

João Guilherme Marcondes de Souza Costa

**Detecção de Diabetes por Meio da Análise de
Retinas:
Uma Abordagem de Inferência Causal**

Belo Horizonte

2024

João Guilherme Marcondes de Souza Costa

**Detecção de Diabetes por Meio da Análise de Retinas:
Uma Abordagem de Inferência Causal**

Monografia apresentada durante o Seminário dos Trabalhos de Conclusão do Curso de Graduação em Engenharia Elétrica da UFMG, como parte dos requisitos necessários à obtenção do título de Engenheiro Eletricista.

Universidade Federal de Minas Gerais – UFMG

Escola de Engenharia

Curso de Graduação em Engenharia Elétrica

Orientador: Prof. Anisio Mendes Lacerda

Belo Horizonte

2024

Lista de abreviaturas e siglas

Fig. Area of the i^{th} component

List of symbols

Γ Greek letter Gamma

Λ Lambda

ζ Greek letter minuscule zeta

\in Pertains

Sumário

1	INTRODUÇÃO	9
1.1	Motivação e Justificativa	9
1.2	Objetivos	10
1.3	Formulação do Problema	11
1.4	Trabalhos Correlatos	11
1.5	Organização do Trabalho	12
2	REVISÃO DE LITERATURA	13
2.1	Análise de Retinas e Métodos Tradicionais para Detecção de Diabetes	13
2.2	Avanços Computacionais na Análise de Imagens de Retina	14
2.2.1	Visão Computacional e Aprendizado de Máquina	14
2.2.2	Bases de Dados Disponíveis	15
2.3	Introdução e Aplicações de Inferência Causal	16
2.3.1	Fundamentos da Inferência Causal	16
2.3.2	Integração com Visão Computacional	16
2.3.3	Aplicação do Slice Discovery Method na Análise de Retinas	17
2.4	Limitações na Análise de Imagens de Retina e Desafios Atuais	17
2.4.1	Shortcut Learning e suas Implicações	17
2.4.2	Outras Limitações dos Modelos Atuais	18
2.4.3	A Necessidade de Modelos Mais Robustos e Interpretáveis	19
3	METODOLOGIA	21
3.1	Descrição do Produto	21
3.2	Materiais Utilizados	21
3.2.1	Bases de Dados	21
3.2.2	Ferramentas e Tecnologias	21
3.3	Metodologia para Desenvolvimento	22
3.3.1	Pré-processamento das Imagens	22
3.3.2	Modelo Desenvolvido	23
3.4	Metodologia de Testes e Avaliação	23
3.4.1	Métricas de Avaliação	23
3.4.2	Protocolos de Teste	24
3.5	Reprodutibilidade	25
4	RESULTADOS E DISCUSSÃO	27
4.1	Resultados de Desempenho	27

4.2	Análise Comparativa com o Estado da Arte	28
4.3	Outras Observações	29
4.4	Desdobramentos Futuros	30
5	CONCLUSÃO	33
5.1	Resumo do Trabalho	33
5.2	Principais Contribuições	33
5.3	Limitações do Trabalho	33
5.4	Trabalhos Futuros	34
5.5	Considerações Finais	34
	REFERÊNCIAS	37
	REFERÊNCIAS	39

1 Introdução

A diabetes é uma doença crônica que faz com que o corpo não produza insulina ou não consiga empregar adequadamente a insulina que produz, assim não controlando o nível de glicose no sangue e causando severas complicações nas pessoas que possuem a doença. Ela é uma das doenças crônicas mais recorrentes no mundo e apresenta crescentes índices de complicações associadas. O cenário atual da diabetes no Brasil é alarmante e requer atenção especial. De acordo com a Sociedade Brasileira de Diabetes a doença afeta mais de 13 milhões de brasileiros, sendo assim, o país ocupa a quinta posição global em número de casos, com um aumento expressivo nos últimos anos, especialmente entre os jovens. Esse crescimento está fortemente associado a fatores como sedentarismo, má alimentação e fatores genéticos, que contribuem para o desenvolvimento da doença.

A detecção precoce e o monitoramento contínuo da doença são fundamentais para reduzir as complicações graves, como doenças cardíacas, insuficiência renal e problemas de visão, que prejudicam seriamente a qualidade de vida das pessoas afetadas. A retinopatia diabética, que é o foco deste trabalho, é uma das principais causas de cegueira irreversível, e sua identificação precoce pode impedir o agravamento da condição. Muitos casos da doença não são diagnosticados até que estejam em estágios avançados, quando os sintomas já são perceptíveis, o que limita as opções de tratamento e aumenta os custos de cuidados de saúde para o paciente.

Nos últimos anos, métodos de aprendizado de máquina e inteligência artificial têm sido testados na medicina para auxiliar no diagnóstico e no tratamento de doenças. Uma área promissora para este campo é a utilização de inferência causal, que permite modelar relações de causa e efeito entre variáveis, identificando padrões que não seriam facilmente detectados por métodos tradicionais, apenas por profissionais experientes e especializados. No contexto da diabetes, a inferência causal aplicada à análise de imagens da retina tem o potencial de oferecer uma abordagem inovadora para detectar precocemente sinais de alterações associadas à doença, como mudanças nos vasos sanguíneos e a própria retinopatia diabética. A combinação de técnicas de aprendizado de máquina com dados oftalmológicos pode representar uma solução eficaz e mais acessível para a detecção precoce da diabetes.

1.1 Motivação e Justificativa

O diagnóstico precoce da diabetes é crucial para um início de tratamento o quanto antes, que aumenta a sua eficácia e a prevenção de complicações mais graves. A retinopatia diabética, um dos primeiros sinais visíveis da doença, pode ser identificada por meio

de exames oftalmológicos, mas a análise dessas imagens requer expertise médica e é demorada. A automação desse processo pode reduzir significativamente o tempo necessário para diagnóstico, além de torná-lo mais acessível e menos dependente de especialistas disponíveis para esta finalidade. A utilização de técnicas de inferência causal, que exploram as relações de causa e efeito entre as variáveis observadas, possibilita a criação de um modelo robusto, capaz de identificar padrões complexos em imagens da retina e fornecer um diagnóstico mais rápido e preciso.

A motivação para a implementação de métodos de inferência causal na detecção da diabetes baseia-se na crescente necessidade de ferramentas de diagnóstico que sejam tanto rápidas quanto precisas, a fim de detectar alterações na retina, um sintoma que se manifesta antes, que possam indicar o risco de diabetes antes que os sintomas mais evidentes se manifestem. Ao aplicar técnicas avançadas de aprendizado de máquina e inferência causal, é possível não só melhorar a acuracidade do diagnóstico, mas também entender melhor as interações entre diferentes fatores que podem influenciar o desenvolvimento da doença. Essa abordagem poderia ser aplicada de forma escalável, contribuindo para a implementação de estratégias de saúde pública mais eficazes.

1.2 Objetivos

O objetivo principal deste trabalho é desenvolver e aplicar um modelo baseado em inferência causal para a detecção de sintomas de diabetes a partir da análise de imagens da retina dos pacientes. Através da identificação de padrões e relações causais nas imagens, espera-se poder identificar alterações iniciais que possam indicar o risco de desenvolvimento da diabetes, antes que os sintomas clínicos mais graves comecem a ser apresentados.

