



**UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO
CURSO DE ENGENHARIA DE COMPUTAÇÃO**

THYAGO MACHADO RODRIGUES

IDENTIFICAÇÃO BIOMÉTRICA OCULAR ATRAVÉS DE REDES NEURAS SIAMESAS

São Luís - MA

2023

THYAGO MACHADO RODRIGUES

IDENTIFICAÇÃO BIOMÉTRICA OCULAR ATRAVÉS DE REDES NEURAIAS SIAMESAS

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia da Computação da Universidade Estadual do Maranhão, como pré-requisito para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: Lúcio Flávio de Albuquerque Campos

São Luís - MA

2023

Rodrigues, Thyago Machado.

Identificação biométrica ocular através de redes neurais siamesas /
Thyago Machado Rodrigues. – São Luís, 2023.

50 f.

Monografia (Graduação em Engenharia de Computação) – Universidade
Estadual do Maranhão, 2023.

Orientador: Prof. Dr. Lúcio Flávio de Albuquerque Campos.

1.Redes siamesas. 2.Íris. 3.Classificação. I.Título.

CDU: 004.9:617.7

THYAGO MACHADO RODRIGUES

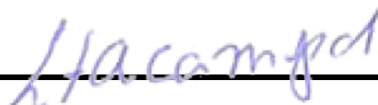
IDENTIFICAÇÃO BIOMÉTRICA OCULAR ATRAVÉS DE REDES NEURAIS SIAMESAS

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia da Computação da Universidade Estadual do Maranhão, como pré-requisito para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: Lúcio Flávio de Albuquerque Campos

Aprovada em: de janeiro de 2023.

BANCA EXAMINADORA



Prof. Dr. Lúcio Flávio de Albuquerque Campos (Orientador)

Universidade Estadual do Maranhão



Prof. Dr. Antônio Fernando Lavareda Jacob Junior (Examinador)

Universidade Estadual do Maranhão



Prof. Pedro Brandão Neto (Examinador)

Universidade Estadual do Maranhão

AGRADECIMENTOS

Agradeço a Deus que permitiu que tudo isso acontecesse, ao longo de minha vida, e não somente nestes anos como universitário, mas que em todos os momentos é o maior mestre que alguém pode conhecer.

LISTA DE FIGURAS

Figura 1. Exemplos de características biométricas.	18
Figura 2. Comparação entre as características biométricas.	18
Figura 3. Anatomia Ocular.	19
Figura 4. Aplicação do Filtro de Canny.	21
Figura 5. Aplicação da Operação de Erosão.	23
Figura 6. Aplicação da Operação de Dilatação.	23
Figura 7. Aplicação da Operação de Dilatação.	24
Figura 8. Aplicação da Transformada de Hough.	25
Figura 9. Parâmetros da Reta.	26
Figura 10. Exemplo de Arquitetura de uma Rede Perceptron de uma Saída.	28
Figura 11. Funções de Ativação Comuns.	29
Figura 12. Exemplo de Arquitetura de uma Rede Convolutacional.	29
Figura 13. Exemplo de Arquitetura de uma Rede Siamesa.	32
Figura 14. Perda por Energia para valores semelhantes (vermelho) e divergentes (azul).	33
Figura 15. Exemplos de Imagens Ruidosas.	35
Figura 16. Localização da Íris.	38
Figura 17. Estrutura do Extrator de Características.	39
Figura 18. Estrutura da Rede Simples.	40
Figura 19. Estrutura da Rede Densa.	41
Figura 20. Um dos Treinamentos Consistentes Encontrados.	42
Figura 21. Treinamento (Rede Simples).	42
Figura 22. Treinamento (Rede Densa).	43
Figura 23. Matriz de Confusão (Rede Densa).	44

LISTA DE TABELAS

Tabela 1. Detalhes do Dataset.	35
Tabela 2. Nomenclatura das Imagens.	36
Tabela 3. Comparação da Acurácia dos Modelos.	44

LISTA DE EQUAÇÕES

Equação 1. Definição da Transformação Erosão	22
Equação 2. Definição da Transformação Dilatação	23
Equação 3. Definição da Transformação Abertura	24
Equação 4. Função de Ativação	28
Equação 5. Definição da Convolução	30
Equação 6. Definição da Convolução para Imagens	30
Equação 7. Equação Geral de Distância	32
Equação 8. Contrastive Loss	32
Equação 9. Binarycrossentropy	33

LISTA DE ACRÔNIMOS

BGR	Blue, Green, Red
CNN	Convolutional Neural Network
RGB	Red, Green, Blue
RNC	Rede Neural Convolutcional

