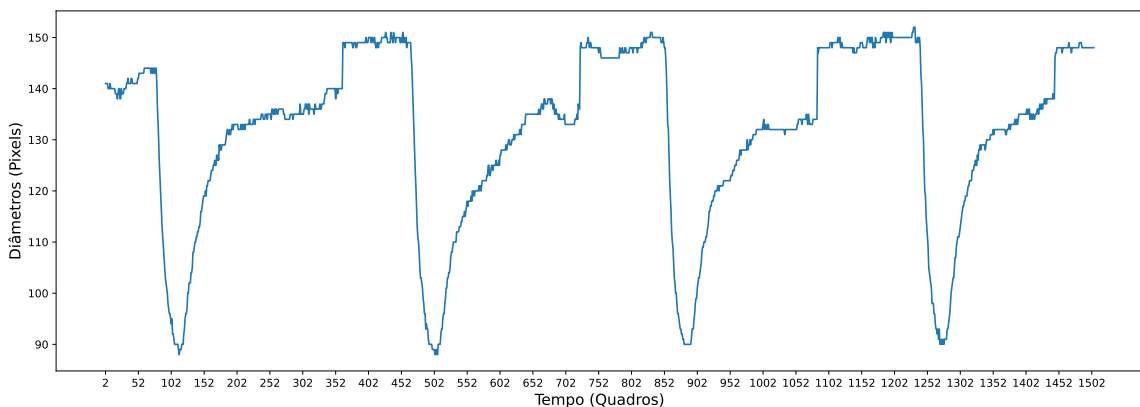


Figura 5.3: Exemplo da curva do sinal pupilar segmentado, após a remoção dos períodos redundantes.

Ainda na fase de pré-processamento, tem-se que, a padronização de um conjunto de dados é um requisito comum para muitos classificadores: eles podem apresentar baixo desempenho se as características individuais não estiverem ajustadas para uma escala comum. Por isso, a classificação foi feita usando-se a normalização `MinMaxScaler` da biblioteca `scikit-learn`.

O `MinMaxScaler` normaliza as características ajustando-as para um intervalo entre zero e um, descrito pela Equação 5-2. Este método subtrai o valor mínimo da característica e, em seguida, divide pelo intervalo entre os valores máximo e mínimo da mesma. Uma vantagem do `MinMaxScaler` é que ele preserva a forma da distribuição original das características, sem reduzir a influência de valores extremos.

$$X_{\text{std}} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (5-2)$$

5.2.5 Filtragem

O sinal pupilar pode ser afetado por ruídos devido a piscadas ou mudanças na direção do olhar dos voluntários. Apesar das instruções para manter o foco em um ponto específico no pupilômetro e evitar piscar, ainda foram observadas instâncias de piscadas e desvios de olhar.

Neste estudo, foram exploradas três abordagens distintas de filtragem. A primeira abordagem, denominada Sem Filtragem (SF), consiste na análise direta do sinal sem qualquer modificação. A segunda, chamada Filtragem Customizada (FC), emprega um algoritmo desenvolvido especificamente para esta pesquisa, que remove leituras anômalas que ultrapassam os limites normais da fisiologia pupilar e suaviza variações bruscas no sinal. A terceira abordagem, Filtragem Passa-Baixa (PB), foca em atenuar componentes de ruído de alta frequência presentes no sinal. Esses três métodos foram avaliados para verificar se contribuem para aumentar a precisão e a confiabilidade na classificação dos sinais.

O filtro personalizado desenvolvido nesta pesquisa para manejar mudanças abruptas no sinal que excedem um limiar pré-estabelecido funciona da seguinte forma: inicialmente, o limiar é definido em um valor baixo, especificamente 3. Todas as alterações no sinal que ultrapassem esse limiar são eliminadas, a menos que a sua remoção resulte em uma descontinuidade no sinal. Uma descontinuidade é identificada quando há a necessidade de eliminar 20 quadros consecutivos. Se ocorrer uma descontinuidade, o limiar é aumentado em uma unidade e o processo é repetido. Este ciclo continua até que o sinal resultante esteja livre de mudanças abruptas e descontinuidades. Detalhes adicionais sobre esta abordagem podem ser encontrados no pseudo-código no Apêndice B.

Por fim, o filtro Passa-Baixa (PB) da biblioteca SciPy em Python foi avaliado. Este filtro permite a passagem de sinais com frequências inferiores a uma determinada frequência de corte, enquanto atenua, reduzindo a amplitude dos sinais com frequências superiores a essa frequência. Ele foi implementado utilizando a função `filtfilt`, que aplica o filtro do início ao fim e do fim ao início do sinal, eliminando as altas frequências sem causar deslocamentos de fase.

O filtro foi ajustado empiricamente, com uma frequência de corte de 0,6, a qual foi normalizada pela frequência de Nyquist (Nyq) Seção 2.5.1, calculada como 0,5 vezes a taxa de quadros da gravação (30 fps). O filtro foi configurado com ordem 3 para promover atenuação das frequências indesejadas sem comprometer a integridade do sinal.

5.2.6 Extração de Características

A extração de características é um processo que transforma os dados de entrada em um subespaço mais compacto, mantendo a maioria das informações cruciais do

conjunto de dados original [Khalid, Khalil e Nasreen 2014]. Esta técnica é capaz de melhorar significativamente o desempenho de algoritmos de classificação, especialmente quando comparada ao uso direto dos dados brutos.

Neste estudo, adotaram-se duas abordagens distintas quanto à extração de características. Inicialmente, o conjunto completo de diâmetros pupilares foi avaliado sem a realização de qualquer extração. Posteriormente, foram extraídas dez características reconhecidas na literatura científica, aplicadas individualmente às quatro fases de estimulação do sinal, resultando em um total de 40 características. Adicionalmente, mais duas características foram extraídas do sinal completo, elevando o número total de características analisadas para 42.

As características extraídas de cada seção de estimulação do sinal são:

1. Diâmetro Inicial (DI): O diâmetro médio da pupila em repouso antes do estímulo luminoso, tipicamente o diâmetro máximo. Calculado usando a média de vinte quadros 0,3 segundos antes da estimulação.
2. Máxima Contração (MC): O menor diâmetro observado na seção, que ocorre geralmente durante a aplicação do estímulo. A contração máxima é calculada identificando-se o menor diâmetro pós-estímulo, excluindo valores abaixo de 40 píxeis.
3. Tempo até a Máxima Contração (TMC): A duração necessária para alcançar a máxima contração após o início do estímulo. É determinada pelo quantidade de quadros até que ocorra o menor diâmetro.
4. Amplitude de Contração Absoluta (ACA): A diferença entre o maior diâmetro (diâmetro inicial) e o menor (máxima contração). Calculada subtraindo-se DI menos MC.
5. Amplitude de Contração Relativa (ACR): A razão entre o menor e o maior diâmetro. Calculado dividindo-se a CM pelo DI.
6. Velocidade de Contração (VC): Razão da amplitude pelo tempo para a máxima contração. Calculado dividindo-se ACA por TMC.
7. Latência (LAT): O tempo entre o início do estímulo e o início efetivo da contração. Determinado pela primeira instância em que o diâmetro reduz 5% do DI.
8. Diâmetro de Re-dilatação após 6 segundos (DR6s): Tamanho da pupila 6 segundos após o estímulo. Calculado a partir do primeiro diâmetro válido em uma janela de tolerância de 10 diâmetros pós-estímulo; foi atribuído -1 quando nenhum valor válido foi encontrado dentro da janela de tolerância.
9. Tempo de Re-dilatação (TRD): Tempo para a pupila retornar a 80% do diâmetro inicial pós-estímulo. Calculado encontrando a posição do quadro que excede 80% do DI.

10. Complexidade Fractal de Higuchi (CFH): Uma medida não linear de dimensão fractal em séries temporais. Calculada ao longo de 20 quadros, 0,3 segundos antes da estimulação, descartando valores inválidos e analisando até 5 escalas de séries temporais (kmax). Ngo *et al.* [Ngo et al. 2021] apresenta mais detalhes sobre a complexidade fractal de Higuchi e sua significância na diferenciação entre pacientes controle e glaucomatosos.

Segue a descrição das características aplicadas ao sinal completo:

11. Diâmetro no Final do Exame (DFE): representa o tamanho da pupila no final do exame. Calculado como a média dos últimos 10 diâmetros de cada curva, descartando diâmetros menores que 40 píxeis.
12. Complexidade Fractal de Higuchi no Sinal (CFHS): é a complexidade de Higuchi, neste caso, aplicada em toda a curva pupilar. Calculado considerando a curva inteira, desconsiderando diâmetros impossíveis e, como na análise por seção, definindo o parâmetro K-max = 5.

Com a extração de características do sinal pupilar concluída, os próximos passos são realizar o balanceamento das amostras e selecionar as características mais eficazes para análises futuras.

5.2.7 Balanceamento

Um desbalanceamento em um conjunto de dados de duas classes ocorre quando a classe minoritária está significativamente sub-representada em comparação com a classe majoritária. Nas análises realizadas, foi definido que há desbalanceamento entre as classes sempre que a diferença nas proporções excedem 10%. Portanto, o balanceamento é considerado desnecessário se a diferença percentual do quantitativo de amostras entre as classes não ultrapassar esse limite.

Nas avaliações onde foram identificados desequilíbrio entre as classes, foram utilizadas técnicas de subamostragem e sobre amostragem para equilibrar a quantidade de amostras. A subamostragem consiste na remoção aleatória de amostras da classe majoritária até que seu tamanho seja equivalente ao da classe minoritária. Em contraste, a sobre amostragem aumenta a classe minoritária adicionando amostras sintéticas, criadas a partir das amostras existentes vizinhas, até que as classes atinjam tamanhos equivalentes. Para a sobre amostragem, foi empregado o algoritmo SMOTE (Synthetic Minority Over-sampling Technique), que é um método de aumento de dados que gera novas instâncias considerando a proximidade entre as amostras existentes [He e Ma 2013], nessa pesquisa o número de vizinhos considerados foi $k = 5$.

5.2.8 Seleção de Características

Seleção de características (SC) é um processo voltado a identificar características pertinentes enquanto elimina dados irrelevantes, redundantes ou ruidosos. Características irrelevantes falham em contribuir com informações significativas, enquanto características redundantes não oferecem percepções adicionais além do que já é obtido das características selecionadas [Kumar e Minz 2014]. O propósito da seleção de características é tríplice: (1) melhorar a precisão da classificação e prevenir o *overfitting*, (2) desenvolver modelos mais enxutos, e (3) aumentar a interpretabilidade dos modelos para o entendimento humano.

Métodos de seleção de características dividem-se em três categorias principais: métodos de filtro, de *wrapper* e embutidos. Métodos de filtro, como o SelectKBest, utilizam critérios gerais como correlação para descartar características irrelevantes independentemente de algoritmos de aprendizado de máquina. Métodos de *wrapper* utilizam classificadores para avaliar o desempenho e escolher a combinação ótima de características. Métodos embutidos integram a seleção de características diretamente no processo de aprendizado de máquina.

Nessa pesquisa, utilizou-se um método de cada categoria, inicialmente o método SelectKBest, que é um método baseado em filtro que avalia a relação individual entre cada característica e a variável de saída usando um teste estatístico específico (neste estudo análise de variância - ANOVA) para selecionar os melhores k atributos.

Posteriormente, empregou-se a Eliminação Recursiva de Características (Recursive Feature Elimination - RFE), uma técnica baseada em *wrapper*. O método RFE é uma abordagem iterativa que começa com todas as características e remove progressivamente as menos significativas com base no desempenho do modelo. O *Random Forest* foi usado para fazer as eliminações até que o número alvo de características seja alcançado.

Por fim, o método Lasso (Least Absolute Shrinkage and Selection Operator) com validação cruzada foi empregado. Trata-se uma regressão linear, descrita na Equação 5-3, que impõe uma penalidade correspondente à soma dos valores absolutos dos coeficientes de regressão, conhecida como norma L1. A função objetivo, que o Lasso busca minimizar, envolve y , o vetor de resposta; X , a matriz de dados contendo n observações; β , que são os coeficientes do modelo; e λ , o parâmetro de regularização que determina a intensidade da penalidade aplicada aos coeficientes.

$$\text{Minimizar } \frac{1}{2n} \|y - X\beta\|^2 + \lambda \|\beta\|_1 \quad (5-3)$$

O processo de seleção de características foi elaborado para identificar as vinte melhores características, explicitamente implementado definido para os métodos SelectKBest e RFE. Ao contrário destes, o seletor LassoCV foi configurado com um nível de tole-

rância de 0,01, permitindo-lhe seleccionar um número flexível de características. Os vinte melhores coeficientes foram então usados no segundo passo para seleccionar as 20 melhores características. Após a conclusão da seleção de características dos dados do sinal pupilar, a fase subsequente envolve classificar o sinal, completo ou fatiado.

5.2.9 Fatiamento do Sinal

No contexto do processamento de sinais, fatiar o sinal refere-se ao processo de seleccionar e extrair um segmento específico de um sinal maior. É semelhante a cortar uma imagem, mas em vez de seleccionar uma parte de uma imagem, selecciona-se uma parte de uma série temporal ou forma de onda. Ele pode ser usado para isolar um intervalo de tempo específico ou para remover partes indesejadas do sinal para análise ou processamento. Nesta pesquisa, dividiu-se o sinal em quatro segmentos correspondentes à cor da estimulação: vermelho, azul, verde e branco, conforme definido no protocolo de gravação.

Esta abordagem foi adotada para reduzir o número total de características, funcionando como uma forma de seleção de características. Como resultado, quando não foi realizada a extração de características, o fatiamento as reduziu de 1500 para 375. Quando foi realizada a extração de características, o fatiamento as reduziu de 40 para 10. Nessa análise, as características que se referem ao sinal completo, como DEE e HFCS foram deixadas de fora.

5.2.10 Classificadores

Para avaliar a contribuição das técnicas de filtragem, extração, balanceamento e seleção na classificação do glaucoma, foram utilizados os seguintes classificadores:

(A) Classificador Linear:

Análise Discriminante Linear (Linear Discriminant Analysis - LDA): Identifica uma combinação linear de características que melhor diferencia duas ou mais classes.

(B) Classificador Baseado em Vizinhos:

K-Vizinhos Mais Próximos (k-Nearest Neighbors - KNN): Classifica uma entrada considerando a classe majoritária de seus vizinhos mais próximos, onde 'k' indica o número de vizinhos considerados na análise.

(C) Classificadores Baseados em Árvores:

Árvore de Decisão (Decision Tree - DT): Utiliza uma estrutura de árvore de decisão para classificação. O objetivo principal ao dividir um nó em uma árvore de decisão é criar subgrupos mais homogêneos ou 'puros' com relação à variável de resposta, ou seja, que pertençam à mesma classe.

Floresta Aleatória (Random Forest - RF): Utiliza um conjunto de árvores de decisão. Árvores Extra (Extra Tree - ET): São uma variação do método de florestas aleatórias. Neste modelo, as divisões dos nós das árvores são feitas de maneira aleatória, diferentemente das florestas aleatórias tradicionais que selecionam a melhor divisão com base na maximização da pureza.

(D) Classificadores Baseados em Boosting:

AdaBoost (AB): Integra múltiplos classificadores ‘fracos’ para aumentar a força da classificação, ajusta os novos modelos para corrigir os erros dos modelos anteriores.

Máquina de Incremento Gradiente (Gradiente Boosting Machine - GBM): Assim como o AdaBoost constrói o modelo final de forma incremental adicionando modelos fracos, geralmente árvores de decisão de pequena profundidade, uma de cada vez. Melhora a classificação ajustando-se ao gradiente negativo da função de perda (daí o nome “Gradient Boosting”) dos modelos anteriores.