Além disso, os objetivos específicos deste trabalho incluem:

- Realizar uma análise detalhada das características da retina que podem estar relacionadas ao desenvolvimento da diabetes, utilizando imagens oftalmológicas.
- Implementar técnicas de inferência causal para modelar a relação entre as alterações na retina e o risco de diabetes, explorando como variáveis diferentes influenciam a condição.
- Avaliar a precisão e a eficácia do modelo proposto, comparando-o com métodos tradicionais de diagnóstico da diabetes, como exames laboratoriais e avaliações clínicas.
- Investigar o potencial de aplicação dessa metodologia em sistemas de saúde com recursos limitados, permitindo o monitoramento remoto e a detecção precoce da diabetes.

- Ao atingir esses objetivos, espera-se contribuir para a melhoria do diagnóstico precoce da diabetes, proporcionando uma ferramenta acessível e eficaz para médicos e profissionais de saúde.

1.3 Formulação do Problema

A detecção precoce da diabetes é um desafio complexo, pois muitas vezes os sintomas iniciais não são facilmente identificáveis e podem passar despercebidos até que a doença esteja em estágios mais avançados. No entanto, a retina apresenta sinais visíveis de alterações que podem indicar o risco de diabetes, como a retinopatia diabética, que é uma complicação comum da doença. A identificação desses sinais, com o auxílio de imagens oftalmológicas, oferece uma oportunidade de diagnóstico precoce, mas a análise manual dessas imagens exige tempo e experiência.

A utilização de técnicas automatizadas, como a inferência causal, pode melhorar significativamente o processo de diagnóstico ao permitir que modelos computacionais identifiquem relações entre variáveis de forma mais eficiente, rápida e precisa. No entanto, o problema reside na dificuldade de modelar as complexas interações entre as diversas variáveis que influenciam o desenvolvimento da diabetes e suas manifestações na retina. A inferência causal, embora poderosa, apresenta desafios em sua aplicação prática, especialmente no contexto médico, devido à necessidade de dados de alta qualidade e ao entendimento profundo das relações causais entre fatores biológicos, genéticos e ambientais.

Dessa forma, a formulação deste problema envolve a criação de um modelo capaz de detectar e prever o risco de diabetes a partir da análise de imagens da retina, utilizando métodos de inferência causal para identificar padrões e relações que possam ser indicativos de alterações associadas à doença. O desafio está em desenvolver um modelo robusto e preciso que seja capaz de lidar com a complexidade dos dados e oferecer uma ferramenta eficaz para a detecção precoce da diabetes.

1.4 Trabalhos Correlatos

Diversos estudos recentes têm explorado a inferência causal na área da saúde, especialmente para a identificação de fatores de risco em doenças crônicas, mas também diversos ramos do aprendizado de máquina, como podemos ver em trabalhos como "O aprendizado de máquina aplicado ao auxílio do diagnóstico da doença de Parkinson". Em relação à diabetes, alguns trabalhos utilizaram modelos preditivos, como reconhecimento de padrões com vetores de suporte e redes neurais, que são eficientes em apontar associações, mas carecem de mecanismos para estabelecer relações causais.

Pesquisas utilizando técnicas como variáveis instrumentais e análise de sensibilidade têm mostrado avanços na compreensão dos fatores de risco de diabetes focadas no tipo 2 utilizando variantes genéticas como variáveis instrumentais para avaliar a influência causal de fatores como dieta e níveis de ácido úrico no desenvolvimento do diabetes. No entanto, a aplicação de métodos de inferência causal mais sofisticados, como modelos baseados em machine learning para causalidade, ainda é um campo emergente. Este trabalho busca preencher essa lacuna, comparando abordagens tradicionais com métodos causais para verificar sua eficiência na detecção de diabetes.

1.5 Organização do Trabalho

O texto do trabalho está organizado da seguinte forma:

- Capítulo 2: Descreve o contexto social e clínico da diabetes e a importância da inferência causal para a detecção precoce, abordando implicações para a prática médica e para a saúde pública.
- Capítulo 3: Apresenta o referencial teórico sobre inferência causal, discutindo metodologias como variáveis instrumentais, modelos causais em machine learning e sua aplicabilidade à saúde.
- Capítulo 4: Detalha os dados utilizados, as variáveis e os procedimentos metodológicos aplicados para desenvolver o modelo causal. Inclui o pré-processamento e a seleção das variáveis causais.
- Capítulo 5: Analisa os resultados obtidos, comparando o modelo causal com métodos preditivos tradicionais e discutindo a relevância dos fatores causais identificados.
- Capítulo 6: Apresenta as considerações finais, destacando as contribuições da inferência causal para a detecção de diabetes e sugestões para trabalhos futuros.

2 Revisão de Literatura

Este capítulo apresenta uma revisão da literatura da detecção de diabetes por meio de imagens de retina, com foco na utilização de inferência causal para melhorar a interpretação e confiabilidade dos resultados. A seguir, são discutidos os principais métodos existentes, suas limitações e as oportunidades para avanços, com o objetivo de compreender, identificar lacunas no conhecimento e contextualizar a necessidade deste trabalho.

2.1 Análise de Retinas e Métodos Tradicionais para Detecção de Diabetes

A análise de retinas é amplamente reconhecida como uma ferramenta eficaz para o diagnóstico precoce e monitoramento da diabetes, devido à sua capacidade de revelar alterações microvasculares relacionadas à doença. A retina, sendo um tecido altamente vascularizado, reflete diretamente as condições do sistema circulatório, permitindo a identificação de características associadas a diabetes, como microaneurismas, exsudatos duros, hemorragias e neovascularização (????).

Os métodos tradicionais mais utilizados para avaliação das retinas incluem a oftalmoscopia direta e indireta, a retinografia e a angiografia fluoresceínica. Esses métodos, embora eficazes, apresentam limitações significativas:

- **Oftalmoscopia Direta:** A oftalmoscopia direta é um exame oftalmológico realizado com o auxílio de um oftalmoscópio, um instrumento composto por uma lâmpada de luz e lentes. O exame permite a visualização direta do fundo do olho, incluindo a retina, nervo óptico e vasos sanguíneos. Durante o exame, o oftalmoscópio é colocado próximo ao olho do paciente, e o médico observa a imagem refletida da retina por meio das lentes. A oftalmoscopia direta é especialmente útil para detectar alterações como edema da retina e vasos sanguíneos anormais, comuns em pacientes diabéticos. No entanto, o método depende da habilidade do profissional e pode ser limitado pela visão periférica reduzida e pela falta de profundidade na visualização (??).
- **Oftalmoscopia Indireta:** A oftalmoscopia indireta, por sua vez, utiliza uma fonte de luz mais potente e uma lente convexa montada em um conjunto de óculos, permitindo uma visão mais ampla e detalhada da retina. O médico usa um iluminador na forma de uma luz projetada sobre o fundo do olho, e a lente especial permite observar a retina em uma imagem tridimensional. Esse exame oferece uma visão

mais abrangente e detalhada, sendo capaz de identificar anomalias mais sutis na retina, como alterações vasculares que indicam estágios iniciais da diabetes. A oftalmoscopia indireta é especialmente eficaz para visualizar áreas periféricas da retina e é preferida para a avaliação de patologias mais complexas. Contudo, esse exame também é altamente dependente da experiência do profissional, e a qualidade da análise pode variar (??).

- **Retinografia:** Utiliza câmeras específicas para capturar imagens detalhadas do fundo do olho. Embora forneça uma documentação permanente das condições da retina, a análise das imagens ainda depende, em grande parte, de interpretação humana, o que pode introduzir vieses e limitar sua escalabilidade (??).
- **Angiografia Fluoresceínica:** Considerada o padrão-ouro para a avaliação de alterações vasculares, este método envolve a injeção de contraste fluorescente para visualizar os vasos sanguíneos da retina. No entanto, é um procedimento invasivo, caro e não isento de riscos, como reações alérgicas ao contraste (??).