RESUMO

O uso da tecnologia para tornar os processos mais eficientes tem levado a uma necessidade crescente de identificação automática dos indivíduos. Métodos tradicionais de identificação, como assinaturas e cartões, estão se tornando menos eficazes e estão sendo substituídos por métodos de identificação biométrica, que utilizam características físicas ou comportamentais dos indivíduos. Este trabalho propõe o desenvolvimento de modelos de reconhecimento biométrico baseado em íris, que utiliza técnicas de aprendizado de máquina, especificamente redes siamesas e convolucionais. Dois modelos de redes siamesas foram construídos, sendo um modelo mais simples e outro mais robusto. Experimentos realizados em uma base de dados que conta com imagens de íris sob diversas condições de captura, como ruído e oclusões, propiciaram um modelo de acurácia de 84.51% em dados de teste.

Palavras-chave: Redes Siamesas. Íris. Classificação.

ABSTRACT

The use of technology to make processes more efficient has led to a growing need for automatic identification of individuals. Traditional identification methods, such as signatures and cards, are becoming less effective and are being replaced by biometric identification methods, which use physical or behavioral characteristics of individuals. This work proposes the development of iris-based biometric recognition models, which uses machine learning techniques, specifically siamese and convolutional networks. Two models of Siamese networks were built, a simpler model and a more robust one. Experiments carried out in a database that has iris images under different capture conditions, such as noise and occlusions, provided an accuracy model of 84.51% in test data.

Keywords: *Siamese Networks. Iris. Classification.*

SUMÁRIO

1. INTRODUÇÃO	13
1.1. OBJETIVOS	14
1.1.1. Objetivo Geral	14
1.1.2. Objetivos Específicos	14
1.2. METODOLOGIA	14
1.3. ESTRUTURA DO TRABALHO	15
2. FUNDAMENTAÇÃO TEÓRICA	16
2.1. SISTEMAS BIOMÉTRICOS	16
2.1.1. Região Ocular	18
2.1.2. Íris	19
2.2. PROCESSAMENTO DE IMAGENS	21
2.2.1. Identificação de Bordas	21
2.2.2. Transformações Morfológicas	22
2.2.2.1. Erosão	22
2.2.2.2. Dilatação	23
2.2.2.3. Abertura	24
2.2.3. Detecção de Círculos	24
2.2.3.1. Transformada de Hough	25
2.2.4. Limpeza de Ruídos	26
2.2.5. Normalização.	27
2.3. APRENDIZADO DE MÁQUINA	27
2.3.1. Métodos de Classificação	27
2.3.1.1. Perceptron Multicamadas	27
2.3.1.2. Redes Neurais Convolucionais	29
2.3.1.3. Redes Neurais Siamesas	31
2.3.2. Métricas	32
2.3.2.1. Contrastive Loss	32
2.3.2.2. Binary Crossentropy	33
3. PROJETO	34
3.1. FRAMEWORK EXPERIMENTAL	34
3.1.1. Base de Dados	34
3.1.2. Pré-processamento	36
3.1.3. Localização da Íris	37
3.1.4. Arquitetura da Proposta	39
3.1.4.1. Extrator de Características	39
3.1.4.2. Rede Simples	40
3.1.4.3. Rede Densa	40
3.2. RESULTADOS	41

4. CONSIDERAÇÕES FINAIS	45
SUGESTÕES PARA TRABALHOS FUTUROS	46
REFERÊNCIAS	47
GLOSSÁRIO	49
APÊNDICE A - Repositório Github	50

1. INTRODUÇÃO

Os humanos têm usado características físicas, incluindo o rosto, a voz e a marcha, i.e., o ato de caminhar, para identificarem-se por milhares de anos. A ideia de usar várias medidas e indicadores corporais para identificar pessoas, especialmente criminosos, foi desenvolvida por Alphonse Bertillon, chefe da divisão de identificação criminal do departamento de polícia de Paris em meados do século XIX, e, em seguida, colocada em prática. Sua abordagem ganhou popularidade rapidamente, porém foi ofuscada por uma descoberta que era muito mais significativa e prática no final do século XIX: a distinção entre impressões humanas em mídia digital. Após essa descoberta, muitos dos principais departamentos responsáveis pela aplicação da lei propuseram a ideia do primeiro "registro" impressões digitais dos criminosos e armazená-las em um banco de dados, com o objetivo de, posteriormente, em locais de onde ocorreram delitos, levantar as impressões digitais, geralmente fragmentadas, chamadas de latentes, e compará-las com as impressões digitais previamente cadastradas em banco para a identificação de criminosos (JAIN; ROSS; PRABHAKAR, 2004).