(E) Classificador Baseado em Probabilidade:

Naive Bayes (NB): Aplica o teorema de Bayes com a suposição de independência entre as características.

(F) Classificador Baseado em Máquina de Vetores de Suporte:

Máquina de Vetores de Suporte (Support Vector Machine - SVM): Utiliza funções de kernel (linear, polinomial, RBF, sigmoid, etc.) para identificar fronteiras ótimas de separação entre as classes.

(G) Classificadores Baseados em Redes Neurais:

Rede Neural Totalmente Conectada (Fully Connected Neural Network- FCN): É composta por neurônios em cada camada totalmente conectados a todos os neurônios na camada subsequente. É conhecida como densa devido às conexões completas entre as camadas.

Rede Neural Convolutiva Unidimensional (One Dimensional Convolutional Neural Network - 1D-CNN): Projetadas para dados com uma estrutura de grade, como séries temporais (1D). Podem ser efetivas na identificação de padrões locais em conjuntos de dados como áudio, séries temporais financeiras e sequenciamento genético.

Rede Neural Transformers (TRANS): Introduzida por Vaswani *et al.* em “Attention is All You Need” (2017), essa rede usa o mecanismo de “autoatenção”, que permite a cada parte de uma sequência interagir com todas as outras partes, capturando inter-relações complexas. Permite avaliar a relevância de diferentes palavras ou características no contexto de uma palavra, ou característica específica.

Os classificadores tradicionais empregados neste estudo foram obtidos da biblioteca Python scikit-learn. As redes neurais artificiais, por sua vez, foram desenvolvidas

utilizando o framework TensorFlow. A representação arquitetônica das redes neurais mencionadas estão disponíveis no Apêndice C.

5.2.11 Classificação

O estudo envolveu uma avaliação detalhada que classificou seis diferentes abordagens de agrupamento de dados, detalhadas na seção 5.2, três métodos de filtragem (sem filtragem, filtrando com um filtro personalizado e com um filtro passa-baixa), duas estratégias de balanceamento (sobre amostragem e subamostragem) e duas estratégias de extração de características (com e sem extração de características). Resultando em 72 cenários de classificação. Para cada cenário, quatro técnicas distintas de seleção de características foram examinadas: sem seleção de características, SelectKBest, RFE e LassoCV.

Esta análise abrangente foi realizada para identificar padrões nos dados de pupilometria e determinar quais abordagens oferecem a maior precisão na classificação do glaucoma.

A Tabela 5.2 apresenta os principais parâmetros dos classificadores avaliados na pesquisa, as opções de cada parâmetro e a indicação da configuração que foi efetivamente utilizada.

5.2.12 Validação

A validação cruzada (Cross Validation - CV) utiliza o método de partição *K-fold* para avaliar a generalização de modelos de aprendizado de máquina. Nela o conjunto de dados é dividido em K subconjuntos aproximadamente iguais e mutuamente exclusivos (dobras). Essas dobras são então utilizadas em K rodadas de treinamento e avaliação do modelo. Em cada rodada, $K-1$ subconjuntos são usados para treinamento, enquanto o subconjunto restante é usado para testar o modelo. Este ciclo é repetido até que cada subconjunto tenha sido o conjunto de teste uma vez. Após a conclusão dessas iterações, os resultados do desempenho de cada dobra são agregados para produzir uma medida média da eficácia do modelo [Kohavi et al. 1995].

Neste estudo, optou-se pela validação cruzada com $K = 5$, um valor que proporciona um bom equilíbrio entre a variabilidade das análises e a economia de recursos computacionais, em comparação com valores maiores de K .

Dado que há múltiplos vídeos de um mesmo voluntário, para evitar o vazamento de informações durante a validação cruzada, adotou-se a estratégia de validação cruzada estratificada. Essa abordagem assegura que todos os vídeos de um mesmo voluntário sejam exclusivamente alocados ou no grupo de treinamento, ou no grupo de testes, mantendo assim a integridade dos dados nas diferentes fases de validação.

Tabela 5.2: Principais parâmetros usados pelos classificadores no processo de classificação do sinal pupilar.

PARÂMETROS DOS CLASSIFICADORES	
Classificador	Parâmetros
LDA	solver: {'svd', 'lsqr', 'eigen'}, usado='svd' n_components: int, usado=None
KNN	n_neighbors: int, usado=5, weights: {'uniform', 'distance'}, usado='uniform', algorithm: {'auto', 'ball_tree', 'kd_tree', 'brute'}, usado='auto', metric: string, usado='minkowski'
DT	criterion: {'gini', 'entropy'}, usado='gini', splitter: {'best', 'random'}, usado='best', max_depth: int, usado=None
RF	n_estimators: int, usado=100, criterion: {'gini', 'entropy'}, usado='gini', max_depth: int, usado=None
ET	n_estimators: int, usado=100, criterion: {'gini', 'entropy'}, usado='gini', max_depth: int, usado=None
AB	base_estimator: object, usado=None, n_estimators: int, usado=50 learning_rate: float, usado=1.0
GBM	loss: {'deviance', 'exponential'}, usado='deviance' learning_rate: float, usado=0.1, n_estimators: int, usado=100
NB	-
SVM	C: float, usado=1.0, kernel: {'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'}, usado='rbf', degree: int, usado=3, gamma: {'scale', 'auto'}, usado='scale'
FCN	epochs=100, batch_size=8, dropout=0.25, callbacks=early_stop
1D-CNN	epochs=100, batch_size=8, filters=64, dropout=0.25 callbacks=early_stop
TRANS	epochs=100, batch_size=8 num_heads=4, head_size=64 ff_dim=32, num_transformer_blocks=4 mlp_units={64} dropout=0.25, mlp_dropout=0.4, callbacks=early_stop

Análise Discriminante Linear (Linear Discriminant Analysis - LDA). K-Vizinhos Mais Próximos (k-Nearest Neighbors - KNN). Árvore de Decisão (Decision Tree - DT). Floresta Aleatória (Random Forest - RF). Árvores Extra (Extra Tree - ET). AdaBoost (AB). Máquina de Incremento Gradiente (Gradient Boosting Machine - GBM). Naive Bayes (NB). Máquina de Vetores de Suporte (Support Vector Machine - SVM). Rede Neural Totalmente Conectada (Fully Connected Neural Network - FCN). Rede Neural Convolutacional Unidimensional (One Dimensional Convolutional Neural Network - 1D-CNN). Rede Neural Transformer (TRANS).

5.2.13 Avaliação dos Resultados

Para discernir o impacto de diferentes metodologias nos resultados de classificação e identificar a abordagem mais eficaz, realizou-se duas avaliações para cada classificador: uma avaliação estatística e uma análise baseada na precisão ótima.

5.2.14 Avaliação Estatística

Após completar a fase de classificação, foi realizada uma análise estatística dos resultados obtidos pelos classificadores, essa análise foi feita buscando determinar se os

resultados da classificação utilizando determinada abordagem era superior aos demais. Inicialmente, para determinar a normalidade das distribuições de resultados, utilizou-se o teste de Shapiro-Wilk, eficaz para conjuntos de dados pequenos como os que estão sendo usados, em detrimento do teste de Kolmogorov-Smirnov, que é mais adequado para conjuntos de dados maiores [Razali, Wah et al. 2011]. Quando os resultados dos classificadores seguiram uma distribuição normal, aplicou-se o teste t de Student pareado [Efron e Hastie 2021] para análise. Por outro lado, para resultados distribuídos de forma não normal, foi utilizado o teste de Wilcoxon [Conover 1999]. Ambos com um nível de significância de 0,05.

5.2.15 Avaliação pela Melhor Acurácia

Embora a análise estatística ofereça uma visão geral do desempenho médio de diversos classificadores, ela pode não captar inteiramente a eficácia de uma abordagem específica que se destaque significativamente, especialmente se os outros classificadores, na mesma abordagem, não alcançarem bons resultados.

Por isso, complementou-se a análise estatística com uma avaliação exclusivamente pela melhor acurácia, que destaca o classificador que alcançou a melhor acurácia para a abordagem avaliada.

Essa análise dupla permite avaliar a consistência dos vários classificadores, e destacar os que tem melhor desempenho.

5.2.16 Métricas de Avaliação

Para fazer a avaliação, foram consideradas as métricas: Acurácia, que mostra a proporção de casos corretamente identificados do total de casos examinados. Sensibilidade (Taxa de Verdadeiros Positivos - TPR), que quantifica a proporção de casos positivos reais (indivíduos com glaucoma) corretamente identificados como tal. Especificidade (Taxa de Verdadeiros Negativos - TNR), que mede a porcentagem de casos negativos verdadeiros (pacientes controle saudáveis) corretamente identificados. E a Pontuação F1, que é uma métrica que combina sensibilidade e precisão em uma única medida, representando a média harmônica de sensibilidade e especificidade.

5.3 Resultados

Nesse estudo, realizou-se uma análise exploratória do sinal pupilar, abordando (1) três técnicas distintas de filtragem - sem filtragem, filtragem com um filtro personalizado e usando um filtro passa-baixa; (2) dois métodos de balanceamento de dados - *undersampling* e *oversampling*; (3) duas abordagens para extração de características - sem extração de características e extração de quarenta e duas características mencionadas na literatura; e (4) quatro estratégias para seleção de características - sem seleção de características, e usando SelectKbest, RFE e o LassoCV.

Em seguida, os dados foram processados usando nove classificadores convencionais: SVM, KNN, DT, RF, ET, LDA, NB, AB e o GBM. Adicionalmente, os dados foram submetidos a três classificadores avançados baseados em arquiteturas de rede neural, incluindo uma FCN, uma 1D-CNN e uma rede Transformer.

Os resultados da classificação foram analisados estatisticamente, considerando as métricas de desempenho dos classificadores, e também por meio da avaliação pela melhor acurácia de cada classificador. A Figura 5.4 traz o fluxo detalhado de como foi feita a avaliação das abordagens de aprendizado de máquina aplicadas aos dados da pupilometria.

Nas subseções seguintes, detalharemos os resultados da análise estatística, enfatizando os efeitos e a eficácia da filtragem de sinal, extração de características, balanceamento de dados e seleção de características.

5.3.1 Resultado da Análise Estatística do Sinal Completo

5.3.1.1 Análise Estatística das Técnicas de Filtragem

O sinal pupilar foi classificado empregando-se três metodologias distintas de filtragem: (1) sem qualquer filtragem do sinal pupilar, (2) usando um filtro customizado projetado para eliminar mudanças abruptas no sinal pupilar, e (3) aplicando um filtro passa-baixa ao sinal pupilar.

A Tabela 5.3 apresenta uma análise comparativa dessas estratégias de filtragem. Ela destaca os casos em que cada uma das três abordagens de filtragem foi significativamente equivalente ou superior às outras, considerando os seis grupos de dados, as técnicas de balanceamento de dados de *undersampling* e *oversampling*. A comparação é feita para cenários com e sem extração de características, considerando técnicas de seleção de características. Ao final da tabela tem-se uma sumarização que mostra que das 96 combinações possíveis em 49 delas os melhores resultados vieram sem se fazer filtragem, 19 vezes com a filtragem customizada, 20 vezes com a filtragem passa-baixa e 9 vezes as abordagens de filtragem foram equivalentes.

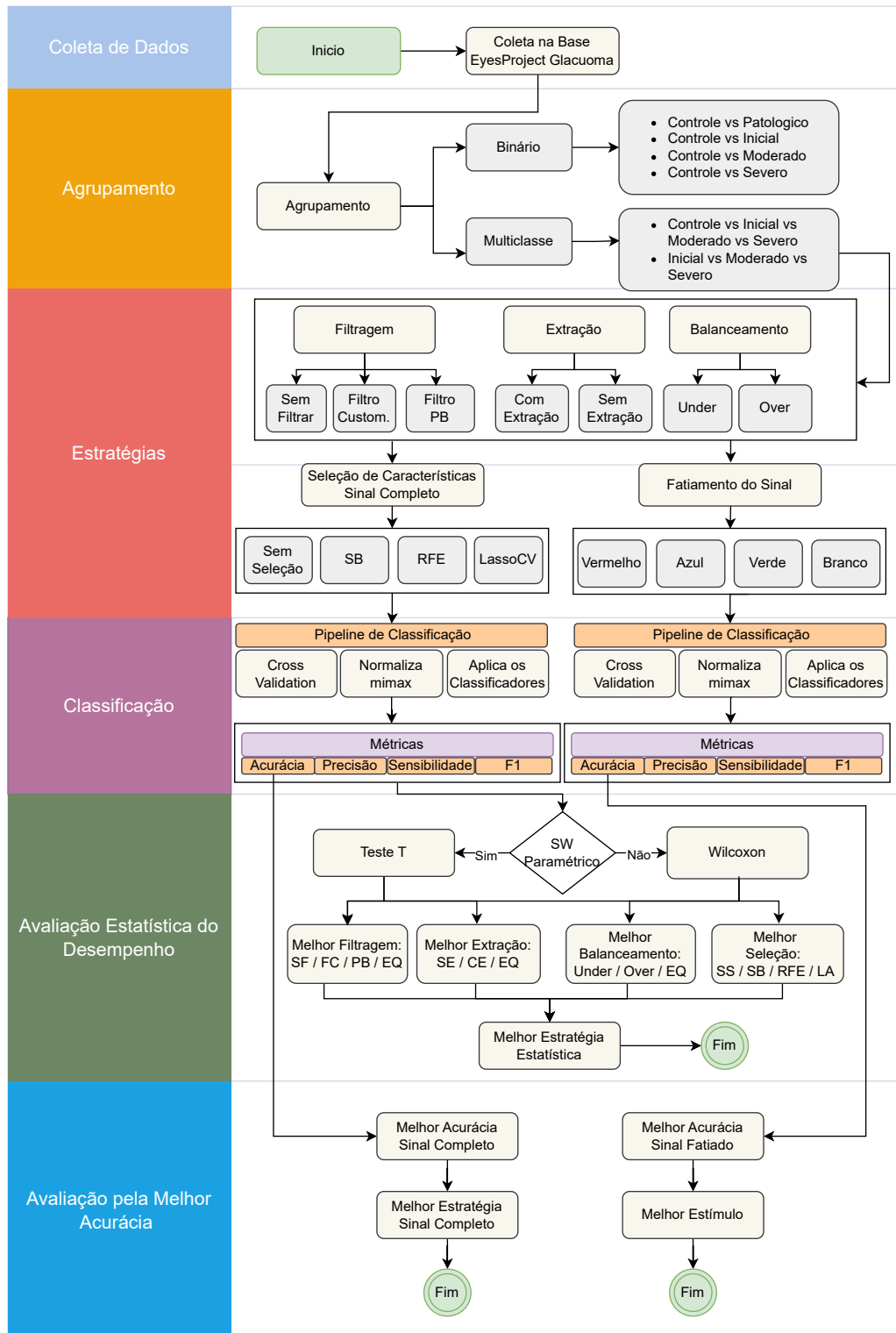


Figura 5.4: Fluxograma da aplicação das técnicas de aprendizado de máquina.

5.3.1.2 Análise Estatística da Eficácia da Extração de Características

Essa parte da análise se concentrou em determinar a abordagem mais eficaz para a classificação do sinal pupilar: se realizar a extração de características específicas,

Tabela 5.3: Análise estatística das metodologias de filtragem destacando a abordagem mais eficaz ou identificando os equivalentes para cada agrupamento.