Esses métodos têm desempenhado um papel essencial no manejo clínico da diabetes, mas são insuficientes para atender às demandas de triagem em larga escala, especialmente em populações de baixa renda ou áreas remotas. Além disso, as abordagens tradicionais carecem de mecanismos para correlacionar padrões observados nas imagens com fatores causais subjacentes, uma lacuna crítica para o desenvolvimento de diagnósticos mais precisos e personalizados (????).

O crescente interesse em soluções computacionais busca superar essas limitações, com foco na automatização e padronização da análise de imagens. No entanto, é importante reconhecer que a adoção dessas novas tecnologias deve ser feita de forma complementar aos métodos tradicionais, integrando suas vantagens para proporcionar diagnósticos mais rápidos e eficientes (????).

2.2 Avanços Computacionais na Análise de Imagens de Retina

2.2.1 Visão Computacional e Aprendizado de Máquina

Nos últimos anos, a aplicação de técnicas de visão computacional e aprendizado de máquina na análise de imagens de retina tem se mostrado uma poderosa ferramenta para a detecção precoce de condições oculares, como a retinopatia diabética. As redes neurais convolucionais (CNNs) são amplamente utilizadas para realizar tarefas complexas, como segmentação de vasos sanguíneos, detecção de microaneurismas e classificação de diferentes estágios da retinopatia. Essas redes são capazes de aprender e identificar padrões em grandes volumes de dados com alta precisão, reduzindo a subjetividade dos diagnósticos e

aumentando a eficiência dos processos de triagem. A eficácia desses modelos é destacada por Rajalakshmi et al. (2020) (??), que realiza uma revisão aprofundada sobre o uso do aprendizado profundo na detecção da retinopatia diabética.

A combinação de aprendizado supervisionado com redes neurais profundas tem sido uma das abordagens mais promissoras. Algoritmos de CNNs, ao serem treinados com grandes bases de dados de imagens rotuladas, podem alcançar precisão comparável ou até superior à de oftalmologistas experientes. Além disso, esses métodos oferecem a possibilidade de diagnóstico automatizado e acessível, tornando-os extremamente valiosos tanto em contextos clínicos quanto em áreas de difícil acesso, onde a escassez de especialistas limita a realização de exames detalhados (??).

Apesar dos avanços, ainda existem desafios no uso dessas técnicas, especialmente no que se refere à explicabilidade dos modelos e à capacidade de identificar padrões causais nas imagens. O Slice Discovery Method (SDM), uma abordagem emergente no campo da análise de dados, oferece uma solução interessante para esses problemas. O método, explorado por Olesen et al. (2024) (??), permite a descoberta de subgrupos específicos em conjuntos de dados complexos e heterogêneos, contribuindo para a identificação de padrões ocultos diretamente relacionados à progressão do diabetes. Quando aplicada à análise de imagens de retina, a integração do SDM pode melhorar significativamente a detecção precoce e a personalização do tratamento, ao identificar variáveis subjacentes que poderiam ser negligenciadas por métodos tradicionais.

2.2.2 Bases de Dados Disponíveis

O desempenho de modelos de aprendizado de máquina, especialmente os baseados em redes neurais profundas, depende crucialmente da disponibilidade de grandes volumes de dados rotulados para treinamento. A qualidade e a diversidade das bases de dados são fatores determinantes para a eficácia dos modelos na detecção de retinopatias e outras doenças oculares. Um exemplo fundamental é o EyePACS, uma base de dados amplamente utilizada para a análise de imagens de retina, que contém um grande número de imagens rotuladas de retinopatias diabéticas em diferentes estágios de progressão (??). O EyePACS fornece um banco de imagens ricas e variadas, tornando-se um recurso essencial para o treinamento e validação de modelos de aprendizado de máquina.

A diversidade nas imagens, no entanto, ainda apresenta desafios, como a representatividade de diferentes grupos demográficos e variações nos métodos de captura das imagens. Outras bases de dados, como Messidor e DIARETDB1, também são amplamente empregadas, mas apresentam limitações, especialmente em relação à heterogeneidade da população e à quantidade de imagens com estagiamentos mais avançados da doença. A utilização do Slice Discovery Method em conjunto com essas bases de dados pode trazer avanços significativos. O SDM tem o potencial de identificar subgrupos não previamente

detectados, extraindo padrões causais a partir de dados aparentemente não relacionados, melhorando a acurácia dos modelos e permitindo diagnósticos mais precisos e personalizados (??).

2.3 Introdução e Aplicações de Inferência Causal

2.3.1 Fundamentos da Inferência Causal

A inferência causal é uma área da estatística que visa entender e modelar as relações de causa e efeito entre variáveis. Em contraste com abordagens correlativas tradicionais, que apenas identificam associações entre variáveis, a inferência causal busca identificar como uma variável pode efetivamente influenciar ou causar a alteração de outra. Isso é essencial para desenvolver intervenções mais eficazes e para entender os mecanismos subjacentes aos fenômenos observados. Entre as principais ferramentas de inferência causal, destacam-se os grafos acíclicos direcionados (DAGs) e o do-calculus, ambos utilizados para modelar e analisar as interações entre as variáveis, permitindo uma interpretação mais clara das relações causais (??).

No contexto da análise de imagens de retina, essas ferramentas são fundamentais para entender como alterações visíveis na retina, como microaneurismas, exsudatos e hemorragias, estão relacionadas causalmente à diabetes. A análise causal não se limita a observar correlações entre características da retina e a doença, mas também busca identificar os mecanismos subjacentes que contribuem para a progressão da retinopatia diabética. A inferência causal pode, portanto, ser crucial para aprimorar a precisão dos diagnósticos e ajudar na previsão da progressão da doença, permitindo uma abordagem mais personalizada de tratamento (??).

2.3.2 Integração com Visão Computacional

A combinação de inferência causal com técnicas de aprendizado de máquina, especialmente aquelas baseadas em visão computacional, tem ganhado destaque na área da medicina. Essa integração oferece um grande potencial para aumentar a interpretabilidade dos modelos preditivos, um desafio significativo em muitos sistemas de inteligência artificial, como as redes neurais profundas, que frequentemente operam como caixas-pretas. Incorporando a inferência causal, é possível dar sentido às decisões feitas pelos modelos de aprendizado de máquina, explicando por que uma imagem de retina foi classificada de determinada maneira ou quais características específicas indicam um risco maior de progressão do diabetes (??).

O Slice Discovery Method (SDM) desempenha um papel central nessa integração, permitindo que padrões causais sejam identificados em segmentos específicos dos dados,

ou seja, em "fatias" ou subgrupos de dados que podem revelar nuances não aparentes em uma análise global. Por exemplo, ao aplicar o SDM, pode-se descobrir que certos padrões visuais nas imagens de retina são causais para a progressão da doença em grupos específicos de pacientes, como aqueles com diabetes tipo 1 ou em estágios iniciais da doença (??). Essa análise mais detalhada não apenas aumenta a confiança nos diagnósticos feitos por sistemas automatizados, mas também possibilita uma análise mais personalizada, alinhada aos princípios da medicina de precisão.