Embora a biometria tenha se tornado popular através de seu uso extensivo na aplicação da lei para identificar criminosos (por exemplo, estrangeiros trabalhando sem autorização, autenticação de segurança para funcionários realizando tarefas sensíveis, determinação de paternidade, teste de polígrafo, e identificação de presos e condenados), agora é cada vez mais usado para estabelecer o reconhecimento de pessoas em uma ampla gama de aplicações civis (JAIN; ROSS; PRABHAKAR, 2004).

Nesse contexto, aplicações de reconhecimento de indivíduos por meio da biometria ocular, focado em regiões do olho, especificamente a íris, têm sido tradicionalmente consideradas como a biometria ocular mais popular e amplamente aceita devido à sua alta precisão e estabilidade ao longo da vida do indivíduo (RATTANI; DERAKHSHANI, 2017).

Ademais, métodos de identificação de indivíduos baseados em aprendizado de máquina, em especial as redes neurais, geralmente baseados em classificação de amostras baseadas em seus classes, têm desempenhando um papel importante para a resolução de problemas do cotidiano. A classificação em aprendizado de máquina compreende o processo em que uma amostra, com n características, é rotulada com um classe. O algoritmo de aprendizado busca compreender como essas características se relacionam com a classe da amostra, no entanto, para o contexto de

identificação de indivíduos, cada pessoa se transformaria em uma classe, implicando em um número de classes gigantescas e prejudicando o processo de aprendizado. Este problema, no entanto, pode ser abordado utilizando a arquitetura de redes siamesas, a qual trabalha com pares de amostra e busca a relação entre esses pares para com a classe designada (VARGAS; ZHANG; IZQUIERDO, 2020).

1.1. OBJETIVOS

1.1.1. Objetivo Geral

O objetivo deste trabalho se concentra no desenvolvimento de um modelo de rede neural profunda capaz de, dada a entrada de uma imagem contendo a região ocular, comparar com uma amostra existente e avaliar o grau de compatibilidade entre as imagens, apontando para uma relação genuína ou de impostor.

1.1.2. Objetivos Específicos

Compreende-se que para prosseguir na direção do objetivo deste trabalho, os seguintes objetivos/etapas são necessárias:

- Definir rótulos para a aprendizagem de máquina;
- Aplicar técnicas de pré-processamentos à luz da literatura;
- Aplicar diferentes arquitetura de redes profundas;
- Avaliar modelos;

1.2. METODOLOGIA

A pesquisa que será abordada neste projeto seguirá uma abordagem quantitativa, objetivando a busca por modelos mais bem ajustados, de natureza aplicada e de caráter experimental, objetivando explorar diferentes abordagens que compreendem o processo de identificação de indivíduos através da íris.

Para a execução deste projeto, planeja-se iniciar com uma busca de técnicas de pré-processamento de imagens voltadas à problemática, bem como buscar por técnicas comuns de

machine learning também direcionadas para o problema proposto para esta pesquisa, confrontando-as através das métricas comuns de treinamento, a fim de obter uma análise quantitativa do desempenho dos modelos e, por conseguinte, apresentar o modelo mais bem ajustado. Por fim, expor o modelo ao ambiente não controlado, testando-o com entradas sujeitas à alterações do ambiente.

1.3. ESTRUTURA DO TRABALHO

Este relatório está estruturado da forma como segue. No Capítulo 2 é apresentada a fundamentação teórica. O Capítulo 3 apresenta o desenvolvimento do trabalho. Finalmente no Capítulo 4 são apresentadas as conclusões e considerações finais.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. SISTEMAS BIOMÉTRICOS

O componente essencial de um sistema biométrico é uma rotina de reconhecimento de padrões que coleta dados biométricos de um indivíduo, extrai um conjunto de recursos dos dados e compara o conjunto de recursos com o modelo definido no banco de dados (JAIN; ROSS; PRABHAKAR, 2004).

Uma das principais vantagens da utilização de sistemas biométricos é a maior segurança contra fraudes, devido ao fato de que a identificação por meio de cartões de identificação ou senhas pode ser mais facilmente burlada, seja por meio de processos de clonagens, ou utilização de técnicas para descobrir de senhas, sem contar que a posse pode ser perdida e pode ser esquecida pelo próprio indivíduo.

O objetivo básico de um sistema biométrico pode ser a identificação ou validação, dependendo do seu uso.

Um sistema de identificação individual funciona realizando uma pesquisa em todo o banco de dados para identificar o usuário baseado exclusivamente em sua entrada biométrica no sistema (VIZONI, 2019), aplicando-os em comparação com registros em banco de dados em momentos distintos, para definir a identidade de alguém (DUARTE et al., 2021). Neste sistema, uma pessoa que quer ser reconhecida afirma sua identidade, normalmente usando um PIN, nome do usuário, ou cartão inteligente. O sistema então faz uma comparação um-para-um para determinar se a afirmação é verdadeira ou falsa (JAIN; ROSS; PRABHAKAR, 2004).