FILTRAGEM										
Agrupamento	Balanceamento	Extração	SS	SB	RFE	LA	Sumarização (Freq)			
			Melhor Filtragem				SF	FC	PB	EQ
Controle vs Patológico	Under	SE	EQ	SF	SF	SF	3	0	0	1
		CE	SF	SF	SF	EQ	3	0	0	1
	Over	SE	EQ	SF	SF	FC	2	1	0	1
		CE	SF	FC	SF	SF	3	1	0	0
SubTotal:							11	2	0	3
Controle vs Inicial	Under	SE	SF	SF	SF	SF	4	0	0	0
		CE	FC	EQ	EQ	EQ	0	1	0	3
	Over	SE	SF	SF	SF	SF	4	0	0	0
		CE	SF	SF	SF	FC	3	1	0	0
SubTotal:							11	2	0	3
Controle vs Moderado	Under	SE	SF	SF	FC	FC	2	2	0	0
		CE	EQ	SF	SF	PB	2	0	1	1
	Over	SE	SF	FC	SF	PB	2	1	1	0
		CE	PB	PB	PB	PB	4	0	0	0
SubTotal:							10	3	2	1
Control vs Severe	Under	SE	SF	SF	SF	FC	3	1	0	0
		CE	FC	FC	FC	FC	0	4	0	0
	Over	SE	SF	SF	SF	SF	4	0	0	0
		CE	PB	PB	PB	PB	0	0	4	0
SubTotal:							7	5	4	0
Controle vs Inicial vs Moderado vs Severo	Under	SE	FC	FC	EQ	FC	0	3	0	1
		CE	SF	SF	SF	SF	4	0	0	0
	Over	SE	PB	PB	PB	PB	4	0	0	0
		CE	SF	SF	PB	PB	2	2	0	0
SubTotal:							10	5	0	1
Inicial vs Moderado vs Severo	Under	SE	SF	SF	SF	SF	4	0	0	0
		CE	FC	FC	SF	FC	1	3	0	0
	Over	SE	SF	SF	SF	PB	3	1	0	0
		CE	PB	PB	PB	PB	0	0	4	0
SubTotal:							8	4	4	0
Total:							49	18	20	9

Subamostragem (Undersampling - Under), Sobre amostragem (Oversampling - Over), Sem Extração (SE), Com Extração (CE), Sem Seleção (SS), SelectKBest (SB), Eliminação Recursiva de Características (Recursive Feature Elimination - RFE), Seleção de Características LassoCV (LA), Sem Filtragem (SF), Filtragem Customizada (FC), Filtro Passa Baixa (PB), Estatisticamente Equivalentes (EQ), Frequência da Melhor Estratégia (Freq).

identificadas na literatura, do sinal pupilar melhora o desempenho da classificação ou se fornecer aos classificadores o sinal completo dos diâmetros pupilares sem uma síntese prévia de dados se mostra mais benéfico.

A Tabela 5.4 fornece um desdobramento estatístico, baseado em métricas de classificação, das instâncias onde utilizar o sinal completo com todos os diâmetros foi equivalente ou mais eficaz do que realizar a extração de características do sinal. A sumarização apresentada no final da tabela aponta que de um total de 144 combinações possíveis, 27 vezes as técnicas foram equivalentes, 53 vezes os melhores resultados vieram sem se realizar a extração e 64 vezes extrair as características trouxe melhor resultado.

Tabela 5.4: Análise estatística das metodologias de extração de características, destacando a abordagem mais eficaz ou identificando os equivalentes para cada agrupamento.

EXTRAÇÃO								
Agrupamento	Balanceamento	Filtragem	SS	SB	RFE	LA	Sumarização	
			Melhor Extração				Estratégia	Freq
Controle vs Patológico	Under	SF	CE	SE	SE	SE	SE	7
		FC	EQ	EQ	SE	SE	CE	2
		PB	CE	EQ	SE	SE	EQ	3
	Over	NF	CE	SE	SE	SE	SE	7
		FC	EQ	EQ	SE	SE	CE	2
		LP	CE	EQ	SE	SE	EQ	3
Controle vs Inicial	Under	NF	CE	EQ	SE	SE	SE	7
		FC	CE	SE	SE	SE	CE	4
		LP	CE	CE	SE	SE	EQ	1
	Over	NF	SE	SE	SE	SE	SE	7
		FC	EQ	EQ	SE	SE	CE	3
		LP	EQ	CE	SE	CE	EQ	2
Control vs Moderate	Under	NF	CE	EQ	CE	EQ	SE	1
		FC	EQ	EQ	SE	EQ	CE	6
		LP	CE	CE	CE	CE	EQ	5
	Over	NF	SE	CE	SE	CE	SE	6
		FC	EQ	CE	SE	SE	CE	5
		LP	SE	CE	CE	SE	EQ	1
Control vs Severe	Under	NF	CE	CE	EQ	CE	SE	0
		FC	CE	CE	CE	CE	CE	10
		LP	CE	CE	CE	EQ	EQ	2
	Over	NF	SE	CE	SE	SE	SE	3
		FC	CE	CE	CE	CE	CE	8
		LP	CE	CE	EQ	CE	EQ	1
Controle vs Inicial vs Moderado vs Severo	Under	NF	EQ	SE	SE	SE	SE	7
		FC	CE	CE	EQ	EQ	CE	2
		LP	SE	SE	SE	SE	EQ	3
	Over	NF	EQ	CE	SE	EQ	SE	1
		FC	CE	CE	CE	CE	CE	7
		LP	EQ	CE	EQ	CE	EQ	4
Inicial vs Moderado vs Severo	Under	NF	SE	EQ	SE	SE	SE	5
		FC	CE	CE	CE	CE	CE	5
		LP	EQ	CE	SE	SE	EQ	2
	Over	NF	SE	CE	SE	CE	SE	2
		FC	CE	CE	CE	CE	CE	10
		LP	CE	CE	CE	CE	EQ	0
Total:							SE	53
							CE	64
							EQ	27

Subamostragem (Undersampling - Under), Sobre amostragem (Oversampling - Over), Sem Filtragem (SF), Filtragem Customizada (FC), Filtragem Passa Baixa (PB), Sem Seleção de Características (SS), Seleção de Características SelectKBest (SB), Eliminação Recursiva de Características (Recursive Feature Elimination - RFE), Seleção de Características LassoCV (LA), Abordagem Com Extração de características é a melhor (CE), Abordagem Sem Extração é a melhor (SE), As Abordagens são Estatisticamente Equivalentes (EQ), Frequencia da melhor estratégia (Freq).

5.3.1.3 Análise Estatística das Técnicas de Balanceamento

Essa subseção da pesquisa teve como objetivo explorar o impacto e a influência do balanceamento das amostras nos resultados da classificação.

A Tabela 5.5 apresenta uma comparação detalhando as instâncias nas quais as técnicas de *oversampling* e *undersampling*, aplicadas a dados com e sem extração de características e combinadas com vários métodos de seleção de características, provaram ser equivalentes ou superiores. A sumarização ao final da tabela mostra que das 144

combinações possíveis de cruzamento de agrupamentos e das técnicas em 122 delas o balanceamento por sobre amostragem foi superior, em 5 delas a subamostragem venceu e 17 vezes as abordagens foram estatisticamente equivalentes.

Tabela 5.5: Avaliação estatística das técnicas de balanceamento por sobre amostragem e subamostragem, destacando a abordagem mais eficaz ou identificando os equivalentes para cada agrupamento.

BALANCEAMENTO								
Agrupamento	Extração	Filtragem	SS	SB	RFE	LA	Sumarização	
			Melhor Balanceamento				Estratégia	Freq
Controle vs Patológico	SE	SF	Over	EQ	Over	Under	Over	4
		FC	Over	EQ	EQ	EQ	Under	1
		PB	Over	Over	Over	EQ	EQ	7
	CE	SF	EQ	Over	EQ	EQ	Over	4
		FC	Over	Over	Over	Under	Under	2
		PB	EQ	EQ	Over	Under	EQ	5
Controle vs Inicial	SE	SF	Over	Over	Over	Under	Over	8
		FC	Over	Over	Over	EQ	Under	1
		PB	Over	EQ	Over	EQ	EQ	3
	CE	SF	EQ	Over	Over	Over	Over	7
		FC	Over	Over	EQ	Under	Under	1
		PB	Over	Over	EQ	EQ	EQ	4
Controle vs Moderado	SE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	0
	CE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	0
Controle vs Severo	SE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	0
	CE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	0
Controle vs Inicial vs Moderado vs Severo	SE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	0
	CE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	3
Inicial vs Moderado vs Severo	SE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	0
	CE	SF	Over	Over	Over	Over	Over	12
		FC	Over	Over	Over	Over	Under	0
		PB	Over	Over	Over	Over	EQ	0
Total:							Over	122
							Under	5
							EQ	17

Sem Extração (SE), Com Extração (CE), Sem filtragem (SF), Filtragem Customizada (FC), Filtragem Passa Baixa (PB), Subamostragem (Undersampling - Under), Sobre amostragem (Oversampling - Over), Sem Seleção de Características (SS), Seleção de Características SelectKBest (SB), Eliminação Recursiva de Características (Recursive Feature Elimination - RFE), Seleção de Características LassoCV (LA), Estatisticamente Equivalentes (EQ), Frequência da melhor estratégia (Freq).

5.3.1.4 Análise Estatística da Seleção de Características

Nesta análise, nosso objetivo foi entender a influência da Seleção de Características no processo de classificação. Especificamente, pretendia-se determinar se realizar a seleção de características antes da classificação melhora os resultados e, em caso afirmativo, quais técnicas demonstram desempenho superior.

A Tabela 5.6 detalha as instâncias nas quais uma técnica específica de seleção de características superou as outras, considerando ambos os métodos de balanceamento de *undersampling* e *oversampling* e em contextos com ou sem extração de características. A sumarização ao final da tabela mostra que das 72 combinações possíveis, em 37 delas a seleção RFE foi superior, 20 vezes a seleção LA foi superior, 6 vezes a seleção SelectKBest foi melhor e 9 vezes o melhor resultado foi obtido sem fazer a seleção de características.

Tabela 5.6: Análise da técnica de seleção estatística de características destacando a abordagem mais eficaz para cada agrupamento. (Parte 1).

SELEÇÃO DE CARACTERÍSTICAS (Parte 1)								
Agrupamento	Filter	Extração	Filtragem	Melhor Extração de Características	Sumarização (Freq)			
					SS	SB	RFE	LA
Controle vs Patológico	Under	SE	SF	LA	0	0	0	3
			FC	LA				
	PB		LA					
	CE	SF	RFE	0	0	2	1	
FC		RFE						
PB	FC	LA	0	0	3	0		
	PB	RFE						
Controle vs Inicial	Under	SE	SF	RFE	0	0	2	1
			FC	LA				
	PB		RFE					
	CE	SF	RFE	0	0	3	0	
FC		RFE						
PB	FC	RFE	0	1	1	1		
	PB	RFE						
Controle vs Moderado	Under	SE	SF	LA	0	0	1	2
			FC	RFE				
	PB		RFE					
	CE	SF	RFE	0	0	1	2	
FC		LA						
PB	FC	LA	1	0	0	2		
	PB	LA						
Controle vs Moderado	Over	SE	SF	SS	1	0	0	2
			FC	LA				
	PB		LA					
	CE	SF	LA	0	0	2	1	
FC		RFE						
PB	FC	RFE	0	0	2	1		
	PB	RFE						

A tabela continua...

Tabela 5.6: Análise da técnica de seleção estatística de características destacando a abordagem mais eficaz para cada agrupamento. (Parte 2).

SELEÇÃO DE CARACTERÍSTICAS (Parte 2)								
Agrupamento	Balanceamento	Extração	Filtragem	Melhor Extração de Características	Sumarização (Freq)			
					SS	SB	RFE	LA
Controle vs Severo	Under	SE	SF	RFE	0	0	3	0
			FC	RFE				
	PB	RFE						
	CE	SF	RFE	1	0	2	0	
FC	RFE							
PB	SS	RFE						
Controle vs Severo	Over	SE	SF	RFE	0	0	1	2
			FC	LA				
	PB	LA						
	CE	SF	SS	1	0	1	1	
FC	LA							
PB	RFE	SS						
Controle vs Inicial vs Moderado vs Severo	Under	SE	SF	RFE	0	1	2	0
			FC	RFE				
	PB	SB						
	CE	SF	RFE	0	2	1	0	
FC	SB							
PB	SB	RFE						
Controle vs Inicial vs Moderado vs Severo	Over	SE	SF	RFE	0	0	0	3
			FC	RFE				
	PB	RFE						
	CE	SF	SS	3	0	0	0	
FC	SS							
PB	SS	SS						
Inicial vs Moderado vs Severo	Under	SE	SF	RFE	0	0	3	0
			FC	RFE				
	PB	RFE						
	CE	SF	RFE	0	1	2	0	
FC	RFE							
PB	SB	RFE						
Inicial vs Moderado vs Severo	Over	SE	SF	RFE	1	0	2	0
			FC	RFE				
	PB	SS						
	CE	SF	RFE	2	0	1	0	
FC	SS							
PB	SS	SS						
Total					9	6	37	20

Subamostragem (Undersampling - Under), Sobre amostragem (Oversampling - Over), Com Extração (CE), Sem Extração (SE), Sem Filtragem (SF), Filtragem Customizada (FC), Filtragem Passa Baixa (PB), Sem Seleção de Características (SS), Seleção de Características SelectKBest (SB), Eliminação Recursiva de Características (Recursive Feature Elimination - RFE), Seleção de Características LassoCV (LA), Frequencia da melhor estratégia (Freq.).

5.3.2 Análise pela Melhor Acurácia

Avaliar o classificador de melhor desempenho pode destacar um classificador que tenha um desempenho destacado dentro da metodologia específica. Analisar o melhor classificador e identificar a abordagem que propiciou obter a melhor classificação pode reforçar as conclusões obtidas a partir da análise estatística.

5.3.2.1 Avaliação do Sinal Completo pela Melhor Acurácia

Nessa etapa, foi verificado para cada agrupamento qual a melhor acurácia percentual de classificação e qual abordagem levou à melhor a ela considerando o sinal completo.

A Tabela 5.7 apresenta a melhor acurácia percentual de classificação alcançada, juntamente com a abordagem específica que contribuiu para esse resultado ótimo. Nela é

possível ver que os melhores resultados foram alcançados sem se realizar a filtragem e sem se realizar a extração de características em todos os seis grupos de diagnósticos. Quanto à seleção, os resultados foram mistos, mas ficou evidente que não compensa realizar a seleção de características SelectkBest. Quanto ao balanceamento, o Over foi geralmente melhor, especialmente para os agrupamentos com maior desbalanceamento. Quanto aos classificadores, os resultados foram variados, sendo o LDA apontado como mais eficiente para o grupo Controle vs Patológico.

Tabela 5.7: Melhores acurácias alcançadas por classificadores agrupados por interesse e abordagem, utilizando o sinal completo.

MELHORES ACURÁCIAS - SINAL COMPLETO						
Agrupamento	Abordagem					Melhor Acur (%)
	Filtragem	Extração	Seleção	Balanceamento	Classificador	
Controle vs Patológico	SF	SE	LA	Under	LDA	73,90
Controle vs Inicial	SF	SE	LA	Under	NB	72,97
Controle vs Moderado	SF	SE	SS	Over	SVM	98,10
Controle vs Severo	SF	SE	RFE	Over	RF	98,73
Controle vs Inicial vs Moderado vs Severo	SF	SE	SS	Over	ET	79,21
Inicial vs Moderado vs Severo	SF	SE	SS	Over	ET	92,21

Sem filtragem (SF), Sem Extração (SE), Sem Seleção de Características (SS), Eliminação Recursiva de Características (Recursive Feature Elimination - RFE), Seleção de Características LassoCV (LA), Subamostragem (Undersampling - Under), Sobre amostragem (Oversampling - Over), Análise Discriminante Linear (Linear Discriminant Analysis - LDA), Bayes Ingênuo (Naive Bayes - NB), Máquina de Vetores de Suporte (Support Vector Machine (SVM), Floresta Aleatória (Random Forest - RF), Árvores Extra (Extra Tree - ET), Acurácia (Acur).