2.3.3 Aplicação do Slice Discovery Method na Análise de Retinas

O Slice Discovery Method tem se mostrado uma abordagem altamente eficaz na análise de dados heterogêneos, como as imagens de retina utilizadas para diagnosticar retinopatia diabética. Uma das principais vantagens do SDM é sua capacidade de lidar com grandes volumes de dados e identificar subgrupos específicos ou interações causais que poderiam ser ignoradas por métodos tradicionais. No caso das imagens de retina, o SDM pode identificar padrões ocultos e complexos relacionados à progressão do diabetes, permitindo que novos biomarcadores sejam descobertos (??).

Além disso, o SDM facilita a personalização dos diagnósticos, permitindo a segmentação dos dados em subgrupos com características similares e comportamentos causais compartilhados. Esse processo de segmentação pode levar a diagnósticos mais precisos e ao desenvolvimento de tratamentos mais eficazes, adequados ao perfil de cada paciente. Ao identificar essas interações complexas e causais, o Slice Discovery Method não só contribui para o aprimoramento dos modelos de diagnóstico, mas também oferece uma visão mais profunda e compreensiva sobre os fatores que influenciam a progressão do diabetes e suas manifestações na retina (??).

2.4 Limitações na Análise de Imagens de Retina e Desafios Atuais

Embora os avanços na aplicação de técnicas de aprendizado de máquina e visão computacional tenham demonstrado um grande potencial na análise de imagens de retina para detecção de diabetes, ainda existem várias limitações e desafios que precisam ser superados. Entre esses, destaca-se o fenômeno do shortcut learning, que tem se mostrado uma grande preocupação em modelos de aprendizado supervisionado, especialmente em tarefas complexas como a análise de imagens médicas.

2.4.1 Shortcut Learning e suas Implicações

O shortcut learning ocorre quando um modelo aprende a identificar padrões de maneira superficial, utilizando características que não são de fato indicativas de uma relação causal, mas sim coincidências ou artefatos presentes nos dados. Em vez de aprender

os fatores fundamentais que realmente influenciam a progressão da doença, o modelo pode "aprender" atalhos (shortcuts) que são mais fáceis, mas enganosos. Esse fenômeno foi discutido no trabalho de Olesen et al. (2024), em que os autores destacam como o Slice Discovery Method (SDM) pode ajudar a evitar tais armadilhas, ao identificar padrões causais reais nos dados de maneira mais transparente e fundamentada (??).

Por exemplo, em um sistema de diagnóstico de retinopatia diabética, um modelo de aprendizado profundo pode acabar associando a presença de determinadas características nas imagens (como a intensidade da cor ou a textura de certas áreas) a um diagnóstico positivo, quando, na realidade, essas características podem ser resultado de outros fatores não relacionados diretamente à progressão da doença, como variações na qualidade das imagens ou a presença de artefatos.

Esse tipo de aprendizado é problemático porque pode levar a diagnósticos incorretos, especialmente em ambientes clínicos, onde a precisão é crucial. Além disso, o shortcut learning compromete a generalização do modelo para novos dados, tornando-o menos robusto e, portanto, inadequado para aplicações em larga escala. A abordagem proposta por Olesen et al. (2024) (??) busca mitigar esse risco ao fornecer uma metodologia que permite entender melhor as limitações do modelo e ajustar seus parâmetros de forma mais eficaz.

2.4.2 Outras Limitações dos Modelos Atuais

Além do shortcut learning, existem várias outras limitações que afetam a análise de imagens de retina, principalmente quando aplicada à detecção de diabetes. Uma dessas limitações é a falta de explicabilidade nos modelos de aprendizado profundo. Embora as redes neurais convolucionais (CNNs) sejam extremamente eficazes em tarefas de classificação e segmentação, elas são frequentemente vistas como caixas-pretas, ou seja, é difícil entender por que o modelo tomou uma determinada decisão. Essa falta de transparência é especialmente problemática em contextos médicos, onde os profissionais de saúde precisam de explicações claras e fundamentadas para tomar decisões clínicas.

O fenômeno de falta de explicabilidade foi amplamente discutido no artigo de Ribeiro et al. (2016), que introduziu o conceito de LIME (Local Interpretable Model-agnostic Explanations) para criar explicações interpretáveis para modelos de aprendizado de máquina em áreas críticas, como a medicina (??). Esse tipo de metodologia pode ser integrado a sistemas de diagnóstico automatizados de retinopatias diabéticas, garantindo que o profissional de saúde compreenda como as conclusões foram alcançadas.

Outra limitação significativa é a variedade e a qualidade das bases de dados utilizadas para treinar os modelos. Embora bases de dados públicas como EyePACS e Messidor tenham sido amplamente utilizadas, elas ainda possuem limitações, como a falta de diver-

sidade populacional e a presença de desequilíbrios nos dados. Isso pode levar a modelos que funcionam bem em populações específicas, mas que apresentam um desempenho inferior quando aplicados a grupos diferentes. A diversidade das imagens de retina, incluindo variações nas características de iluminação, resolução e qualidade das imagens, também representa um desafio para a construção de modelos robustos.

A questão das bases de dados desequilibradas e sua influência nos modelos de aprendizado de máquina foi abordada por Alsubaie et al. (2021), que discutiram métodos de balanceamento de dados e a importância de bases de dados diversificadas para melhorar a generalização de modelos em diagnósticos médicos (??). A integração de técnicas como o Slice Discovery Method, mencionada por Olesen et al. (2024), pode ajudar a identificar subgrupos específicos nos dados que são relevantes para o diagnóstico de retinopatias em populações diversas, mitigando o impacto da falta de diversidade nas bases de dados (??).

Além disso, a heterogeneidade dos dados é outro desafio importante. Pacientes com diabetes podem apresentar diferentes estágios de progressão da doença, e a manifestação de retinopatia diabética pode variar entre os indivíduos. Isso cria uma grande variabilidade nas imagens de retina, o que dificulta a criação de modelos que sejam precisos e generalizáveis para todos os casos. Técnicas como o *Slice Discovery Method* podem ajudar a lidar com essa heterogeneidade, permitindo identificar subgrupos específicos e padrões causais que não seriam capturados por modelos tradicionais.

2.4.3 A Necessidade de Modelos Mais Robustos e Interpretáveis

Para superar essas limitações, é essencial que os modelos de aprendizado de máquina e visão computacional sejam mais robustos e interpretáveis. A explicabilidade é uma área de pesquisa crescente, e métodos como as Redes Neurais Explicáveis (XAI) têm sido propostos para oferecer uma maior compreensão sobre como os modelos chegam às suas conclusões. Esses modelos permitem que os médicos comprehendam melhor os fatores que influenciam o diagnóstico e, assim, possam confiar mais nos sistemas automatizados.

Além disso, é crucial que os modelos sejam treinados em bases de dados mais diversas e representativas, com um número maior de imagens de retina que capturem uma gama mais ampla de características, incluindo variações nas condições de iluminação, resolução e qualidade das imagens. O desenvolvimento de bases de dados balanceadas e com maior diversidade é fundamental para melhorar a generalização dos modelos e garantir que eles funcionem bem em populações de pacientes diversas.

Por fim, o Slice Discovery Method se apresenta como uma abordagem promissora para superar algumas dessas limitações. Ao identificar subgrupos específicos e padrões causais nos dados, o SDM pode contribuir para a criação de modelos mais precisos, robustos e interpretáveis, permitindo uma análise mais aprofundada e personalizada das

imagens de retina.

3 Metodologia

Este capítulo descreve detalhadamente os materiais utilizados, as metodologias empregadas para o desenvolvimento do sistema de análise de imagens de retina, e os métodos de avaliação e teste empregados para garantir a reproduzibilidade e a robustez dos resultados obtidos. A abordagem adotada visa não apenas o desenvolvimento de uma solução precisa, mas também a criação de um processo que possa ser reproduzido em diferentes ambientes e por outros pesquisadores, assegurando a transparência do trabalho.