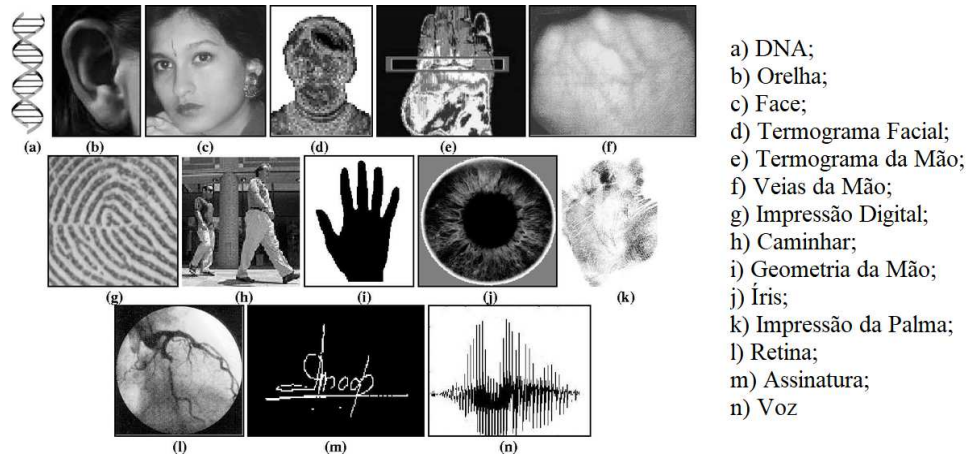
Em contrapartida, uma abordagem de validação tem como objetivo a questão da comprovação da identidade de um usuário com base em uma entrada anterior (VIZONI, 2019). O sistema reconhece uma pessoa procurando todos os modelos de usuário no banco de dados para uma correspondência. Este tipo de verificação de identidade é, portanto, um método empírico que consiste em definir o indivíduo como sendo o mesmo ou, também, reafirmar o ato de conhecer, através de certas características, considerando que uma pessoa é a mesma que já conhecemos em outro momento (DUARTE et al., 2021).

Sendo um sistema biométrico capaz de realizar a análise de características biológicas ou comportamentais de indivíduos, para garantir seu bom desempenho, isto é, conseguir diferenciar indivíduos com uma margem de erro aceitável, e estabelecer o reconhecimento correto de dados de entrada genuínos, é ideal que as características satisfaçam os seguintes requisitos (JAIN; ROSS; PRABHAKAR, 2004):

- **Universalidade:** Garantir que cada pessoa possua a característica;
- **Unicidade:** Garantir que cada indivíduo seja distinto dos demais;
- **Permanência:** Apresentar baixa mutabilidade ao longo do tempo;
- **Coletabilidade:** Ser mensurável quantitativamente;
- **Performance:** Promover de modo eficaz e eficiente o reconhecimento possibilitado pela característica, incluindo recursos necessários, operações, e fatores ambientais que venham a influenciar o desempenho;
- **Aceitabilidade:** Ter aceitação das pessoas que utilizarem aquela característica como forma de identificação;
- **Segurança:** Ser robusta contra fraudes, isto é, a característica não deve ser facilmente violada.

A Figura 1 ilustra algumas características biométricas que, ao levar em conta os pré-requisitos descritos anteriormente, produzem bons resultados e são, portanto, utilizados para a construção de vários tipos de sistemas biométricos.

Figura 1: Exemplos de características biométricas.



Fonte: Adaptado de (JAIN; ROSS; PRABHAKAR, 2004)

Cada característica biométrica apresenta suas peculiaridades e diferem quanto sua aderência aos requisitos mencionados anteriormente, captura de biométrica, representações dos dados, etc. Mediante às características biométricas apresentadas, a Figura 2 ilustra uma comparação entre algumas dessas características quanto aos requisitos propostos, ponderando-os em relação a aderência às respectivas exigências previamente discutidas, classificando-as em baixo (B), médio (M) e alto (A).

Figura 2: Comparação entre as características biométricas.