5.3.2.2 Avaliação do Sinal após o Fatiamento pela Melhor Acurácia

Esta análise segmenta o sinal em quatro partes distintas, cada uma correspondendo a um período de estimulação, para determinar se uma seção específica do sinal é mais proeminente para fins de classificação. Ela também busca identificar a cor de estimulação mais eficaz para a triagem do glaucoma. Considerando a metodologia de classificação aplicada, os resultados para cada segmento são apresentados na Tabela 5.8. Nela é possível ver que o segmento azul se mostrou o melhor em três dos 6 grupos avaliados e as abordagens sem filtragem e sem extração continuaram a ser as melhores para a maioria dos grupos.

Tabela 5.8: Melhores acurácias de classificação alcançadas para cada agrupamento por interesse e abordagem, utilizando um segmento de sinal associado a uma cor de estimulação.

MELHOR ACURÁCIA - SINAL SEGMENTADO						
Agrupamento	Abordagem					Melhor Acur (%)
	Filtragem	Extração	Segmento	Balanceamento	Classificador	
Controle vs Patológico	PB	CE	Branca	Over	KNN	64,88
Controle vs Inicial	PB	SE	Azul	Over	ET	65,76
Controle vs Moderado	SF	SE	Verde	Over	ET	94,93
Controle vs Severo	SF	SE	Azul	Over	ET	97,04
Controle vs Inicial vs Moderado vs Severo	SF	SE	Vermelha	Over	GBM	76,31
Inicial vs Moderado vs Severo	SF	SE	Azul	Over	ET	92,68

Sem filtragem (SF), Filtro Passa Baixa (PB), Sem Extração (SE), Com Extração (CE), Sobreamostragem (Over - Oversampling), K-Vizinhos Mais Próximos (k-Nearest Neighbors - KNN), Máquina de Incremento Gradiente (Gradient Boosting Machine - GBM), Árvores Extras (Extra Tree - ET), Acurácia (Acur).

5.3.3 Implementando a Técnica de Classificação Mais Promissora

Após avaliar vários métodos de aprendizado de máquina, realizou-se novamente a classificação do grupo de dados Controle vs. Patológico, utilizando a abordagem identificada como a mais promissora segundo a análise estatística e segundo a análise pela melhor acurácia.

O agrupamento de dados Controle vs. Patológico foi selecionado para essa análise final por ser o mais geral dos agrupamentos, o qual efetivamente distingue entre indivíduos saudáveis e aqueles com qualquer grau de glaucoma. Embora esta abordagem não informe o grau específico do glaucoma, ela está alinhada visando auxiliar a triagem do glaucoma, fase em que a graduação detalhada da doença ainda é menos crítica.

Quanto à filtragem, houve consenso entre as duas análises de que usar os dados sem filtragem é a melhor opção. Para extração de características a análise estatística não foi conclusiva já que 64 vezes a classificação com extração foi a melhor, 53 vezes a classificação sem extração foi melhor e 27 elas foram equivalentes. Isso mostra que não houve uma prevalência cabal de uma técnica sobre as demais, porém como a análise pela melhor acurácia apontou que sem fazer a extração os resultados são melhores, essa abordagem foi considerada a melhor.

Quanto ao balanceamento, a análise estatística aponta para o *oversampling* como a melhor técnica de balanceamento de dados, mas como para os grupos Controle vs. Patológico a análise pela melhor acurácia aponta para estratégia *under*, as duas técnicas foram novamente testadas e se mostraram equivalentes, possivelmente em função do pequeno desbalanceamento para registrado nesse grupo específico.

Quanto a seleção de características, a análise estatística apontou para o uso da técnica RFE, já a análise pela melhor acurácia com o sinal completo apontou para o uso da técnica LassoCV (LA). Sendo assim, as duas técnicas foram novamente avaliadas e a seleção usando LA se mostrou mais promissora.

Para ilustrar a organização dos dados por classe no grupo Controle vs. Patológico, a Figura 5.5 apresenta um gráfico de violino que mostra a distribuição das 20 principais características selecionadas pelo seletor LassoCV. Este gráfico destaca a distribuição dos dados de cada característica em relação às classes, conforme detalhado na Seção 2.8.5.

Para exibir a correlação entre as características selecionadas pelos diferentes métodos, a Figura 5.6 apresenta os mapas de calor gerados com os seletores LassoCV (A) e RFE (B). Eles mostram a correlação entre as características, sendo os quadrantes mais escuros preferíveis, pois indicam menor interdependência entre as características. Ao comparar os mapas, nota-se que o mapa A é ligeiramente mais escuro que o mapa B, o que corrobora a melhor acurácia obtida com o seletor de características LassoCV.

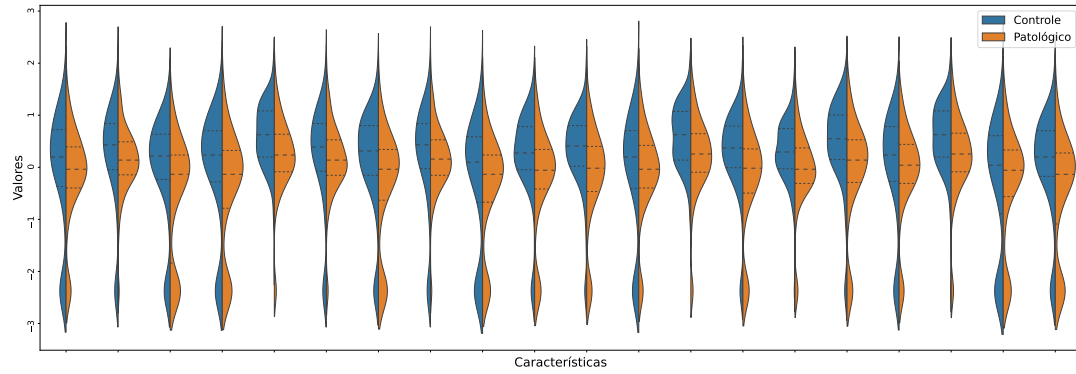


Figura 5.5: Gráfico de violino ilustrando a distribuição normalizada (por escore-Z) dos valores das 20 principais características do seletor de características LassoCV.

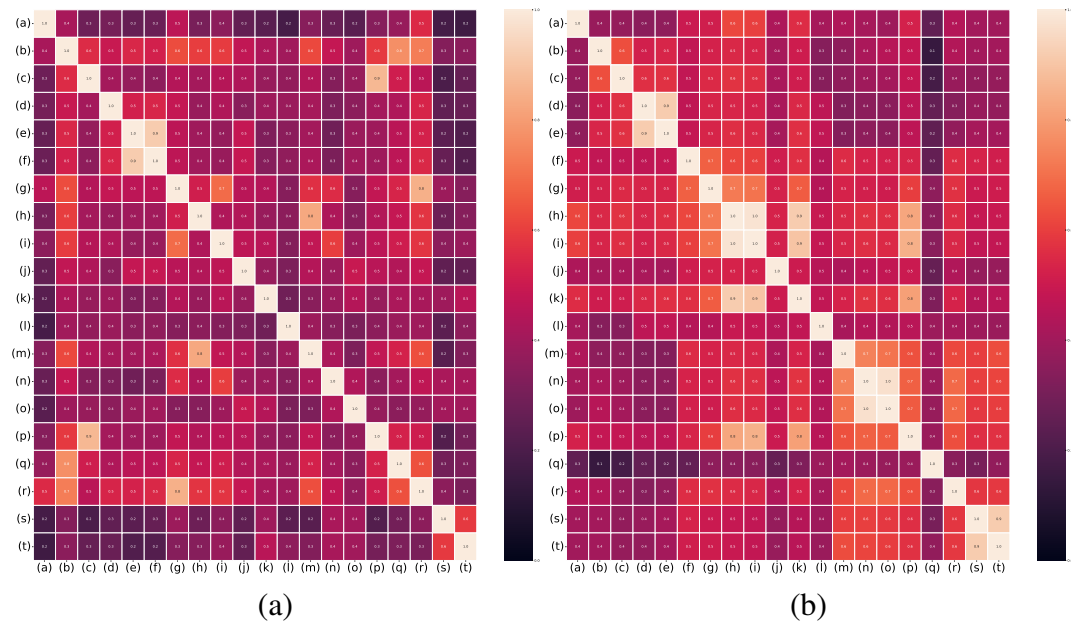


Figura 5.6: Mapa de calor das 20 principais características selecionadas pelo seletor de características LassoCV (a) e pelo seletor RFE (b).

A Figura 5.7 apresenta um gráfico *boxplot* com os níveis de acurácia alcançados pelos classificadores na abordagem ótima aplicada ao grupo de dados Controle vs. Patológico. A partir dela, é possível ver que o classificador LDA alcançou a melhor acurácia, com média de 73,9%. Por fim, a Figura 5.8 permite avaliar os resultados da classificação por meio de uma matriz de confusão que ilustra a acurácia e as classificações erradas do LDA e a área sob a curva ROC, que efetivamente demonstra a capacidade do classificador em discriminar entre os exames do grupo controle e patológico.

Tendo reunido os resultados da classificação das técnicas de aprendizado de máquina aplicadas, avançamos agora para analisar esses resultados e tirar conclusões sobre a eficácia e as implicações das técnicas usadas.

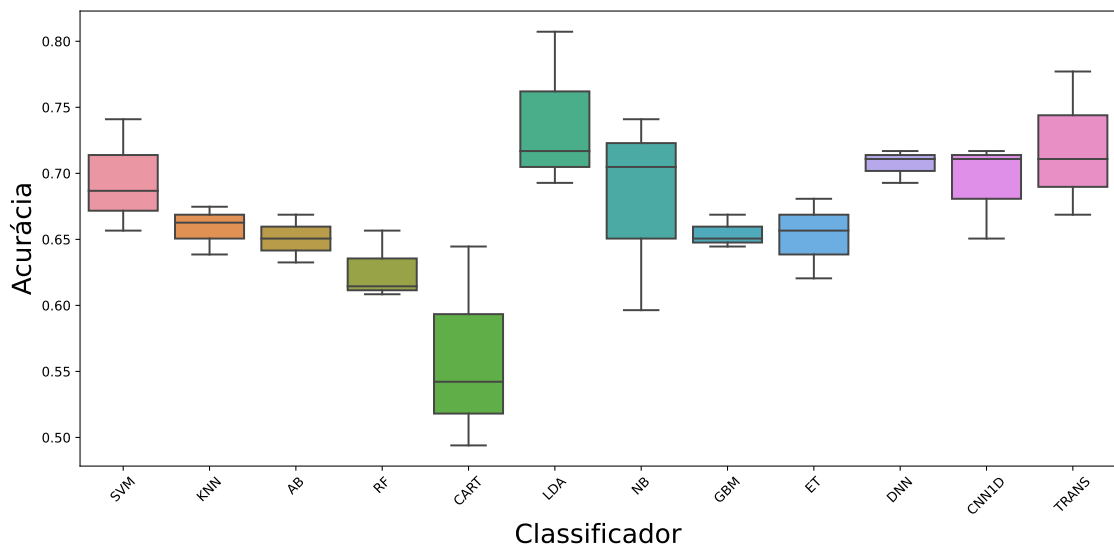


Figura 5.7: Gráfico *boxplot* apresentando a acurácia de cada um dos classificadores avaliados, para o agrupamento Controle vs Patológico, usando a abordagem mais eficaz.

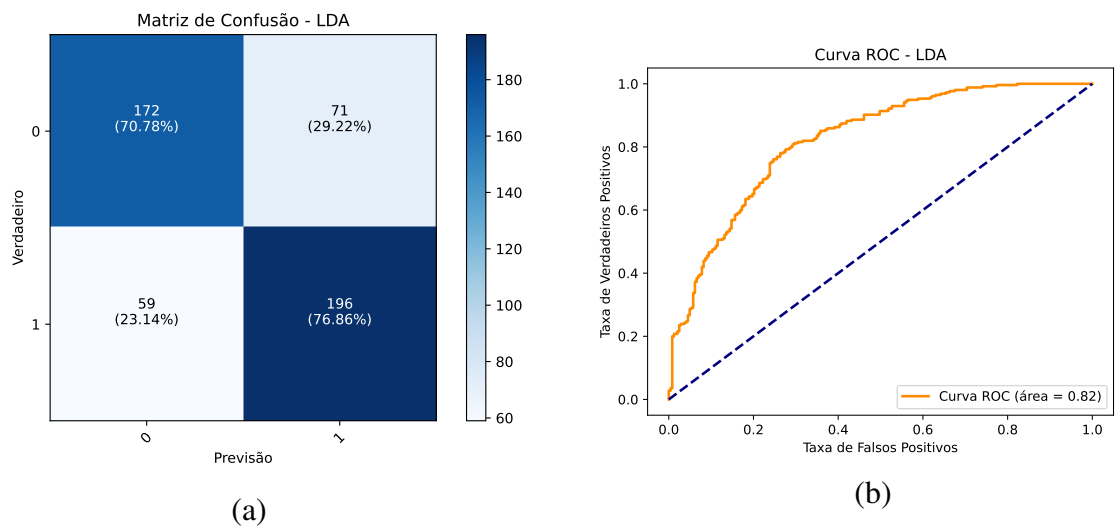


Figura 5.8: Matriz de Confusão (a) e Curva ROC (b) exibindo o desempenho do classificador LDA para o agrupamento Controle vs. Patológico usando a abordagem mais eficaz.

5.4 Discussão

Esse estudo investigou técnicas de aprendizado de máquina para classificar o glaucoma com base nos sinais do reflexo pupilar à luz. Isso envolveu explorar diferentes abordagens, como filtragem, balanceamento de amostras e técnicas de extração e seleção de características. Abaixo está um resumo detalhado de nossos principais achados.

5.4.1 Sobre a Filtragem

A análise estatística na Tabela 5.3 indicou que se abster de filtrar o sinal produz geralmente resultados superiores, independentemente das outras técnicas implementadas.

Esse resultado foi corroborado pela análise de melhor acurácia do sinal inteiro na Tabela 5.7, destacando que sinais não filtrados mantiveram sua integridade e foram mais eficazes para classificação.

5.4.2 Extração de Características

A análise estatística na Tabela 5.4 produziu resultados mistos, variando a eficácia com base nos métodos de balanceamento e seleção de características.

No entanto, a análise de melhor acurácia na Tabela 5.7 sugeriu uma preferência por utilizar todos os diâmetros de sinal, insinuando que a extração extensiva de características pode não ser necessária.

5.4.3 Sobre o Balanceamento de Amostras

Os resultados, tanto da análise estatística quanto da análise pela melhor acurácia, indicam que a técnica de *oversampling*, melhora o desempenho dos classificadores, especialmente quando há um desbalanceamento significativo. Quando o desbalanceamento é menor, em torno de 10%, as técnicas podem ser consideradas equivalentes.