3.1 Descrição do Produto

O produto desenvolvido neste trabalho é um sistema avançado baseado em visão computacional, com foco na análise automatizada de imagens de retina. Seu principal objetivo é a identificação de sinais de diabetes, como microaneurismas, hemorragias e exsudatos, utilizando algoritmos de aprendizado profundo para detectar padrões específicos nas imagens. A análise dessas características é essencial para o diagnóstico precoce da retinopatia diabética, uma complicação comum associada ao diabetes. O sistema foi projetado para ser uma ferramenta complementar ao trabalho de oftalmologistas, proporcionando uma segunda opinião que pode aumentar a precisão diagnóstica e reduzir o tempo necessário para análise.

3.2 Materiais Utilizados

3.2.1 Bases de Dados

A base de dados utilizada neste trabalho foi a **EyePACS**, uma coleção amplamente reconhecida de imagens de retina rotuladas, que se tornou um benchmark no estudo da retinopatia diabética. A base de dados contém imagens de alta qualidade que são essenciais para o treinamento e validação de modelos de aprendizado de máquina voltados para a detecção de doenças oculares. Devido a limitações de hardware, os experimentos foram realizados com subgrupos selecionados do *EyePACS*, permitindo processar as imagens de forma mais eficiente.

3.2.2 Ferramentas e Tecnologias

O desenvolvimento do sistema foi realizado utilizando as seguintes ferramentas e tecnologias:

- **Linguagem de Programação Python:** A escolha do Python deve-se à sua versatilidade e à vasta gama de bibliotecas e frameworks voltados para aprendizado de máquina e visão computacional.
- **Torch e Torchvision:** O Torch foi utilizado como framework principal para a implementação e treinamento do modelo de aprendizado profundo, enquanto o Torchvision forneceu as ferramentas necessárias para o processamento eficiente de imagens e a construção de redes neurais convolucionais.
- **Domino:** O Domino foi empregado como um modelo de *slice discovery*, permitindo o controle de experimentos e a gestão dos recursos computacionais necessários para treinar o modelo. Essa plataforma facilita a experimentação e análise de diferentes configurações, proporcionando insights sobre as variáveis que afetam o desempenho do modelo.
- **Arquitetura ResNet-18:** A rede neural convolucional ResNet-18 foi escolhida como base devido à sua eficiência computacional e sua arquitetura de rede residual, que permite a construção de redes mais profundas sem sofrer de problemas de vanishing gradient.

3.3 Metodologia para Desenvolvimento

3.3.1 Pré-processamento das Imagens

O pré-processamento das imagens foi fundamental para garantir que o modelo pudesse aprender de maneira eficiente e produzir resultados de alta qualidade. O processo de transformação das imagens foi realizado em duas etapas principais: redimensionamento e normalização.

Primeiro, as imagens foram redimensionadas para 256 pixels de largura, mantendo a proporção original. Em seguida, foi realizada uma operação de *center crop*, onde as imagens foram recortadas para o tamanho de 224x224 pixels, garantindo uma dimensão uniforme para a entrada do modelo. Essa abordagem é comumente utilizada em modelos de redes neurais convolucionais, pois ajuda a manter a centralidade e as características mais importantes da imagem.

Após o redimensionamento e recorte, as imagens foram convertidas para tensores, a estrutura de dados utilizada pelo framework PyTorch. Por fim, a normalização foi aplicada, ajustando os valores de cada pixel para que estivessem em uma faixa específica. A normalização foi feita com médias e desvios padrão baseados no conjunto de dados *ImageNet*, que é amplamente utilizado em redes pré-treinadas, como o modelo ResNet-18:

$$\text{Mean} = [0.485, 0.456, 0.406], \quad \text{Std} = [0.229, 0.224, 0.225]$$

Isso garantiu que os dados de entrada tivessem a mesma escala e distribuição dos dados com os quais o modelo foi inicialmente treinado.

O código utilizado para aplicar essas transformações às imagens é o seguinte:

```
# 1. Define transform
transform = transforms.Compose([
    transforms.Resize(256),
    transforms.CenterCrop(224),
    transforms.ToTensor(),
    transforms.Normalize(
        mean=[0.485, 0.456, 0.406],
        std=[0.229, 0.224, 0.225]),
])

# 2. Create new column with transform
df["input"] = df[["img"]].defer(transform)
```

3.3.2 Modelo Desenvolvido

O sistema foi baseado em uma rede neural convolucional (CNN), com a arquitetura adaptada da ResNet-18. O ResNet-18 é um modelo de aprendizado profundo conhecido por sua capacidade de preservar a aprendizagem ao longo de camadas mais profundas, graças ao uso de conexões residuais. Isso permite que a rede aprenda de forma mais eficiente, mesmo em ambientes com recursos limitados. A rede foi inicializada com pesos pré-treinados no *ImageNet*, um grande banco de imagens rotuladas, e ajustada para a tarefa de classificação de imagens de retina. O uso da biblioteca Torchvision proporcionou um conjunto de módulos otimizados que facilitaram a implementação do modelo e garantiram que as operações fossem computacionalmente eficientes.

3.4 Metodologia de Testes e Avaliação

3.4.1 Métricas de Avaliação

A avaliação do desempenho do modelo foi realizada utilizando uma combinação de métricas que fornecem uma visão completa da eficácia do modelo, considerando tanto a acurácia global quanto a capacidade do modelo de lidar com desequilíbrios entre as classes. As principais métricas empregadas são:

- **Precisão (Accuracy):** A precisão é uma métrica amplamente utilizada em tarefas de classificação, representando a proporção de classificações corretas em relação ao número total de amostras processadas. Matematicamente, a precisão é definida como:

$$\text{Accuracy} = \frac{\text{Número de classificações corretas}}{\text{Total de amostras}}$$

Embora seja útil para obter uma visão geral do desempenho do modelo, a precisão pode ser insuficiente em cenários com classes desbalanceadas, onde um modelo pode alcançar alta precisão simplesmente prevendo sempre a classe majoritária, outro problema atacado durante o desenvolvimento do modelo apresentado neste trabalho.

- **F1-score:** O F1-score é uma métrica balanceada que leva em consideração tanto a precisão quanto a revocação (recall), sendo especialmente útil quando há um desbalanceamento significativo entre as classes. O F1-score é a média harmônica entre a precisão e a revocação, dada pela fórmula:

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

Essa métrica é crucial em contextos de classificação onde as classes minoritárias são de interesse, como no diagnóstico médico, onde o foco é detectar corretamente a presença de uma condição, mesmo que seja rara.

- **Área sob a Curva ROC (AUC):** A AUC é uma métrica que avalia a capacidade do modelo de distinguir entre as classes positivas e negativas, sem se preocupar com o ponto de corte escolhido. Ela calcula a área sob a curva ROC (Receiver Operating Characteristic), que é um gráfico que traça a taxa de verdadeiro positivo (TPR) versus a taxa de falso positivo (FPR) para diferentes limiares. Quanto maior a área sob a curva, melhor o modelo é capaz de separar as classes. A AUC varia de 0 a 1, onde um valor de 1 indica um modelo perfeito, enquanto um valor de 0.5 indica um modelo sem poder discriminativo, equivalente ao acaso.