Característica Biométrica	Universalidade	Unicidade	Continuidade	Coletabilidade	Performance	Aceitabilidade	Segurança
DNA	A	A	A	B	A	B	A
Face	A	B	M	A	B	A	B
Termograma facial	A	A	B	A	M	A	A
Impressão digital	M	A	A	M	A	M	M
Geometria da mão	M	M	M	A	M	M	M
Veias da mão	M	M	M	M	M	M	A
Íris	A	A	A	M	A	B	A

Fonte: Adaptado de (VIZONI, 2019)

2.1.1. Região Ocular

O termo "região ocular", também conhecido como região periocular, refere-se à área do rosto que inclui o olho e sua íris como componentes principais, porém aliado de características

auxiliares, como as bochechas, pálidos, sobrancelhas, e pele ao seu redor. Como as informações oculares podem ser capturadas pelo mesmo sensor usado para capturar imagens de rostos e íris, características desta região têm sido utilizadas para construir sistemas biométricos que busquem características únicas em indivíduos. O traço biométrico periocular também pode ser vantajoso como modalidade complementar, quando outros traços oculares apresentam grandes variações entre as classes, portanto, sistemas podem ser usados sozinhos, em conjunto com rostos ou íris, ou ambos (RATTANI; DERAKHSHANI, 2017).

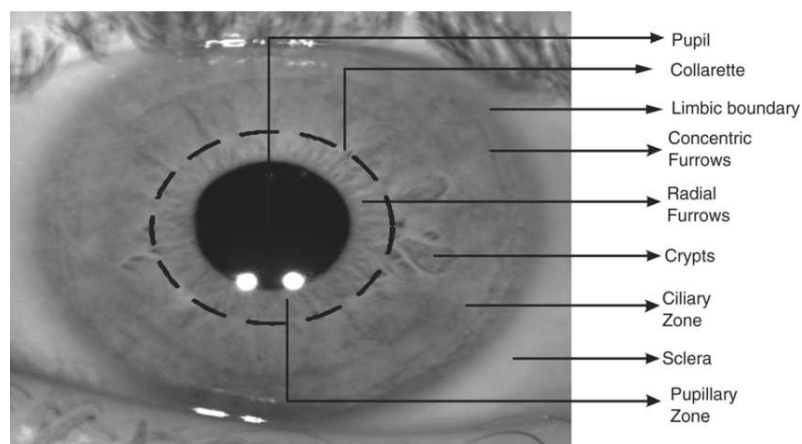
Dessa forma, a biometria da região do olho pode ser apresentada como uma alternativa em circunstâncias em que possa ser difícil de reconhecer um rosto ou uma íris, seja devido à obstrução da imagem ou inapropriado posicionamento do sensor no momento da captura (VIZONI, 2019).

2.1.2. Íris

A íris é a área do olho que corresponde a um diafragma circular fino e tem uma perfuração central conhecida como pupila. O objetivo da íris é regular a quantidade de luz entrando através da pupila por meio de os músculos esfínter e dilatação, que mudam o diâmetro da pupila dependendo da quantidade de luz presente no meio ambiente (VIZONI, 2019). A Figura 3 aponta alguns dos principais componentes de interesse na anatomia ocular para a classificação biométrica.

As texturas da íris são únicas, divididas em zona pupilar e zona ciliar. É um órgão interno protegido e permanece estável ao longo da vida do indivíduo. É difícil modificar a textura da íris (PATIL, 2013).

Figura 3: Anatomia Ocular.



Fonte: (GHALI et al., 2017)

A precisão do sistema de reconhecimento de íris é maior do que outros sistemas biométricos, como assinatura, impressão digital e reconhecimento facial. Essa precisão depende da precisão dos dados da imagem capturada da íris. O ambiente em que os dados da imagem da íris capturada no momento da captura pode ser dividido em ambientes restritos e menos restritos. Em um ambiente limitado, a distância entre a câmera e o reflexo do objeto é mantida ao mínimo (menos que um metro), e tanto a câmera quanto o reflexo estão no mesmo olho, logo compreendem imagens que apresentam menos ruídos (PATIL, 2013).

Em ambientes menos restritos, os dados das imagens da íris são afetados por vários tipos de ruídos, como a oclusão da íris devido a pálpebra e cílios, o ângulo fora do foco, o reflexão equilátero, variações nos raios, a desfocalização do movimento, etc. O desempenho do sistema de reconhecimento de ruído depende da qualidade dos dados de entrada e se degrada na presença de ruído. As condições de raios podem diferir em aplicações onde a cooperação do usuário não é esperada. A imagem da íris contém ruído nesses ambientes menos restritos, e como resultado de sua presença, o desempenho do sistema de reconhecimento de íris se deteriora (PATIL, 2013).

Na construção de sistemas de reconhecimento de íris, as imagens degradadas representam um grande problema. As circunstâncias circundam quando uma imagem de íris capturada em um ambiente menos restrito e controlado produz imagens não ideais com qualidade de imagem ruim, como reflexão especulativa, variações de iluminação e baixo contraste etc (PATIL, 2013).