5.4.4 Seleção de Características

As análises mostraram que a eficácia de diferentes técnicas de seleção de características variou bastante conforme o agrupamento, sem que um único método demonstrasse dominância sobre os demais. Embora a análise estatística indique que a Eliminação Recursiva de Características (RFE) seja a técnica mais adequada, a análise de acurácia sugere que o método LassoCV (LA) oferece um melhor desempenho. As análises, entretanto, foram unânimes em apontar o método SelectKBest como o menos indicado. Isso provavelmente ocorreu porque SelectKBest seleciona características com base em testes estatísticos individuais. Esse método avalia cada característica de forma independente, sem considerar as interações entre elas. Por outro lado, RFE e LassoCV levam em conta as interações entre as características, resultando em uma seleção mais eficaz.

5.4.5 Sobre o Fatiamento do Sinal

A análise pela melhor acurácia usando o sinal segmentado, correspondente a cada uma das quatro estimulações, não produziu conclusões definitivas sobre qual parte do sinal é mais crítica, conforme mostrado na Tabela 5.8. A eficácia de cada cor de estímulo variou conforme os diferentes tipos de agrupamentos. No entanto, dos seis agrupamentos analisados, o segmento do sinal estimulado pela cor azul apresentou a maior acurácia de classificação em três deles. Assim, a estimulação com o azul se destaca como a mais promissora.

Outra constatação que se pode chegar é que as acurácias observadas com o sinal fatiado foram ligeiramente menores em comparação com o uso do sinal completo em praticamente todos os agrupamentos. Embora a diferença de desempenho tenha sido pequena, isso sugere que, usar o sinal completo pode ser mais vantajoso. As estratégias de não filtrar o sinal, evitar a extração de características e aplicar o balanceamento de *oversampling* permaneceram como as mais eficazes praticamente em todos os grupos.

5.4.6 Análise dos Classificadores

Classificadores tradicionais demonstraram um desempenho geral satisfatório na classificação dos seis grupos estudados, como pode ser visto na Tabela 5.7. Esses algoritmos se mantiveram eficientes e raramente foram superados por técnicas baseadas em redes neurais artificiais.

Como esperado, verificou-se que a classificação do glaucoma em estágios avançados apresenta maior precisão em comparação à identificação do glaucoma nos estágios iniciais. Isso se deve ao fato de que as características patológicas associadas ao glaucoma em fases avançadas são mais pronunciadas, enquanto as alterações nos estágios iniciais tendem a ser mais sutis e, conseqüentemente, mais difíceis de serem capturadas pelos modelos de classificação.

Por fim, os resultados sugerem que preservar o espectro completo de dados do sinal sem filtragem ou extração de características pode ser mais vantajoso para a classificação. Isso se aplicou particularmente à base de dados usada nessa pesquisa, que provavelmente possui um nível de ruído aceitável. Também observou-se o impacto significativo e positivo do balanceamento usando a estratégia *oversampling* com SMOTE, e da seleção de características com as técnicas RFE e Lasso na melhoria da delimitação da fronteira de decisão dos classificadores.

5.4.7 Recomendações Práticas

A avaliação de diversas técnicas de aprendizado de máquina, para dados oriundos da pupilometria e classificação do glaucoma, sugere reflexões tais como:

Considerar o uso do conjunto de dados não filtrado, pois, para os dados analisados nesta pesquisa, essa abordagem resultou nos melhores desempenhos.

Utilizar todos os diâmetros disponíveis, pois, em nossos testes, essa abordagem se mostrou mais eficaz para a classificação do que a extração de características pré-determinadas encontradas na literatura.

Em casos de desequilíbrio na quantidade de amostra dos dados, pode ser vantajoso realizar o balanceamento por meio da criação de amostras sintéticas (*oversampling*) com o SMOTE.

O uso de técnicas de seleção de características, especialmente o LassoCV, pode contribuir para a melhoria dos resultados da classificação.

5.4.8 Limitações do Estudo

A base de dados usada na pesquisa possui uma proporção relativamente baixa de pacientes com glaucoma moderado e severo. Esta limitação, entretanto, é mitigada por representar adequadamente os casos de glaucoma inicial, o que é essencial, pois a pupilometria é especialmente útil na triagem do glaucoma em estágios iniciais.

A abordagem de extração de características pode produzir resultados aprimorados com a descoberta e inclusão de novas características do sinal pupilar, indicando potencial para mais pesquisas nesta área.

Explorar técnicas adicionais de seleção de características e avaliar estratégias alternativas de balanceamento de dados poderia aprimorar ainda mais a avaliação.

Em relação à classificação, embora os principais classificadores tenham sido avaliados, a explorar outras arquiteturas de rede e fazer o ajuste fino dos hiperparâmetros poderiam resultar em desempenhos ainda melhores.

5.4.9 Limitações da Avaliação do PLR

A ampla gama de diagnósticos baseados no reflexo pupilar à luz destaca sua sensibilidade a diversos fatores, como substâncias químicas, condições médicas e fatores fisiológicos, incluindo privação de sono e estresse. Distinguir os efeitos desses fatores ao diagnosticar o glaucoma permanece um desafio considerável.

5.5 Considerações Finais do Capítulo

Este capítulo apresentou considerações acerca da aplicação de múltiplas técnicas de aprendizado de máquina. Mostrou quais delas se mostraram vantajosas e por fim apresentou recomendações para estudos que desejam usar o reflexo pupilar à luz para apoio a diagnóstico, nesse estudo os testes foram feitos com uma base de dados voltada para o diagnóstico do glaucoma.

Conclusões

6.1 Acerca da Revisão da Literatura

A revisão da literatura mostrou que os protocolos de estimulação recomendável para o exame pupilar é aguardar 10 minutos para adaptação inicial da pupila e usar estímulos com 1 segundo de duração, em ambiente com baixa luminosidade. A técnica do reflexo consensual é frequentemente usada para minimizar reflexos nas imagens. A Tabela 3.3 apresenta alguns métodos de detecção e segmentação pupilar, destacando o algoritmo PuReST e a abordagem baseada em Redes Neurais Convolucionais proposta por Miron et al.

Ela mostrou também que o Reflexo Pupilar à Luz pode auxiliar 54 tipos diferentes de diagnósticos, sendo o glaucoma o mais estudado. Mostrou ainda que as características mais analisadas são o diâmetro máximo da pupila em repouso, a latência para início da contração, e a velocidade de contração. Por fim, na época da revisão, em meados de 2021, ficou evidente que o uso de inteligência artificial ainda era limitado, com 94% das pesquisas focadas em análise estatística e apenas 6% utilizando técnicas de IA.

6.2 Acerca da Segmentação

Este estudo ofereceu uma nova base de dados com imagens, anotadas e não anotadas da região da pupila, de indivíduos sob várias condições, incluindo consumo de álcool, diabetes mellitus tipo II e glaucoma nos estágios inicial, moderado e severo. A disponibilização desta base pode auxiliar o aprimoramento de aplicações médicas que usem o reflexo pupilar à luz para realizar diagnósticos.

O uso dessa base de dados foi exemplificado através do treinamento, teste e avaliação de algoritmos de detecção de objetos. Entre eles, o modelo Faster RCNN mostrou-se eficiente em termos de tempo de treinamento. Em contrapartida, os modelos da série YOLO, especialmente o YOLOv5, foram superiores em velocidade de inferência. E principalmente, quanto à precisão na detecção pupilar, os experimentos mostraram que

a rede YOLOv7 é a mais precisa, com uma acurácia média de 94,14%, demonstrando alta concordância com as anotações manuais e eficácia na mensuração e localização de pupilas. Por isso esse modelo foi incorporado à plataforma web *Eyes Project*.

O estudo conclui que os modelos de IA, após serem retreinados com imagens anotadas da pupila, superam os métodos não baseados em IA para a detecção e medição pupilar. Portanto, desde que não haja restrições de hardware, as redes neurais representam a melhor opção para essa tarefa.

6.3 Acerca da Classificação

Este estudo explorou o uso de técnicas de aprendizado de máquina para a classificação do glaucoma a partir do reflexo pupilar à luz, investigando técnicas de filtragem, balanceamento de amostras, e técnicas de extração e seleção de características. Os achados principais incluem:

Quanto à filtragem, os melhores resultados foram alcançados sem a filtragem do sinal, indicando que manter a integridade do sinal pode ser mais benéfico para a classificação quando nível de ruídos for aceitável.

Quanto à Extração de Características, utilizar a gama completa de diâmetros do sinal mostrou ser preferível, sugerindo que uma extração extensiva de características pode não ser necessária.

Para balancear as amostras, o uso do *oversampling* com SMOTE tende a melhorar o desempenho da classificação em casos de desbalanceamento de dados.

Quanto à seleção de características, não houve um método de seleção de características que se destacasse significativamente, indicando que a eficácia pode variar dependendo do contexto específico. Entretanto, deve-se considerar a possibilidade de se fazer a seleção com a técnica LassoCV.

Quanto ao uso do sinal completo em relação ao o uso do sinal fatiado, não foram encontradas vantagens claras em usar apenas segmentos específicos do sinal. O uso do sinal completo acabou se mostrando mais vantajoso.

Quanto aos classificadores, as análises mostram que os classificadores tradicionais apresentaram desempenho satisfatório, quase sempre superando algoritmos baseados em redes neurais artificiais. Entretanto, como nenhum classificador foi cabalmente superior aos demais, optou-se por incorporar os classificadores avaliados, tanto os clássicos quanto os baseados em redes neurais, à plataforma web.

Os resultados indicam acurácia de 73,90% na classificação geral do glaucoma, 98,10% para o glaucoma moderado e 98,73% para o severo, evidenciando que a classificação do glaucoma nos estágios moderado e severo pode ser realizada satisfatoriamente com as técnicas e classificadores avaliados. Para o estágio inicial do glaucoma, apesar das

taxas de acerto serem promissoras, melhorar esses índices ainda pode ser considerado um desafio.

6.4 Trabalhos Futuros

Após alcançar taxas de precisão satisfatórias na classificação do glaucoma usando um protocolo com períodos relativamente longos de adaptação ao escuro (10 minutos) e tempo de gravação (4 minutos), um passo subsequente seria explorar a possibilidade de reduzir essas durações, tanto de tempo de acomodação quanto de tempo de gravação, sem que se comprometa a precisão da classificação. Essa otimização tornaria o exame do reflexo pupilar mais prático para aplicações médicas.

Outra área promissora para futuras pesquisas é o desenvolvimento de métodos que permitam a realização da pupilometria sem a necessidade de um equipamento específico, viabilizando sua aplicação em ambientes cotidianos por meio de smartphones, por exemplo. Nesse cenário, os usuários poderiam registrar o reflexo pupilar com a câmera do celular, enviando os dados para um aplicativo que os analisaria e retornaria os resultados. Esse avanço ampliaria a acessibilidade e a conveniência do procedimento de triagem e diagnóstico do glaucoma.

Referências Bibliográficas

- [Adam et al. 2022]ADAM, R. et al. Synthetic training data generation and domain randomization for object detection in the formula student driverless framework. In: IEEE. *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. [S.I.], 2022. p. 1–6.
- [Adhikari, Zele e Feigl 2015]ADHIKARI, P.; ZELE, A. J.; FEIGL, B. The post-illumination pupil response (pipr). *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 56, n. 6, p. 3838–3849, 2015.
- [Adhikari et al. 2016]ADHIKARI, P. et al. Quadrant field pupillometry detects melanopsin dysfunction in glaucoma suspects and early glaucoma. *Scientific reports*, Nature Publishing Group, v. 6, p. 33373, 2016.
- [Al-Madani et al. 2021]AL-MADANI, A. M. et al. Real-time driver drowsiness detection based on eye movement and yawning using facial landmark. In: IEEE. *2021 International Conference on Computer Communication and Informatics (ICCCI)*. [S.I.], 2021. p. 1–4.
- [Amodio et al. 2018]AMODIO, A. et al. Automatic detection of driver impairment based on pupillary light reflex. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, v. 20, n. 8, p. 3038–3048, 2018.
- [An et al. 2018]AN, G. et al. Comparison of machine-learning classification models for glaucoma management. *Journal of healthcare engineering*, Hindawi, v. 2018, 2018.
- [Arévalo-López et al. 2023]ARÉVALO-LÓPEZ, C. et al. Pupillary response to chromatic light stimuli as a possible biomarker at the early stage of glaucoma: a review. *International Ophthalmology*, Springer, v. 43, n. 1, p. 343–356, 2023.
- [Asakawa et al. 2019]ASAKAWA, K. et al. Adaptation time, electroretinography, and pupillography in healthy subjects. *Documenta Ophthalmologica*, Springer, v. 139, n. 1, p. 33–44, 2019.
- [Bafna e Hansen 2021]BAFNA, T.; HANSEN, J. P. Mental fatigue measurement using eye metrics: A systematic literature review. *Psychophysiology*, Wiley Online Library, v. 58, n. 6, p. e13828, 2021.

- [Bartosova et al. 2018]BARTOSOVA, O. et al. Pupillometry as an indicator of l-dopa dosages in parkinson's disease patients. *Journal of Neural Transmission*, Springer, v. 125, n. 4, p. 699–703, 2018.
- [Bayraktar et al. 2023]BAYRAKTAR, S. et al. Evaluation of static and dynamic pupillary functions in early stage primary open angle glaucoma. *Journal of Glaucoma*, LWW, 2023.
- [Becktepe et al. 2019]BECKTEPE, J. S. et al. Pupillary response to light and tasks in early and late onset essential tremor patients. *Parkinsonism & related disorders*, Elsevier, v. 66, p. 62–67, 2019.
- [Bittner et al. 2014]BITTNER, D. M. et al. Repetitive pupil light reflex: potential marker in alzheimer's disease? *Journal of Alzheimer's Disease*, IOS Press, v. 42, n. 4, p. 1469–1477, 2014.
- [Bochkovskiy, Wang e Liao 2020]BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.
- [Bremner 2009]BREMNER, F. Pupil evaluation as a test for autonomic disorders. *Clinical Autonomic Research*, Springer, v. 19, n. 2, p. 88–101, 2009.
- [Bullock et al. 2019]BULLOCK, B. et al. Traits related to bipolar disorder are associated with an increased post-illumination pupil response. *Psychiatry research*, Elsevier, v. 278, p. 35–41, 2019.
- [Bussel, Wollstein e Schuman 2014]BUSSEL, I. I.; WOLLSTEIN, G.; SCHUMAN, J. S. Oct for glaucoma diagnosis, screening and detection of glaucoma progression. *British Journal of Ophthalmology*, BMJ Publishing Group Ltd, v. 98, n. Suppl 2, p. ii15–ii19, 2014.
- [Campbell William W. Barohn 2021]CAMPBELL WILLIAM W. BAROHN, R. J. *O Exame Neurológico*. 8ª. ed. [S.l.]: Guanabara Koogan, 2021. ISBN 8527737698.
- [Canny 1986]CANNY, J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, n. 6, p. 679–698, 1986.
- [Carle et al. 2014]CARLE, C. F. et al. Luminance and colour variant pupil perimetry in glaucoma. *Clinical & Experimental Ophthalmology*, Wiley Online Library, v. 42, n. 9, p. 815–824, 2014.
- [Chan, Hadjiiski e Samala 2020]CHAN, H.-P.; HADJIISKI, L. M.; SAMALA, R. K. Computer-aided diagnosis in the era of deep learning. *Medical physics*, Wiley Online Library, v. 47, n. 5, p. e218–e227, 2020.