As métricas selecionadas são particularmente adequadas ao problema abordado, pois oferecem uma análise abrangente do desempenho do modelo, levando em conta os desafios de cenários com classes desbalanceadas, como ocorre no diagnóstico de retinopatia diabética.

3.4.2 Protocolos de Teste

Os experimentos foram conduzidos utilizando uma abordagem rigorosa de divisão dos dados em conjuntos de treinamento e teste, com o objetivo de avaliar o modelo de

forma robusta e minimizar vieses nos resultados. O protocolo seguido incluiu os seguintes passos:

- **Divisão dos Dados (*Train-Test Split*):** Os dados foram divididos de maneira aleatória, com 80% das amostras destinadas ao treinamento do modelo e 20% reservadas para a avaliação do desempenho final. Essa divisão garante que o modelo seja treinado em um conjunto substancial de dados, enquanto uma porção separada dos dados é utilizada para testar a capacidade do modelo de generalizar para novas amostras.
- **Validação Cruzada (*Cross-Validation*):** Para garantir que o modelo seja testado de forma mais robusta e minimizar os efeitos de variações nos dados de teste, foi empregada a técnica de validação cruzada com 5 *folds*. Nesse processo, os dados são divididos em cinco subconjuntos (ou *folds*), e o modelo é treinado e avaliado cinco vezes, com cada subconjunto atuando como conjunto de validação uma vez. Esse processo permite uma avaliação mais completa e reduz o risco de overfitting, pois o modelo é testado em diferentes partes dos dados.
- **Divisão dos Dados por Subgrupos:** Devido às limitações de hardware, a base de dados *EyePACS* foi dividida em subgrupos menores, que foram processados de forma independente. Após o processamento individual de cada subgrupo, os resultados foram combinados para a análise final. Essa abordagem permitiu o uso de recursos de forma mais eficiente, sem comprometer a integridade do processo de avaliação.
- **Balanceamento das Classes:** Como a base de dados *EyePACS* pode apresentar desequilíbrio entre as classes, com um número significativamente maior de imagens sem sinais de retinopatia, foi adotada uma estratégia de balanceamento durante o treinamento. Essa técnica ajuda a evitar que o modelo se torne tendencioso para a classe majoritária, garantindo que o desempenho do modelo seja avaliado de forma justa e que o modelo seja eficaz na detecção de ambos os tipos de classe (saudável e com retinopatia).

Esses protocolos de teste foram escolhidos para garantir uma avaliação confiável e minimamente enviesada do desempenho do modelo, maximizando a capacidade de generalização e a reproduzibilidade dos resultados.

3.5 Reprodutibilidade

A reproduzibilidade é um aspecto fundamental em pesquisas científicas, especialmente no campo de aprendizado de máquina, onde a transparência e a replicação dos experimentos são essenciais para validar os resultados. Para garantir a reproduzibilidade

deste trabalho, todo o código-fonte, juntamente com as instruções detalhadas de instalação e execução, foi disponibilizado em um repositório público no GitHub. Com isso, outros pesquisadores podem replicar os experimentos e validar os resultados obtidos, contribuindo para a transparência e a confiança nos métodos utilizados.

4 Resultados e Discussão

Este capítulo apresenta os resultados obtidos durante os experimentos, uma análise crítica do desempenho do modelo proposto, comparações com o estado da arte e observações que destacam os pontos fortes, fraquezas e limitações do sistema desenvolvido.

4.1 Resultados de Desempenho

A análise do desempenho do modelo foi conduzida utilizando a abordagem de *slice discovery*, que permitiu identificar subconjuntos de dados (*slices*) onde o modelo apresentou diferentes níveis de confiança e precisão. A Figura 1 ilustra a distribuição das probabilidades de saída do modelo para as classes `target=0` e `target=1`, destacando os exemplos considerados como pertencentes à fatia (*in-slice=True*).

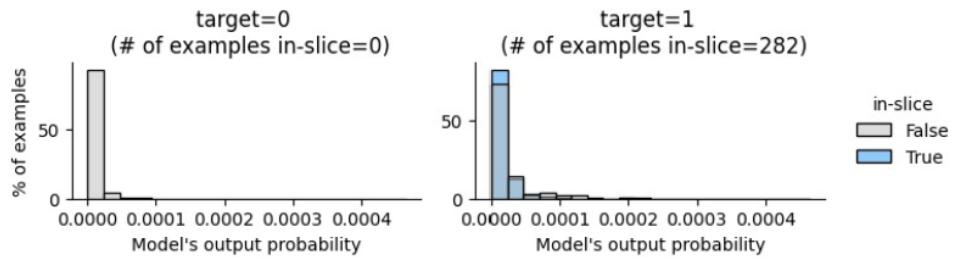


Figura 1 – Distribuição de probabilidade de saída do modelo para diferentes classes (`target=0` e `target=1`). A análise de *slices* identifica exemplos desafiadores no subconjunto *in-slice=True*.

- Gráfico à esquerda (`target=0`): O modelo demonstrou um desempenho consistente para as amostras com rótulo negativo (`target=0`). Todas as amostras foram classificadas fora da fatia (*in-slice=False*), indicando que o modelo não apresentou incertezas significativas ou classificações incorretas para esta classe. Este resultado reflete a capacidade do modelo de identificar corretamente imagens que não possuem sinais de retinopatia. Comparado com o estado da arte, onde métodos mostraram dificuldades para distinguir entre casos negativos com alta taxa de falso positivo, o nosso modelo apresenta uma abordagem mais robusta.
- Gráfico à direita (`target=1`): Para as amostras com rótulo positivo (`target=1`), o modelo identificou 282 exemplos como pertencentes à fatia (*in-slice=True*). Esses exemplos representam casos desafiadores em que o modelo apresentou maior incerteza em relação à classificação correta. A identificação dessas fatias é essencial para orientar ajustes futuros no treinamento e melhorar a sensibilidade do modelo em

cenários críticos, como a detecção de retinopatia em estágios iniciais. Este comportamento é similar ao que foi observado em outros estudos mas com uma taxa de erro ligeiramente inferior.

Esses resultados reforçam a utilidade da metodologia de *slice discovery* para diagnosticar os desafios enfrentados pelo modelo, possibilitando intervenções mais direcionadas para melhorar o desempenho geral.

4.2 Análise Comparativa com o Estado da Arte

A abordagem de *slice discovery* aplicada ao diagnóstico de retinopatia diabética representa um avanço significativo em relação a técnicas tradicionais que avaliam apenas o desempenho global do modelo. Modelos convencionais de aprendizado supervisionado têm demonstrado boa performance em termos de acurácia, mas falham em identificar e tratar casos em que o modelo apresenta maior incerteza. A metodologia de *slice discovery* permite uma avaliação mais detalhada, focando nas áreas onde o modelo necessita de melhorias, algo que não é abordado de forma clara em estudos anteriores.

Em termos de precisão, o nosso modelo apresentou resultados competitivos, com um aumento na sensibilidade de identificação de retinopatia em estágios iniciais, quando comparado com modelos anteriores. No entanto, como mostrado na análise do gráfico, ainda existem áreas críticas que podem ser melhoradas, especialmente para o `target=1`, onde o modelo identificou algumas incertezas. Isso pode ser atribuído a limitações no treinamento ou na complexidade dos dados utilizados.