Além de boa iluminação e equipamentos apropriados para captura de imagens oculares, bons resultados requerem cooperação individual porque a posição da cabeça e a distância entre o sensor e o sujeito têm um direto impacto na capacidade de obter imagens de alta qualidade para extração de informações de íris. Outro fator crucial são as possíveis obstruções da região da íris causadas por pálpebras parcialmente fechadas ou pelo posicionamento do cílio acima dos olhos, o que acaba por impedir a captura adequada de informações de textura (VIZONI, 2019).

A identificação de pessoas a partir do reconhecimento da íris é um dos métodos mais populares, pois atende à maioria dos requisitos de desempenho e reconhecimento correto de entradas genuínas descritas anteriormente (JAIN; ROSS; PRABHAKAR, 2004; VIZONI, 2019).

2.2. PROCESSAMENTO DE IMAGENS

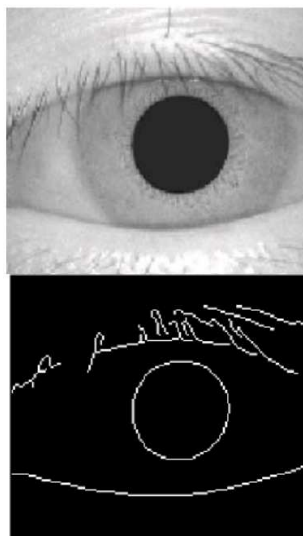
2.2.1. Identificação de Bordas

A identificação de bordas é uma dos primeiros tratamentos a se aplicar para a problemática da autenticação biométrica baseada em íris devido à necessidade de restringir a quantidade de informação contida na imagem, de modo a permanecer apenas as informações de interesse.

Uma vez que a imagem foi adquirida, é necessário detectar o conjunto de fronteiras que a compõem. As fronteiras mais importantes a considerar são as duas circunferências que limitam a íris e pupila, a pálpebra superior, a pálpebra inferior e finalmente as pestanas que possam estar eventualmente sobrepostas com a íris (MALTEZ; DO VALE, [s.d.]). Esta etapa é de grande importância, uma vez que os algoritmos de localização de objetos e/ou entidades podem depender da nitidez das bordas encontradas, que servirão como uma etapa antecedente à identificação de círculos que será descrito posteriormente.

Embora existam métodos alternativos para obter um conjunto de linhas que destaquem as arestas e/ou linhas pertencentes à imagem, um dos métodos mais comuns e de melhor resultado para detectar fronteiras em imagens é o método de Canny. Seu algoritmo é baseado em um gradiente que detecta arestas com uma alta precisão e com uma complexidade computacional aceitável (MALTEZ; DO VALE, [s.d.]).

Figura 4: Aplicação do Filtro de Canny.



Fonte: (MALTEZ; DO VALE, [s.d.])

Este algoritmo permite a detecção e localização de círculos juntamente aos seus centros dentro de uma determinada área, desde que estes respeitem a equação geral para o círculo. Este algoritmo tem um alto nível de complexidade computacional e o tempo de processamento tende a aumentar quando os dados são obviamente incompletos. Esta é uma das razões pelas quais é crucial que, no momento em que a imagem é adquirida, a câmera é posicionada ortogonalmente à íris, resultando na menor quantidade de distorção das circunferências. Entretanto, o algoritmo pode revelar-se incapaz de apresentar resultados corretos também nos casos em que o contraste entre pupila/íris/esclerótica não é suficientemente nítido. Para minimizar estes problemas, a transformada de Hough é normalmente aplicada na imagem obtida através da aplicação do filtro de Canny, pois as arestas na imagem já foram aí evidenciadas (MALTEZ; DO VALE, [s.d.]).

2.2.2. Transformações Morfológicas

De acordo com a morfologia, uma imagem é composta de uma coleção de pixels que são agrupados e têm uma estrutura bidimensional (forma). Certas operações matemáticas em grupos de pixels podem ser usadas para destacar características específicas dos formulários, permitindo a serem contados ou reconhecidos. Os filtros digitais são um dos componentes mais importantes para processar a imagem e obter a melhor resolução a partir dela (DE SOUSA et al., 2020).

2.2.2.1. Erosão

Com A e B como conjuntos de Z^2 , a erosão de A por B , indicada por $A \ominus B$, é definida como:

$$A \ominus B = \{ Z | (B_z) \subseteq A \} \quad \text{Eq. 1}$$

Em outras palavras, a equação informa que a operação de a erosão de A por B é o conjunto de todos os pontos z de forma que B , transladado por z , está contido em A , de tal maneira de que A representa a imagem a alvo e B representa o elemento estruturante (GONZALEZ; WOODS, 2009). A imagem a seguir demonstra a aplicação da erosão em uma imagem binária.