- [Chang et al. 2013]CHANG, D. S. et al. Development and validation of an associative model for the detection of glaucoma using pupillography. *American journal of ophthalmology*, Elsevier, v. 156, n. 6, p. 1285–1296, 2013.
- [Charalel, Lin e Singh 2014]CHARALEL, R. A.; LIN, H. S.; SINGH, K. Glaucoma screening using relative afferent pupillary defect. *Journal of Glaucoma*, LWW, v. 23, n. 3, p. 169–173, 2014.
- [Chen et al. 2022]CHEN, J.-T. et al. Fatigue and arousal modulations revealed by saccade and pupil dynamics. *International journal of environmental research and public health*, MDPI, v. 19, n. 15, p. 9234, 2022.
- [Chen et al. 2011]CHEN, J. W. et al. Pupillary reactivity as an early indicator of increased intracranial pressure: The introduction of the neurological pupil index. *Surgical neurology international*, Scientific Scholar, v. 2, 2011.
- [Choi et al. 2019]CHOI, J. H. et al. Accurate eye pupil localization using heterogeneous cnn models. In: IEEE. *2019 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2019. p. 2179–2183.
- [Cimbalnik et al. 2022]CIMBALNIK, J. et al. Intracranial electrophysiological recordings from the human brain during memory tasks with pupillometry. *Scientific Data*, Nature Publishing Group UK London, v. 9, n. 1, p. 6, 2022.
- [Conover 1999]CONOVER, W. J. *Practical nonparametric statistics*. [S.l.]: john wiley & sons, 1999.
- [Corcoran et al. 2012]CORCORAN, P. M. et al. Real-time eye gaze tracking for gaming design and consumer electronics systems. *IEEE Transactions on Consumer Electronics*, IEEE, v. 58, n. 2, p. 347–355, 2012.
- [Crippa, Domellöf e Kawasaki 2018]CRIPPA, S. V.; DOMELLÖF, F. P.; KAWASAKI, A. Chromatic pupillometry in children. *Frontiers in neurology*, Frontiers Media SA, v. 9, p. 669, 2018.
- [Czajka 2015]CZAJKA, A. Pupil dynamics for iris liveness detection. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 10, n. 4, p. 726–735, 2015.
- [Dalal e Triggs 2005]DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. [S.l.], 2005. v. 1, p. 886–893.

- [Duque-Chica et al. 2018]DUQUE-CHICA, G. L. et al. Inner and outer retinal contributions to pupillary light response: correlation to functional and morphologic parameters in glaucoma. *Journal of glaucoma*, Wolters Kluwer, v. 27, n. 8, p. 723–732, 2018.
- [Eddy, Sanders e Eddy 1983]EDDY, D. M.; SANDERS, L. E.; EDDY, J. F. The value of screening for glaucoma with tonometry. *Survey of ophthalmology*, Elsevier, v. 28, n. 3, p. 194–205, 1983.
- [Efron e Hastie 2021]EFRON, B.; HASTIE, T. *Computer age statistical inference, student edition: algorithms, evidence, and data science*. [S.l.]: Cambridge University Press, 2021.
- [Everingham et al. 2010]EVERINGHAM, M. et al. The pascal visual object classes (voc) challenge. *International journal of computer vision*, Springer, v. 88, p. 303–338, 2010.
- [Fan et al. 2009]FAN, X. et al. Abnormal transient pupillary light reflex in individuals with autism spectrum disorders. *Journal of autism and developmental disorders*, Springer, v. 39, n. 11, p. 1499–1508, 2009.
- [Fawcett 2006]FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006.
- [Felzenszwalb, McAllester e Ramanan 2008]FELZENSZWALB, P.; MCALLESTER, D.; RAMANAN, D. A discriminatively trained, multiscale, deformable part model. In: *IEEE. 2008 IEEE conference on computer vision and pattern recognition*. [S.l.], 2008. p. 1–8.
- [Firmino et al. 2016]FIRMINO, M. et al. Computer-aided detection (cade) and diagnosis (cadx) system for lung cancer with likelihood of malignancy. *Biomedical engineering online*, BioMed Central, v. 15, n. 1, p. 1–17, 2016.
- [Fischler e Bolles 1981]FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, ACM New York, NY, USA, v. 24, n. 6, p. 381–395, 1981.
- [Fisher 1944]FISHER, R. A. *Statistical methods for research workers*. 1944.
- [Fuhl et al. 2015]FUHL, W. et al. Excuse: Robust pupil detection in real-world scenarios. In: SPRINGER. *International Conference on Computer Analysis of Images and Patterns*. [S.l.], 2015. p. 39–51.
- [Fuhl et al. 2015]FUHL, W. et al. Excuse: Robust pupil detection in real-world scenarios. In: *Computer Analysis of Images and Patterns*. Springer International Publishing, 2015. p. 39–51. Disponível em: <<https://doi.org/10.1007/978-3-319-23192-1-4>>.

- [Fuhl, Rosenstiel e Kasneci 2019]FUHL, W.; ROSENSTIEL, W.; KASNECI, E. 500,000 images closer to eyelid and pupil segmentation. In: *Computer Analysis of Images and Patterns, CAIP*. [S.l.: s.n.], 2019.
- [Fuhl et al. 2016]FUHL, W. et al. Pupilnet: Convolutional neural networks for robust pupil detection. *arXiv preprint arXiv:1601.04902*, 2016.
- [Fuhl et al. 2017]FUHL, W. et al. Pupilnet v2. 0: Convolutional neural networks for cpu based real time robust pupil detection. *arXiv preprint arXiv:1711.00112*, 2017.
- [Fuhl et al. 2016]FUHL, W. et al. EISe. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 2016. Disponível em: <<https://doi.org/10.1145/2857491.2857505>>.
- [Fuhl et al. 2016]FUHL, W. et al. Else: Ellipse selection for robust pupil detection in real-world environments. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. [S.l.: s.n.], 2016. p. 123–130.
- [Gaddipati e Sivaswamy 2021]GADDIPATI, D. J.; SIVASWAMY, J. glaucoma assessment from fundus images with fundus to oct feature space mapping. *ACM Transactions on Computing for Healthcare (HEALTH)*, ACM New York, NY, v. 3, n. 1, p. 1–15, 2021.
- [George e Routray 2018]GEORGE, A.; ROUTRAY, A. Escaf: Pupil centre localization algorithm with candidate filtering. *arXiv preprint arXiv:1807.10520*, 2018.
- [Giovannangeli et al. 2024]GIOVANNANGELI, C. J. et al. Pupil light reflex in young elite athletes: autonomic nervous system activity and viscoelastic properties. *Frontiers in Physiology*, Frontiers Media SA, v. 15, p. 1421676, 2024.
- [Girshick 2015]GIRSHICK, R. Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2015. p. 1440–1448.
- [Girshick et al. 2014]GIRSHICK, R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 580–587.
- [González et al. 2017]GONZÁLEZ, M. T. G. et al. Analysis of pupillary response after a stimulus of light to generate characteristical groups. In: IEEE. *2017 International Conference on Electronics, Communications and Computers (CONIELECOMP)*. [S.l.], 2017. p. 1–6.
- [Gou et al. 2017]GOU, C. et al. A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition*, Elsevier, v. 67, p. 23–31, 2017.

- [Gracitelli et al. 2014]GRACITELLI, C. P. et al. A positive association between intrinsically photosensitive retinal ganglion cells and retinal nerve fiber layer thinning in glaucoma. *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 55, n. 12, p. 7997–8005, 2014.
- [Gracitelli et al. 2015]GRACITELLI, C. P. et al. Intrinsically photosensitive retinal ganglion cell activity is associated with decreased sleep quality in patients with glaucoma. *Ophthalmology*, Elsevier, v. 122, n. 6, p. 1139–1148, 2015.
- [Han et al. 2019]HAN, S. Y. et al. Pupil center detection based on the unet for the user interaction in vr and ar environments. In: IEEE. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. [S.l.], 2019. p. 958–959.
- [He e Ma 2013]HE, H.; MA, Y. *Imbalanced learning: foundations, algorithms, and applications*. [S.l.]: John Wiley & Sons, 2013. 47 p.
- [He et al. 2017]HE, K. et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2961–2969.
- [He et al. 2016]HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- [Hodapp 1993]HODAPP, E. Clinical decisions in glaucoma. *The CV Mosby Co*, 1993.
- [Hoyt, Frisen e Newman 1973]HOYT, W. F.; FRISEN, L.; NEWMAN, N. M. Fundoscopy of nerve fiber layer defects in glaucoma. *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 12, n. 11, p. 814–829, 1973.
- [Hu e Lodewijks 2020]HU, X.; LODEWIJKS, G. Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *Journal of safety research*, Elsevier, v. 72, p. 173–187, 2020.
- [Ishibashi et al. 2017]ISHIBASHI, F. et al. The preferential impairment of pupil constriction stimulated by blue light in patients with type 2 diabetes without autonomic neuropathy. *Journal of diabetes research*, Hindawi, v. 2017, 2017.
- [Jaccard 1901]JACCARD, P. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, v. 37, p. 547–579, 1901.
- [Javadi et al. 2015]JAVADI, A.-H. et al. Set: a pupil detection method using sinusoidal approximation. *Frontiers in neuroengineering*, Frontiers, v. 8, p. 4, 2015.

- [Jocher et al. 2022]JOCHER, G. et al. *ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference*. Zenodo, fev. 2022. Disponível em: <<https://doi.org/10.5281/zenodo.6222936>>.
- [Joshi e Gold 2020]JOSHI, S.; GOLD, J. I. Pupil size as a window on neural substrates of cognition. *Trends in cognitive sciences*, Elsevier, v. 24, n. 6, p. 466–480, 2020.
- [Jr e Vessani 2009]JR, R. S.; VESSANI, R. M. Staging glaucoma patient: why and how? *The open ophthalmology journal*, Bentham Science Publishers, v. 3, p. 59, 2009.
- [Kankipati, Girkin e Gamlin 2011]KANKIPATI, L.; GIRKIN, C. A.; GAMLIN, P. D. The post-illumination pupil response is reduced in glaucoma patients. *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 52, n. 5, p. 2287–2292, 2011.
- [Kelbsch et al. 2019]KELBSCH, C. et al. Standards in pupillography. *Frontiers in neurology*, Frontiers Media SA, v. 10, p. 129, 2019.
- [Khalid, Khalil e Nasreen 2014]KHALID, S.; KHALIL, T.; NASREEN, S. A survey of feature selection and feature extraction techniques in machine learning. In: IEEE. *2014 science and information conference*. [S.l.], 2014. p. 372–378.
- [Kitazumi e Nakazawa 2018]KITAZUMI, K.; NAKAZAWA, A. Robust pupil segmentation and center detection from visible light images using convolutional neural network. In: IEEE. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2018. p. 862–868.
- [Kitchenham 2004]KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele University*, v. 33, n. 2004, p. 1–26, 2004.
- [Kohavi et al. 1995]KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. *Ijcai*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.
- [Krizhevsky, Sutskever e Hinton 2012]KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, p. 1097–1105, 2012.
- [Kumar e Minz 2014]KUMAR, V.; MINZ, S. Feature selection: a literature review. *SmartCR*, v. 4, n. 3, p. 211–229, 2014.
- [Lawlor et al. 2017]LAWLOR, M. et al. Patients with normal tension glaucoma have relative sparing of the relative afferent pupillary defect compared to those with open angle glaucoma and elevated intraocular pressure. *Investigative Ophthalmology & Visual Science*,

- The Association for Research in Vision and Ophthalmology, v. 58, n. 12, p. 5237–5241, 2017.
- [Lee et al. 2019]LEE, J. et al. Temporal raphe sign for discrimination of glaucoma from optic neuropathy in eyes with macular ganglion cell–inner plexiform layer thinning. *Ophthalmology*, Elsevier, v. 126, n. 8, p. 1131–1139, 2019.
- [Lehmann, Romano e Casella 1986]LEHMANN, E. L.; ROMANO, J. P.; CASELLA, G. *Testing statistical hypotheses*. [S.l.]: Springer, 1986.
- [Lei et al. 2014]LEI, S. et al. Full-field chromatic pupillometry for the assessment of the postillumination pupil response driven by melanopsin-containing retinal ganglion cells. *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 55, n. 7, p. 4496–4503, 2014.
- [Levatin, Prasloski e Collen 1973]LEVATIN, P.; PRASLOSKI, P.; COLLEN, M. The swinging flashlight test in multiphasic screening for eye disease. *Canadian journal of ophthalmology. Journal canadien d'ophtalmologie*, v. 8, n. 2, p. 356, 1973.
- [Li, Winfield e Parkhurst 2005]LI, D.; WINFIELD, D.; PARKHURST, D. J. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In: IEEE. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*. [S.l.], 2005. p. 79–79.
- [Lin et al. 2017]LIN, T.-Y. et al. Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 2117–2125.
- [Lin et al. 2017]LIN, T.-Y. et al. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2980–2988.
- [Liu et al. 2010]LIU, D. et al. Drowsiness detection based on eyelid movement. In: IEEE. *2010 second international workshop on education technology and computer science*. [S.l.], 2010. v. 2, p. 49–52.
- [Liu et al. 2016]LIU, W. et al. Ssd: Single shot multibox detector. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 21–37.
- [López et al. 2017]LÓPEZ, S. U. et al. Identification of parameters for the study of diabetes from light reflex with controlled stimulus. In: IEEE. *2017 International Conference on Electronics, Communications and Computers (CONIELECOMP)*. [S.l.], 2017. p. 1–6.

- [Lowenstein e Loewenfeld 1958]LOWENSTEIN, O.; LOEWENFELD, I. E. Electronic pupillography: a new instrument and some clinical applications. *AMA archives of ophthalmology*, American Medical Association, v. 59, n. 3, p. 352–363, 1958.
- [Lynch, James e VanDam 2018]LYNCH, G. T.; JAMES, S. M.; VANDAM, M. Pupillary response and phenotype in asd: Latency to constriction discriminates asd from typically developing adolescents. *Autism Research*, Wiley Online Library, v. 11, n. 2, p. 364–375, 2018.
- [Mandal et al. 2021]MANDAL, A. C. et al. A deep-learning approach to pupillometry. In: ZELINSKI, M. E.; TAHA, T. M.; HOWE, J. (Ed.). *Applications of Machine Learning 2021*. SPIE, 2021. Disponível em: <<https://doi.org/10.1117/12.2594315>>.
- [Martucci et al. 2014]MARTUCCI, A. et al. Evaluation of pupillary response to light in patients with glaucoma: a study using computerized pupillometry. *International ophthalmology*, Springer, v. 34, n. 6, p. 1241–1247, 2014.
- [Mathôt 2018]MATHÔT, S. Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, Ubiquity Press, v. 1, n. 1, 2018.
- [Maynard, Zele e Feigl 2015]MAYNARD, M. L.; ZELE, A. J.; FEIGL, B. Melanopsin-mediated post-illumination pupil response in early age-related macular degeneration. *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 56, n. 11, p. 6906–6913, 2015.
- [Metzger e Emborg 2019]METZGER, J. M.; EMBORG, M. E. Autonomic dysfunction in parkinson disease and animal models. *Clinical Autonomic Research*, Springer, v. 29, n. 4, p. 397–414, 2019.
- [Miron et al. 2019]MIRON, C. et al. Efficient pupil detection with a convolutional neural network. In: IEEE. *2019 E-Health and Bioengineering Conference (EHB)*. [S.l.], 2019. p. 1–4.
- [Montgomery e Runger 2010]MONTGOMERY, D. C.; RUNGER, G. C. *Applied statistics and probability for engineers*. [S.l.]: John wiley & sons, 2010.
- [Najjar et al. 2023]NAJJAR, R. P. et al. Handheld chromatic pupillometry can accurately and rapidly reveal functional loss in glaucoma. *British Journal of Ophthalmology*, BMJ Publishing Group Ltd, v. 107, n. 5, p. 663–670, 2023.
- [Najjar et al. 2018]NAJJAR, R. P. et al. Pupillary responses to full-field chromatic stimuli are reduced in patients with early-stage primary open-angle glaucoma. *Ophthalmology*, Elsevier, v. 125, n. 9, p. 1362–1371, 2018.