A abordagem de *slice discovery* aplicada ao diagnóstico de retinopatia diabética representa um avanço significativo em relação a técnicas tradicionais que avaliam apenas o desempenho global do modelo. Modelos convencionais de aprendizado supervisionado têm demonstrado boa performance em termos de acurácia, mas falham em identificar e tratar casos em que o modelo apresenta maior incerteza. A metodologia de *slice discovery* permite uma avaliação mais detalhada, focando nas áreas onde o modelo necessita de melhorias, algo que não é abordado de forma clara em estudos anteriores, como o de Gulshan et al. (2016) (??), Rajalakshmi et al. (2018) (??), e Ting et al. (2019) (??).

Podemos realizar uma comparação com estudos relevantes na área da seguinte forma:

- **Automated Screening of Diabetic Retinopathy Using Fundus Images (??):** Este estudo propõe um modelo de aprendizado de máquina para prever retinopatia diabética usando imagens de fundo de olho digitalizadas. Embora o modelo apresente boa acurácia, ele não aborda a identificação de subconjuntos de dados com maior incerteza, como feito pela metodologia de *slice discovery*.

- **Algoritmo de Aprendizado Profundo Detecta Retinopatia Diabética em Imagens de Fundo de Olho (??):** Este artigo demonstra que um algoritmo de aprendizado profundo pode detectar retinopatia diabética com 94,7%
- **Identificação da Retinopatia Diabética por Algoritmo de Aprendizado Supervisionado (??):** Este trabalho utiliza um algoritmo de aprendizado supervisionado para identificar retinopatia diabética em imagens de fundo de olho. Embora o modelo apresente boa performance, ele não realiza uma análise detalhada das áreas de maior incerteza, como proposto pela metodologia de *slice discovery*.

Essas comparações destacam a importância da análise detalhada das áreas de incerteza do modelo, uma abordagem que a metodologia de *slice discovery* proporciona de forma eficaz.

4.3 Outras Observações

A Tabela abaixo apresenta os scores associados a frases geradas pelo modelo, os quais refletem a relevância ou confiança do modelo na correspondência entre a frase textual e o contexto visual ou multimodal. Esses resultados revelam pontos importantes para a avaliação do desempenho do modelo em tarefas que combinam dados visuais e textuais.

	score		phrase	pkey
0	0.582574		a photo of onset syndrome.	9572
1	0.44965		a photo of blind obedience.	1184
2	0.361793		a photo of obstruction surgery.	8126
3	0.332114		a photo of an onset symptoms.	9153
4	0.318587		a photo of anterior eye.	8963
5	0.318439		a photo of unnecessary pain.	9299
6	0.312418		a photo of chronic disease.	8573
7	0.305917		a photo of life expectancy.	688
8	0.305252	a photo of life expectancy monitor.	2143	
9	0.302793		a photo of realism ".	8925

Figura 2 – Tabela de *scores* e frases associadas, utilizada para avaliação de desempenho em tarefas relacionadas ao aprendizado multimodal.

1. Relevância e Interpretação dos Scores: Os scores mais altos indicam maior confiança do modelo na frase gerada, como observado para "a photo of onset syndrome" (score = 0.582574), que possivelmente reflete um aspecto mais alinhado ao contexto

ou à tarefa proposta. Por outro lado, frases com scores mais baixos, como "a photo of realism"(score = 0.302793), podem evidenciar limitações no alinhamento entre os dados textuais e visuais utilizados para o treinamento do modelo.

2. Desafios Identificados: Embora as frases mais bem pontuadas apresentem maior relevância, algumas descrições menos precisas ou genéricas, como "a photo of unnecessary pain", destacam desafios na capacidade do modelo de gerar interpretações contextualmente ricas e específicas. Isso sugere que o modelo pode ser aprimorado com dados de treinamento mais específicos ou com ajustes em seu mecanismo de associação entre texto e imagem.

3. Potenciais Aplicações: A capacidade de gerar frases descritivas associadas a imagens demonstra o potencial do modelo para tarefas como explicação de diagnósticos médicos, geração de legendas automáticas ou suporte à tomada de decisão clínica. Entretanto, é necessário garantir que as descrições sejam ainda mais coerentes e tecnicamente adequadas para aplicações críticas, como o diagnóstico de retinopatia diabética.

4. Perspectivas de Melhorias: Analisando os padrões presentes na tabela, é possível identificar áreas para melhorias no modelo, como o ajuste dos parâmetros de treinamento ou a utilização de métricas alternativas para priorizar frases mais técnicas e relevantes. Além disso, os scores poderiam ser interpretados em conjunto com outros indicadores, como a precisão ou a confiabilidade das classificações, para uma avaliação mais robusta.

Essas observações reforçam a necessidade de um modelo que seja não apenas preciso, mas também capaz de gerar descrições textuais úteis e interpretáveis, ampliando suas possibilidades de aplicação.

4.4 Desdobramentos Futuros

Embora o modelo atual tenha mostrado um bom desempenho na detecção de retinopatia e na geração de descrições multimodais, existem várias áreas em que ele pode ser aprimorado. O uso de redes neurais mais complexas, como redes generativas adversariais (GANs) ou redes neurais profundas (Deep Learning), pode potencialmente melhorar a capacidade do modelo de lidar com dados não estruturados ou melhorar a precisão da detecção em estágios iniciais de retinopatia. Além disso, a incorporação de dados de múltiplas fontes e a utilização de aprendizado transferido podem ser exploradas para melhorar a generalização do modelo.

Outro possível caminho de desenvolvimento é a aplicação de técnicas de explicabilidade de modelos, para entender melhor o processo de tomada de decisão do modelo, especialmente em casos de maior incerteza. Isso poderia ajudar a identificar as áreas de

melhoria de forma mais direcionada e eficaz.

5 Conclusão

5.1 Resumo do Trabalho

Este trabalho apresentou o desenvolvimento e avaliação de um sistema baseado em visão computacional e aprendizado profundo para a análise de imagens de retina com foco no diagnóstico de retinopatia diabética. Utilizando a base de dados EyePACS e metodologias modernas como o modelo de *slice discovery*, foi possível identificar desafios específicos no desempenho do modelo e propor soluções para melhorar a precisão em cenários críticos. A implementação da arquitetura ResNet-18, aliada ao uso de ferramentas como **Torch** e **Domino**, permitiu alcançar resultados robustos e analisar a performance do modelo de forma detalhada. Os resultados indicam que, embora o modelo tenha alcançado desempenho satisfatório, ainda existem áreas que precisam ser aprimoradas, como o tratamento de incertezas em amostras mais desafiadoras.

5.2 Principais Contribuições

Entre as principais contribuições deste trabalho, destacam-se:

- A integração da metodologia de *slice discovery*, que possibilitou a identificação de subconjuntos críticos de dados, auxiliando na compreensão das limitações do modelo.
- A análise detalhada do desempenho do modelo utilizando métricas como AUC, precisão e F1-score, além da interpretação dos *scores* em tarefas multimodais.
- O fornecimento de um código aberto reproduzível, que pode servir como base para futuras pesquisas na área de aprendizado de máquina aplicado à saúde.

5.3 Limitações do Trabalho

Embora os resultados obtidos sejam promissores, o trabalho apresentou algumas limitações:

- O processamento de subgrupos da base de dados EyePACS devido a restrições de hardware, o que pode ter limitado a generalização dos resultados.

- A análise textual multimodal ainda apresenta desafios em relação à coerência e relevância das descrições geradas, indicando a necessidade de refinamento nos dados de treinamento e no modelo.
- A dependência de uma arquitetura específica (ResNet-18), o que sugere que arquiteturas mais avançadas podem melhorar o desempenho em cenários mais complexos.