Figura 5: Aplicação da Operação de Erosão.



Fonte: (“OpenCV: Morphological Transformations”, [s.d.]

Visualmente, o efeito da erosão em uma imagem é o estreitamento de seus contornos e exclusão de pequenas linhas próximas às bordas.

2.2.2.2. Dilatação

Com A e B como conjuntos de Z^2 , a dilatação de A por B , indicada por $A \oplus B$, é definida como:

$$A \oplus B = \{Z | ((-B)_z) \cap A \neq \emptyset\} \quad \text{Eq. 2}$$

Em outras palavras, a equação informa que a operação de dilatação de A por B é o conjunto de todos os deslocamentos, z , de forma que $-B$ e A se sobreponham por pelo menos um elemento (GONZALEZ; WOODS, 2009). A imagem a seguir demonstra a aplicação da dilatação em uma imagem binária.

Figura 6: Aplicação da Operação de Dilatação.



Fonte: (“OpenCV: Morphological Transformations”, [s.d.]

Visualmente, o efeito da dilatação em uma imagem é o alargamento de seus contornos e incorporação de pequenas linhas próximas às suas bordas.

2.2.2.3. Abertura

Com A e B como conjuntos de Z^2 , a abertura de A pelo elemento estruturante B , indicada por $A \circ B$, é definida como:

$$A \circ B = (A \ominus B) \oplus B \quad \text{Eq. 3}$$

Sendo assim, a abertura de A por B é a erosão de A por B , seguida de uma dilatação do resultado por B . A aplicação da abertura em uma imagem produz a oclusão de pequenos elementos soltos na imagem. A imagem a seguir demonstra a aplicação da abertura em uma imagem binária.

Figura 7: Aplicação da Operação de Dilatação.



Fonte: (“OpenCV: Morphological Transformations”, [s.d.]

2.2.3. Detecção de Círculos

Restringir o espaço amostral para que haja apenas informações relevantes para a solução da problemática é um dos fatores primordiais para o aprendizado de máquina e aplicações envolvendo técnicas de classificação. Ao se tratar da identificação ocular, encontrar os limites internos e externos da íris é o objetivo da fase de localização. A região entre a pupila e a íris serve como limite interno e a região da fronteira entre a pupila e a íris é o limite externo (NETO; NETO, 2019). Existem técnicas de segmentação de imagens que executam a extração de bordas, sendo uma das principais a Transformada de Hough, que será abordada posteriormente.

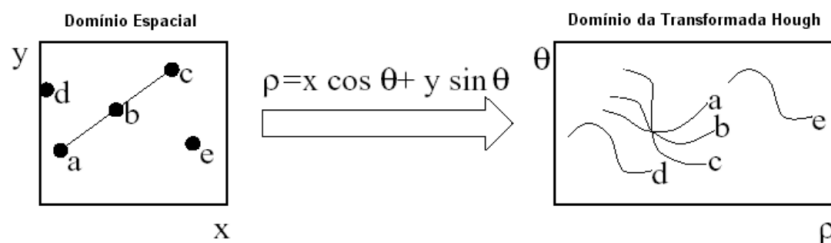
A aquisição e localização da íris dentro da imagem adquirida é um dos passos fundamentais no desenvolvimento de sistemas de reconhecimento biométrico baseado em íris. Para o processo de

identificação, os padrões de textura da íris correspondem à principal característica para a distinção de indivíduos, desta forma, todas as regiões de imagem remanescentes que não pertencem à íris devem ser removidas. O desafio, portanto, reside em localizar a íris e determinar seus limites internos e externos, uma vez que não há um padrão explícito para sua segregação e esses limites são afetados por diferenças individuais na anatomia, iluminação, dilatação da pupila, reflexo e contraste da imagem (NETO; NETO, 2019).

2.2.3.1. Transformada de Hough

A transformada de Hough é um método matemático utilizado para detectar formas em imagens que podem ser parametrizadas, como retas, círculos e elipses, por exemplo. A ideia fundamental por trás das transformações de Hough é aplicar uma transformação a uma imagem de tal forma que cada ponto pertencente à mesma curva seja representado por um único ponto em um novo espaço para a curva desejada (NETO; NETO, 2019). A figura 8 demonstra visualmente a aplicação da transformada em uma amostra de pontos no plano cartesiano.

Figura 8: Aplicação da Transformada de Hough.