- [NCI 2022]NCI, D. o. C. T. *National Cancer Institute at the National Institutes of Health, NCI Dictionary of Cancer Terms*. 2022. (acessado em Janeiro de 2022). Disponível em: <<https://www.cancer.gov/publications/dictionaries/cancer-terms/def/diagnosis>>.
- [Ngo et al. 2021]NGO, Q. C. et al. Pupillary complexity for the screening of glaucoma. *IEEE Access*, IEEE, v. 9, p. 144871–144879, 2021.
- [NIH 2022]NIH, C. R. F. National institute of general medical sciences. 2022. (acessado em Janeiro de 2022). Disponível em: <<https://www.nigms.nih.gov/education/fact-sheets/Pages/circadian-rhythms.aspx>>.
- [Nowak et al. 2016]NOWAK, W. et al. Feature analyses of pupil light reflex to chromatic stimuli in alzheimer's patients. In: IEEE. *2016 2nd International Conference on Frontiers of Signal Processing (ICFSP)*. [S.l.], 2016. p. 58–62.
- [Ong, Hutch e Smirnakis 2019]ONG, C.; HUTCH, M.; SMIRNAKIS, S. The effect of ambient light conditions on quantitative pupillometry. *Neurocritical Care*, Springer, v. 30, n. 2, p. 316–321, 2019.
- [Padilla, Netto e Silva 2020]PADILLA, R.; NETTO, S. L.; SILVA, E. A. D. A survey on performance metrics for object-detection algorithms. In: IEEE. *2020 international conference on systems, signals and image processing (IWSSIP)*. [S.l.], 2020. p. 237–242.
- [Park et al. 2019]PARK, H.-Y. L. et al. Detecting autonomic dysfunction in patients with glaucoma using dynamic pupillometry. *Medicine*, Wolters Kluwer Health, v. 98, n. 11, 2019.
- [Park et al. 2019]PARK, J. C. et al. Electrophysiological and pupillometric measures of inner retina function in nonproliferative diabetic retinopathy. *Documenta Ophthalmologica*, Springer, v. 139, n. 2, p. 99–111, 2019.
- [Park e McAnany 2015]PARK, J. C.; MCANANY, J. J. Effect of stimulus size and luminance on the rod-, cone-, and melanopsin-mediated pupillary light reflex. *Journal of Vision*, The Association for Research in Vision and Ophthalmology, v. 15, n. 3, p. 13–13, 2015.
- [Park, Moss e McAnany 2016]PARK, J. C.; MOSS, H. E.; MCANANY, J. J. The pupillary light reflex in idiopathic intracranial hypertension. *Investigative ophthalmology & visual science*, The Association for Research in Vision and Ophthalmology, v. 57, n. 1, p. 23–29, 2016.
- [Park et al. 2011]PARK, J. C. et al. Toward a clinical protocol for assessing rod, cone, and melanopsin contributions to the human pupil response. *Investigative ophthalmology &*

- visual science*, The Association for Research in Vision and Ophthalmology, v. 52, n. 9, p. 6624–6635, 2011.
- [Pattan et al. 2023]PATTAN, H. F. et al. Assessing the pupillary response in healthy and primary open-angle glaucoma. *Investigative Ophthalmology & Visual Science*, The Association for Research in Vision and Ophthalmology, v. 64, n. 8, p. 2034–2034, 2023.
- [Pillai et al. 2019]PILLAI, M. R. et al. Quantification of rapd by an automated pupillometer in asymmetric glaucoma and its correlation with manual pupillary assessment. *Indian journal of ophthalmology*, Wolters Kluwer–Medknow Publications, v. 67, n. 2, p. 227, 2019.
- [Pinheiro et al. 2015]PINHEIRO, H. et al. Metodologia e dispositivo portátil para avaliação do reflexo pupilar. In: *Twenty-first Americas Conference on Information Systems, Puerto Rico*. [S.l.: s.n.], 2015. p. 10.
- [Pinheiro e Costa 2021]PINHEIRO, H. M.; COSTA, R. M. da. Pupillary light reflex as a diagnostic aid from computational viewpoint: A systematic literature review. *Journal of Biomedical Informatics*, Elsevier, p. 103757, 2021.
- [Pinheiro e Costa 2021]PINHEIRO, H. M.; COSTA, R. M. da. Pupillary light reflex as a diagnostic aid from computational viewpoint: A systematic literature review. *Journal of Biomedical Informatics*, Elsevier BV, v. 117, p. 103757, maio 2021. Disponível em: <<https://doi.org/10.1016/j.jbi.2021.103757>>.
- [Pinheiro et al. 2015]PINHEIRO, H. M. et al. A new approach to detect use of alcohol through iris videos using computer vision. In: SPRINGER. *International Conference on Image Analysis and Processing*. [S.l.], 2015. p. 598–608.
- [Por, Kooten e Sarkovic 2019]POR, E.; KOOTEN, M. van; SARKOVIC, V. Nyquist–shannon sampling theorem. *Leiden University*, v. 1, n. 1, p. 5, 2019.
- [Pradhan et al. 2017]PRADHAN, Z. S. et al. Predicting the magnitude of functional and structural damage in glaucoma from monocular pupillary light responses using automated pupillography. *Journal of Glaucoma*, Wolters Kluwer, v. 26, n. 5, p. 409–414, 2017.
- [Pramanik et al. 2021]PRAMANIK, A. et al. Granulated rcnn and multi-class deep sort for multi-object detection and tracking. *IEEE Transactions on Emerging Topics in Computational Intelligence*, IEEE, 2021.
- [Proakis 2007]PROAKIS, J. G. *Digital signal processing: principles, algorithms, and applications*, 4/E. [S.l.]: Pearson Education India, 2007.

- [Proença et al. 2009]PROENÇA, H. et al. The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 32, n. 8, p. 1529–1535, 2009.
- [Puntmann 2009]PUNTMANN, V. How-to guide on biomarkers: biomarker definitions, validation and applications with examples from cardiovascular disease. *Postgraduate medical journal*, Oxford University Press, v. 85, n. 1008, p. 538–545, 2009.
- [Purves et al. 2008]PURVES, D. et al. *Principles of cognitive neuroscience*. [S.l.]: Sinauer Associates Sunderland, MA, 2008.
- [Quan et al. 2023]QUAN, Y. et al. Binocular head-mounted chromatic pupillometry can detect structural and functional loss in glaucoma. *Frontiers in Neuroscience*, Frontiers Media SA, v. 17, 2023.
- [Razali, Wah et al. 2011]RAZALI, N. M.; WAH, Y. B. et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, v. 2, n. 1, p. 21–33, 2011.
- [Redmon et al. 2016]REDMON, J. et al. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788.
- [Redmon e Farhadi 2018]REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [Ren et al. 2015]REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, v. 28, p. 91–99, 2015.
- [Richter et al. 2017]RICHTER, P. et al. The diagnostic accuracy of chromatic pupillary light responses in diseases of the outer and inner retina. *Graefe's Archive for Clinical and Experimental Ophthalmology*, Springer, v. 255, n. 3, p. 519–527, 2017.
- [Rollins et al. 2014]ROLLINS, M. D. et al. Pupillary effects of high-dose opioid quantified with infrared pupillometry. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, The American Society of Anesthesiologists, v. 121, n. 5, p. 1037–1044, 2014.
- [Rukmini, Milea e Gooley 2019]RUKMINI, A.; MILEA, D.; GOOLEY, J. J. Chromatic pupillometry methods for assessing photoreceptor health in retinal and optic nerve diseases. *Frontiers in neurology*, Frontiers Media SA, v. 10, p. 76, 2019.

- [Rukmini et al. 2015]RUKMINI, A. V. et al. Pupillary responses to high-irradiance blue light correlate with glaucoma severity. *Ophthalmology*, Elsevier, v. 122, n. 9, p. 1777–1785, 2015.
- [Sahayadhas, Sundaraj e Murugappan 2012]SAHAYADHAS, A.; SUNDARAJ, K.; MURUGAPPAN, M. Detecting driver drowsiness based on sensors: a review. *Sensors*, Molecular Diversity Preservation International (MDPI), v. 12, n. 12, p. 16937–16953, 2012.
- [Santini, Fuhl e Kasneci 2018]SANTINI, T.; FUHL, W.; KASNECI, E. Pure: Robust pupil detection for real-time pervasive eye tracking. *Computer Vision and Image Understanding*, Elsevier, v. 170, p. 40–50, 2018.
- [Santini, Fuhl e Kasneci 2018]SANTINI, T.; FUHL, W.; KASNECI, E. Purest: robust pupil tracking for real-time pervasive eye tracking. In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. [S.l.: s.n.], 2018. p. 1–5.
- [Sarezky et al. 2014]SAREZKY, D. et al. Correlation between intereye difference in visual field mean deviation values and relative afferent pupillary response as measured by an automated pupillometer in subjects with glaucoma. *Journal of Glaucoma*, LWW, v. 23, n. 7, p. 419–423, 2014.
- [Sarezky et al. 2016]SAREZKY, D. et al. Correlation between inter-eye difference in average retinal nerve fiber layer thickness and afferent pupillary response as measured by an automated pupillometer in glaucoma. *Journal of Glaucoma*, Wolters Kluwer, v. 25, n. 3, p. 312–316, 2016.
- [Setiawan, Wibirama e Setiawan 2018]SETIAWAN, M. T.; WIBIRAMA, S.; SETIAWAN, N. A. Robust pupil localization algorithm based on circular hough transform for extreme pupil occlusion. In: *IEEE. 2018 4th International Conference on Science and Technology (ICST)*. [S.l.], 2018. p. 1–5.
- [Shapiro e Wilk 1965]SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, Oxford University Press, v. 52, n. 3-4, p. 591–611, 1965.
- [Silva et al. 2017]SILVA, C. R. G. et al. Automated evaluation system for human pupillary behavior. *International Medical Informatics Association (IMIA) and IOS Press, Volume 245: MEDINFO 2017: Precision Healthcare through Informatics*, p. 589 – 593, 2017.
- [Simonyan e Zisserman 2014]SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

- [Stockman e Shapiro 2001]STOCKMAN, G.; SHAPIRO, L. G. *Computer Vision*. [S.l.]: Prentice Hall PTR, 2001. 279–325 p.
- [Student 1908]STUDENT. The probable error of a mean. *Biometrika*, JSTOR, p. 1–25, 1908.
- [Suzuki 2012]SUZUKI, K. A review of computer-aided diagnosis in thoracic and colonic imaging. *Quantitative imaging in medicine and surgery*, AME Publications, v. 2, n. 3, p. 163, 2012.
- [Świrski, Bulling e Dodgson 2012]ŚWIRSKI, L.; BULLING, A.; DODGSON, N. Robust real-time pupil tracking in highly off-axis images. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. [S.l.: s.n.], 2012. p. 173–176.
- [Szegedy et al. 2015]SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9.
- [Talaat et al. 2019]TALAAT, M.-A. et al. Glaucoma detection from retinal images using generic features: Analysis & results. In: *Proceedings of the 2019 2nd International Conference on Watermarking and Image Processing*. [S.l.: s.n.], 2019. p. 10–15.
- [Tatham et al. 2014]TATHAM, A. J. et al. Detecting glaucoma using automated pupillography. *Ophthalmology*, Elsevier, v. 121, n. 6, p. 1185–1193, 2014.
- [Tatham et al. 2014]TATHAM, A. J. et al. Estimation of retinal ganglion cell loss in glaucomatous eyes with a relative afferent pupillary defect. *Investigative Ophthalmology & Visual Science*, The Association for Research in Vision and Ophthalmology, v. 55, n. 1, p. 513–522, 2014.
- [Teasdale e Jennett 1974]TEASDALE, G.; JENNETT, B. Assessment of coma and impaired consciousness: a practical scale. *The Lancet*, Elsevier, v. 304, n. 7872, p. 81–84, 1974.
- [Temel et al. 2019]TEMEL, D. et al. Relative afferent pupillary defect screening through transfer learning. *IEEE Journal of Biomedical and Health Informatics*, IEEE, v. 24, n. 3, p. 788–795, 2019.
- [Tkachenko et al. 2020-2022]TKACHENKO, M. et al. *Label Studio: Data labeling software*. 2020–2022. Open source software available from <https://github.com/heartexlabs/label-studio>. Disponível em: <<https://github.com/heartexlabs/label-studio>>.

- [Traustason et al. 2016]TRAUSTASON, S. et al. Pupillary response to direct and consensual chromatic light stimuli. *Acta ophthalmologica*, Wiley Online Library, v. 94, n. 1, p. 65–69, 2016.
- [Veit et al. 2016]VEIT, A. et al. Coco-text: Dataset and benchmark for text detection and recognition in natural images. *arXiv preprint arXiv:1601.07140*, 2016.
- [Vera-Olmos e Malpica 2017]VERA-OLMOS, F. J.; MALPICA, N. Deconvolutional neural network for pupil detection in real-world environments. In: SPRINGER. *International Work-Conference on the Interplay Between Natural and Artificial Computation*. [S.l.], 2017. p. 223–231.
- [Viola e Jones 2001]VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: IEEE. *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. [S.l.], 2001. v. 1, p. I–I.
- [Wang et al. 2015]WANG, B. et al. Dark adaptation-induced changes in rod, cone and intrinsically photosensitive retinal ganglion cell (iprgc) sensitivity differentially affect the pupil light response (plr). *Graefe's Archive for Clinical and Experimental Ophthalmology*, Springer, v. 253, n. 11, p. 1997–2005, 2015.
- [Wang, Bochkovskiy e Liao 2023]WANG, C.-Y.; BOCHKOVSKIY, A.; LIAO, H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2023. p. 7464–7475.
- [Wang, Bochkovskiy e Liao 2023]WANG, C.-Y.; BOCHKOVSKIY, A.; LIAO, H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2023. p. 7464–7475.
- [Wasserstein e Lazar 2016]WASSERSTEIN, R. L.; LAZAR, N. A. *The ASA statement on p-values: context, process, and purpose*. [S.l.]: Taylor & Francis, 2016. 129–133 p.
- [Wedel, Pieters et al. 2008]WEDEL, M.; PIETERS, R. et al. Eye tracking for visual marketing. *Foundations and Trends® in Marketing*, Now Publishers, Inc., v. 1, n. 4, p. 231–320, 2008.
- [Weinreb, Aung e Medeiros 2014]WEINREB, R. N.; AUNG, T.; MEDEIROS, F. A. The pathophysiology and treatment of glaucoma: a review. *Jama*, American Medical Association, v. 311, n. 18, p. 1901–1911, 2014.