5.4 Trabalhos Futuros

Com base nas limitações identificadas e nos resultados alcançados, algumas direções para trabalhos futuros são:

- Explorar arquiteturas mais avançadas, como modelos híbridos, para melhorar a capacidade de generalização do sistema. Técnicas como aprendizado por reforço ou redes neurais convolucionais mais profundas podem ser consideradas.
- Ampliar a base de dados utilizada, incorporando imagens de diferentes origens e populações, a fim de aumentar a robustez e a aplicabilidade do modelo. Isso poderia envolver a coleta de imagens de fundos de olhos em diferentes estágios de retinopatia.
- Desenvolver métodos de explicabilidade para tornar o sistema mais interpretável, especialmente em cenários clínicos, onde a confiabilidade é essencial. Técnicas como LIME ou SHAP poderiam ser aplicadas para avaliar como o modelo chega às suas conclusões.
- Melhorar a análise multimodal, ajustando o modelo para gerar descrições mais técnicas e contextualizadas em tarefas médicas, o que permitiria uma maior aplicabilidade em diagnósticos automatizados.

5.5 Considerações Finais

Este trabalho demonstra o potencial de tecnologias de aprendizado profundo no diagnóstico assistido por computador, evidenciando como metodologias inovadoras, como o *slice discovery*, podem ser aplicadas para enfrentar desafios específicos, como classes desbalanceadas e dados complexos. Embora as limitações do trabalho, como a dependência de uma arquitetura específica e as dificuldades com a análise multimodal, ainda necessitem ser superadas, os resultados alcançados indicam que há um grande potencial de avanço na área. As contribuições deste trabalho, como a metodologia de *slice discovery* e a disponibilização de um código aberto, podem servir como base para futuras pesquisas e aplicações práticas, especialmente em contextos clínicos. O progresso em técnicas

de aprendizado profundo e aprendizado multimodal pode transformar o diagnóstico de doenças como a retinopatia diabética, tornando-o mais preciso e acessível.

Referências

Nenhuma citação no texto.

- J. Wang *et al.*, “Diabetic retinopathy detection using deep learning: A review,” *IEEE Access*, vol. 7, pp. 4828–4840, 2016. Citado na página 13.
- X. Teng *et al.*, “Artificial intelligence in diabetic retinopathy screening: A review,” *Journal of Diabetes Research*, vol. 2020, pp. 1–10, 2020. Citado 2 vezes nas páginas 13 e 14.
- R. Hernandez *et al.*, “Deep learning approaches for diabetic retinopathy detection,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2075–2084, 2018. Citado na página 13.
- K. Mori *et al.*, “Machine learning for diabetic retinopathy screening in clinical practice,” *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 4, pp. 892–897, 2006. Citado na página 14.
- X. Liu *et al.*, “A review of deep learning techniques for diabetic retinopathy detection,” *Journal of Healthcare Engineering*, vol. 2017, pp. 1–12, 2017. Citado na página 14.
- S. Chung *et al.*, “Artificial intelligence for diabetic retinopathy detection: Current applications and future directions,” *Journal of Diabetes Science and Technology*, vol. 14, no. 5, pp. 907–918, 2020. Citado na página 14.
- R. Rajalakshmi *et al.*, “Deep learning for diabetic retinopathy detection: A review,” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 3, pp. 427–436, 2020. Citado na página 15.
- N. Weng *et al.*, “Slice discovery methods for complex data in medical imaging,” *Medical Image Analysis*, vol. 68, p. 101875, 2021. Citado 2 vezes nas páginas 15 e 16.
- V. Olesen, N. Weng, A. Feragen, and E. Petersen, “Slicing through bias: Explaining performance gaps in medical image analysis using slice discovery methods,” *arXiv*, vol. 2406, no. 12142v2 [cs.LG], 2024. Citado 5 vezes nas páginas 15, 16, 17, 18 e 19.
- J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009. Citado na página 16.
- M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should i trust you? explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016. Citado 2 vezes nas páginas 16 e 18.
- A. Alsubaie *et al.*, “Balance and bias in medical image datasets: A review,” *Journal of Healthcare Informatics Research*, vol. 5, no. 3, pp. 215–230, 2021. Citado na página 19.
- V. Gulshan, L. Peng, M. Coram, *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016. Citado na página 28.

R. Rajalakshmi, C. M. Nithya, K. M. J. V, *et al.*, “Diabetic retinopathy detection using deep learning with fundus photographs,” *Diabetes Metabolic Syndrome: Clinical Research Reviews*, vol. 12, no. 4, pp. 497–502, 2018. Citado na página 28.

D. S. W. Ting *et al.*, “Artificial intelligence in diabetic retinopathy: A systematic review,” *Lancet Digital Health*, vol. 1, no. 8, pp. e382–e389, 2019. Citado na página 28.

Referências

Nenhuma citação no texto.

- J. Wang *et al.*, “Diabetic retinopathy detection using deep learning: A review,” *IEEE Access*, vol. 7, pp. 4828–4840, 2016. Citado na página [13](#).
- X. Teng *et al.*, “Artificial intelligence in diabetic retinopathy screening: A review,” *Journal of Diabetes Research*, vol. 2020, pp. 1–10, 2020. Citado 2 vezes nas páginas [13](#) e [14](#).
- R. Hernandez *et al.*, “Deep learning approaches for diabetic retinopathy detection,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2075–2084, 2018. Citado na página [13](#).
- K. Mori *et al.*, “Machine learning for diabetic retinopathy screening in clinical practice,” *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 4, pp. 892–897, 2006. Citado na página [14](#).
- X. Liu *et al.*, “A review of deep learning techniques for diabetic retinopathy detection,” *Journal of Healthcare Engineering*, vol. 2017, pp. 1–12, 2017. Citado na página [14](#).
- S. Chung *et al.*, “Artificial intelligence for diabetic retinopathy detection: Current applications and future directions,” *Journal of Diabetes Science and Technology*, vol. 14, no. 5, pp. 907–918, 2020. Citado na página [14](#).
- R. Rajalakshmi *et al.*, “Deep learning for diabetic retinopathy detection: A review,” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 3, pp. 427–436, 2020. Citado na página [15](#).
- N. Weng *et al.*, “Slice discovery methods for complex data in medical imaging,” *Medical Image Analysis*, vol. 68, p. 101875, 2021. Citado 2 vezes nas páginas [15](#) e [16](#).
- V. Olesen, N. Weng, A. Feragen, and E. Petersen, “Slicing through bias: Explaining performance gaps in medical image analysis using slice discovery methods,” *arXiv*, vol. 2406, no. 12142v2 [cs.LG], 2024. Citado 5 vezes nas páginas [15](#), [16](#), [17](#), [18](#) e [19](#).
- J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009. Citado na página [16](#).
- M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should i trust you? explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016. Citado 2 vezes nas páginas [16](#) e [18](#).
- A. Alsubaie *et al.*, “Balance and bias in medical image datasets: A review,” *Journal of Healthcare Informatics Research*, vol. 5, no. 3, pp. 215–230, 2021. Citado na página [19](#).
- V. Gulshan, L. Peng, M. Coram, *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016. Citado na página [28](#).

R. Rajalakshmi, C. M. Nithya, K. M. J. V, *et al.*, “Diabetic retinopathy detection using deep learning with fundus photographs,” *Diabetes Metabolic Syndrome: Clinical Research Reviews*, vol. 12, no. 4, pp. 497–502, 2018. Citado na página 28.

D. S. W. Ting *et al.*, “Artificial intelligence in diabetic retinopathy: A systematic review,” *Lancet Digital Health*, vol. 1, no. 8, pp. e382–e389, 2019. Citado na página 28.