- [Weinreb e Khaw 2004]WEINREB, R. N.; KHAW, P. T. Primary open-angle glaucoma. *The lancet*, Elsevier, v. 363, n. 9422, p. 1711–1720, 2004.
- [Weiss, Khoshgoftaar e Wang 2016]WEISS, K.; KHOSHGOFTAAR, T. M.; WANG, D. A survey of transfer learning. *Journal of Big data*, Springer, v. 3, p. 1–40, 2016.
- [Wilcoxon 1992]WILCOXON, F. Individual comparisons by ranking methods. In: *Breakthroughs in statistics: Methodology and distribution*. [S.l.]: Springer, 1992. p. 196–202.
- [Wu et al. 2019]WU, L. et al. Detect glaucoma with image segmentation and transfer learning. In: *Proceedings of the 32nd International Conference on Computer Animation and Social Agents*. [S.l.: s.n.], 2019. p. 37–40.
- [Wu et al. 2019]WU, Y. et al. *Detectron2*. 2019. <https://github.com/facebookresearch/detectron2>.
- [Wu e Medeiros 2018]WU, Z.; MEDEIROS, F. A. Recent developments in visual field testing for glaucoma. *Current opinion in ophthalmology*, Wolters Kluwer, v. 29, n. 2, p. 141–146, 2018.
- [Yano, Zimmer e Ling 2015]YANO, V.; ZIMMER, A.; LING, L. L. Extraction and application of dynamic pupillometry features for biometric authentication. *Measurement*, Elsevier, v. 63, p. 41–48, 2015.
- [Yaquub 2012]YAQUB, M. Visual fields interpretation in glaucoma: a focus on static automated perimetry. *Community eye health*, International Centre for Eye Health, v. 25, n. 79-80, p. 1, 2012.
- [Yetkin et al. 2019]YETKIN, E. et al. Evaluation of static and dynamic pupil characteristics in hyperopic anisometropic amblyopia. *European Journal of Ophthalmology*, SAGE Publications Sage UK: London, England, v. 29, n. 5, p. 486–493, 2019.
- [Yoo, Yang e Hwang 2017]YOO, Y. J.; YANG, H. K.; HWANG, J.-M. Efficacy of digital pupillometry for diagnosis of horner syndrome. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 6, p. e0178361, 2017.
- [Youden 1950]YOU DEN, W. J. Index for rating diagnostic tests. *Cancer*, Wiley Online Library, v. 3, n. 1, p. 32–35, 1950.
- [Zandi et al. 2021]ZANDI, B. et al. Pupilext: Flexible open-source platform for high-resolution pupillometry in vision research. *Frontiers in neuroscience*, Frontiers, p. 603, 2021.

[Zhang et al. 2021]ZHANG, A. et al. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.

[Zhao, Zhang e Zheng 2019]ZHAO, Z.; ZHANG, Y.; ZHENG, Q. Fast localization algorithm of eye centers based on improved hough transform. In: IEEE. *2019 IEEE 7th International Conference on Bioinformatics and Computational Biology (ICBCB)*. [S.l.], 2019. p. 185–189.

[Zhao et al. 2019]ZHAO, Z.-Q. et al. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, IEEE, v. 30, n. 11, p. 3212–3232, 2019.

Instruções para Download do Conjunto de Dados

A plataforma para análise dos dados do glaucoma e o link para download do conjunto de dados, está disponível em:

<https://ronaldocosta.inf.ufg.br/eyes>

A Figura A.1 ilustra o processo de acesso para download. Inicialmente, ao selecionar a Figura (A) presente no website, o usuário é direcionado para a tela de autenticação (B). Nessa tela, deve-se clicar no botão verde identificado como "Ainda não sou parceiro", conforme sinalizado pela seta. Após essa ação, o usuário é encaminhado para a página em desenvolvimento da plataforma. É nesta página que se encontra o link para efetuar o download (C).

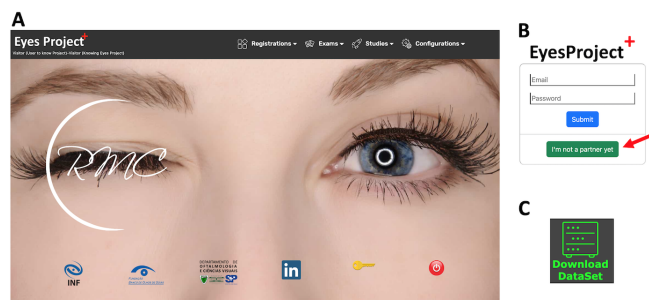


Figura A.1: (A) Link para acesso à plataforma; (B) Tela de acesso para parceiros e não parceiros; (C) Ícone para download do conjunto de dados.

Algoritmo Customizado para Filtragem do Sinal Pupilar

Esta seção do apêndice apresenta o pseudo-código do algoritmo especialmente projetado para filtragem de sinal pupilar.

Algorithm 1 Algoritmo específico para filtragem de sinal.

```

Result: Sinal filtrado
variacao_maxima ← 3;
Repetições ← 100;
sinal_filtrado ← sinal_original;
if discontinuidade(sinal_filtrado, FATOR_DE_DISCONTINUIDADE) is Verdadeiro then
|   return sinal_filtrado
else
|   for  $i \leftarrow 0$  to repeticoes do
|   |   foreach elemento no sinal_filtrado do
|   |   |   if elemento nao e o ultimo e elemento > -1 then
|   |   |   |   if diferenca(elemento, proximo elemento) > variacao_maxima then
|   |   |   |   |   proximo elemento ← -1 foreach subjanela in sublista_sinal_filtrado from proximo elemento do
|   |   |   |   |   |   if difference(elemento, subjanela) > variacao_maxima then
|   |   |   |   |   |   |   proximo proximo elemento ← -1
|   |   |   |   |   |   |   else
|   |   |   |   |   |   |   |   break
|   |   |   |   |   |   |   end
|   |   |   |   |   |   end
|   |   |   |   |   end
|   |   |   |   end
|   |   |   end
|   |   end
|   |   if discontinuidade(trecho_sinal, FATOR_DE_DISCONTINUIDADE) is Verdadeiro then
|   |   |   variacao_maxima ← variacao_maxima + 1
|   |   |   sinal_filtrado ← sinal_original;
|   |   else
|   |   |   break
|   |   end
|   end
|   end
end

```

Arquitetura das Redes Neurais Usadas como Classificadores

Esta seção apresenta diagramas ilustrando a arquitetura de três redes neurais utilizadas como classificadores neste estudo: uma Rede Totalmente Conectada (FCN), uma Rede Neural Convolucional Unidimensional (1D-CNN) e uma Rede Neural Transformer.

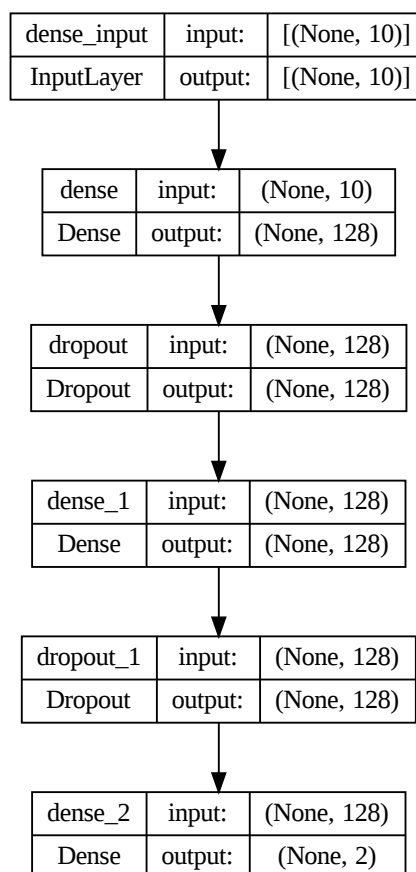


Figura C.1: A arquitetura do classificador Rede Neural Totalmente Conectada (do inglês, FCN).

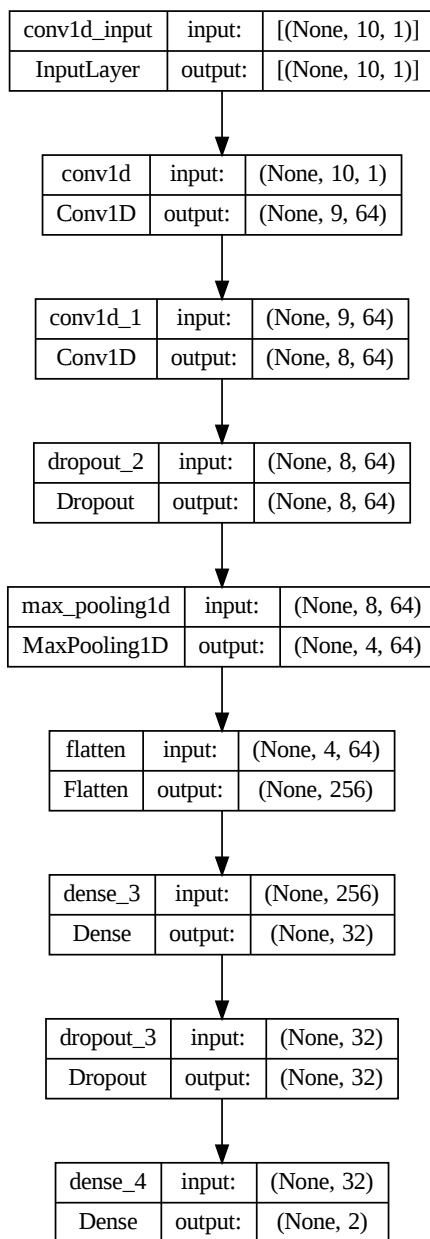


Figura C.2: A arquitetura do classificador de Rede Neural Convolutacional Unidimensional (do inglês, 1D-CNN).

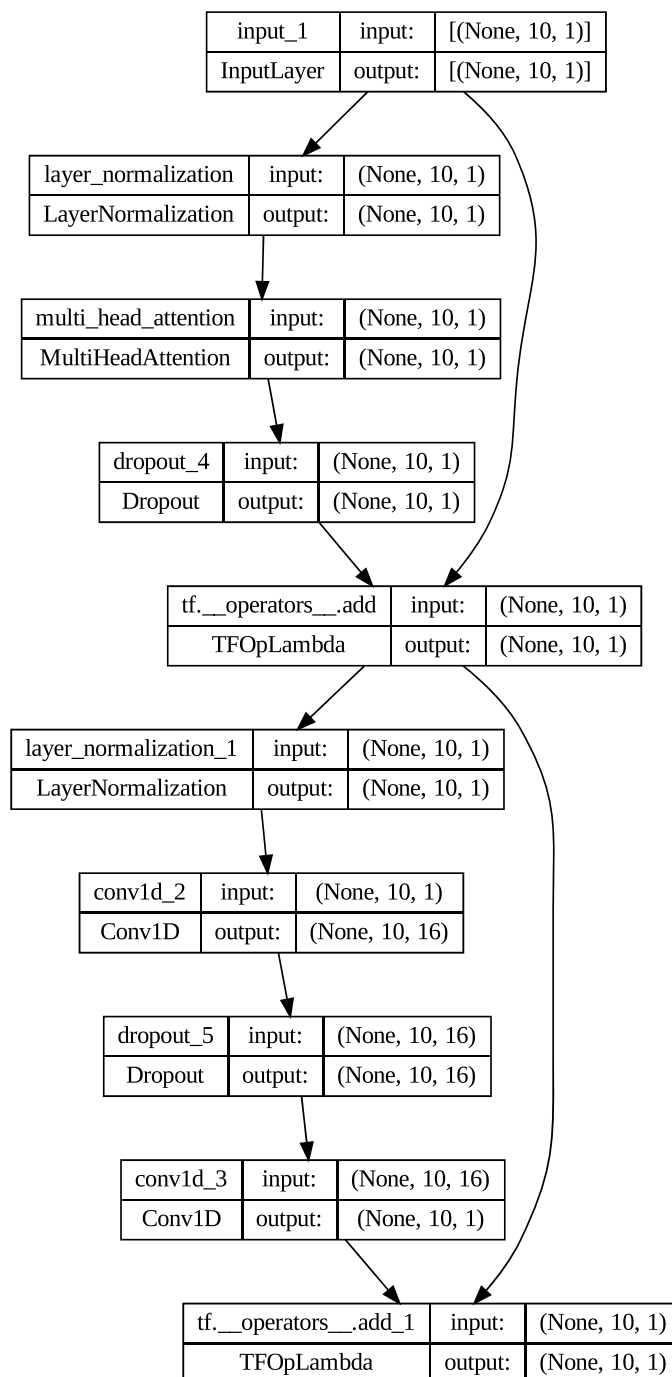


Figura C.3: A arquitetura da primeira cabeça do classificador de Rede Neural Transformer (TRANS).

Contribuições Científicas

Durante a realização desta pesquisa as principais contribuições apresentadas à comunidade científica, com vistas ao aprimoramento do uso do reflexo pupilar à luz no auxílio ao diagnóstico, foram:

1. Contribuição para o desenvolvimento de um portal destinado ao armazenamento, processamento e classificação de exames, construído com o uso do framework Django. Esta plataforma facilita a organização dos dados, a manipulação dos exames (abrangendo pré-processamento, extração de características e classificação) e a execução de experimentos para a classificação desses exames. Acessível em: <https://ronaldocosta.inf.ufg.br/sites/ronaldomc/pages/eyes-en.html>
2. A criação e disponibilização para a comunidade científica de uma base de dados de imagens contendo 10.000 imagens anotadas da pupila e 258.790 imagens não anotadas. Acessível em: <https://ronaldocosta.inf.ufg.br/eyesproject3/dataset10kcreate/>
3. Publicação em coautoria do artigo científico: “*Computer Vision Based Systems for Human Pupillary Behavior Evaluation: A Systematic Review of the Literature.*” (2019), na 43ª Conferência Anual do COMPSAC (Computer Software and Applications Conference). DOI: <https://doi.org/10.1109/COMPSAC.2019.00107>
4. Publicação em coautoria do artigo científico: “*Techniques and Equipment for Automated Pupillometry and its Application to Aid in the Diagnosis of Diseases: A Literature Review.*” (2020), na 44ª Conferência Anual do COMPSAC (Computer Software and Applications Conference). DOI: <https://doi.org/10.1109/COMPSAC48688.2020.00-55>
5. O registro da patente de um dispositivo pupilômetro no Instituto Nacional de Propriedade Industrial (INPI), sob o número de protocolo BR1020190076224, datado de 2020.
6. Publicação do artigo científico: “*Pupillary light reflex as a diagnostic aid from computational viewpoint: A systematic literature review.*” (2021), no periódico *Journal of Biomedical Informatic - JBI*. DOI: <https://doi.org/10.1016/j.jbi.2021.103757> referência nesta tese: [Pinheiro e Costa 2021]

7. Publicação em coautoria do artigo científico: “*An Intelligent System to Improve Diagnostic Support for Oral Squamous Cell Carcinoma.*” (2023), no periódico *Healthcare* DOI: <https://doi.org/10.3390/healthcare11192675>
8. Coautoria na apresentação do trabalho: “*Um sistema Inteligente para suporte ao diagnóstico precoce de carcinoma espinocelular oral com redes neurais artificiais.*” (2024), na forma Pôster, na modalidade Mostra da Pós-Graduação Stricto Sensu e Lato Sensu, no 20º Congresso de Pesquisa, Ensino e Extensão - CONPEEX.
9. Apresentação do trabalho: “*Análise computacional do Reflexo pupilar à luz como ferramenta de auxílio a diagnósticos: uma revisão sistemática da literatura.*” (2024), na forma Pôster, na modalidade Mostra da Pós-Graduação Stricto Sensu e Lato Sensu, no 20º Congresso de Pesquisa, Ensino e Extensão - CONPEEX.
10. Submissão do artigo científico “*A Pupillary Image Dataset: 10,000 Annotated and 258,790 Non-Annotated Images of Patients with Glaucoma, Diabetes, and Subjects Influenced by Alcohol, Coupled with a Segmentation Performance Evaluation.*” (2024), na revista *Computers in Biology and Medicine - CIBM*. [Terceira rodada de revisão] <https://track.authorhub.elsevier.com/?uuid=5632356a-62bf-4b5e-bbca-03fb64b374ad>
11. Publicação do artigo científico “*Evaluating Machine Learning Techniques for Enhanced Glaucoma Screening through Pupillary Light Reflex Analysis.*” (2024), na revista *Array*. DOI: <https://doi.org/10.1016/j.array.2024.100359>