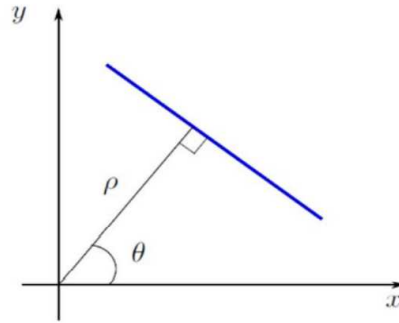
Fonte: (DUARTE, 2003)

O algoritmo de Hough requer um acumulador de dimensão igual ao número de parâmetros desconhecidos na equação da família de curvas que são buscadas. Por exemplo, usar a equação $y = ax + b$ para calcular segmentos de linha requer conhecer dois parâmetros para cada segmento: a e b . As duas dimensões da matriz de acumulação para esta família correspondem aos valores somados para a e b .

Diferentes formas de parametrizar retas podem ser adotadas. Diferentemente da descrita anteriormente, a reta também pode ser descrita pela distância ρ entre a origem e a reta e θ , ângulo que representa a inclinação entre o eixo x e a reta que intercepta a origem e o ponto mais próximo

da reta alvo, resultando em uma equação característica da seguinte forma $\rho = x\cos(\theta) + y\sin(\theta)$ (BERGUES et al., 2015; NETO; NETO, 2019).

Figura 9: Parâmetros da Reta.



Fonte: (BERGUES et al., 2015)

Cada borda de uma imagem é transformada pelo mapeamento para determinar células no espaço de parâmetros, indicadas pelas primitivas definidas através do ponto analisado. Essas células são incrementadas, e indicarão, no final do processo, através da máxima local do acumulador, quais os parâmetros correspondentes à forma especificada.

A transformação apresenta a desvantagem de ser computacionalmente limitada e exigir uma grande quantidade de memória. O processamento computacional aumenta com o acréscimo no tamanho do vetor de acumulação e com a exatidão com que os parâmetros são determinados. A precisão e confiabilidade pelo qual os parâmetros podem ser determinados está diretamente relacionada com a resolução do vetor de acumulação (DUARTE, 2003).

Uma característica da Transformação de Hough é o fato de não permitir a determinação do comprimento e dos pontos de extremidade das linhas existentes em uma imagem. Nesta situação, torna-se necessário utilizar métodos adicionais para atingir esses objetivos (DUARTE, 2003).

2.2.4. Limpeza de Ruídos

Existem vários procedimentos que podem ser usados para remover as obstruções que possam se fazer presente na íris, tais como os procedimentos que objetivam a remoção de reflexos e cílios.

A remoção de eventuais reflexos luminosos condensados em um local podem ser removidos através de um filtro passa baixa na imagem em tons de cinza, pois os altos valores dos pixels RGB serão convertidos em altos valores na escala cinza, resultando em valores mais próximos de 255 (MALTEZ; DO VALE, [s.d.]).

A retirada dos cílios na imagem que possam estar sobrepostas com a íris, no entanto, pode ser realizada com um filtro degrau simples, cortando os pixels com valores próximos ou iguais a zero (pois os cílios são geralmente possuem tonalidades mais escuras) ou a um filtro de Canny nos casos em que a estrutura ou a cor não permitam utilizar o anterior (MALTEZ; DO VALE, [s.d.]).

2.2.5. Normalização.

O processo de normalização tem como objetivo transformar uma imagem, a qual possui dimensões variáveis, e convertê-la a um formato que, além de preservar suas características, permita a comparação entre os dois elementos. Este processo também pode atender pela designação de assinatura ou template.

As técnicas de normalização empregadas compreendem o processamento realizado pelas redes convolucionais em seus mapas de características, as quais serão abordadas nos tópicos posteriores.

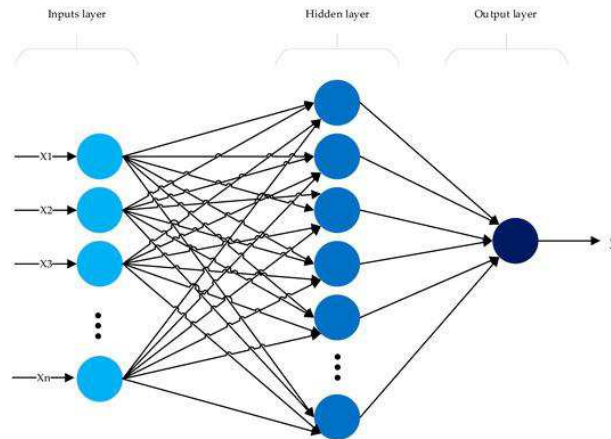
2.3. APRENDIZADO DE MÁQUINA

2.3.1. Métodos de Classificação

2.3.1.1. Perceptron Multicamadas

O perceptron multicamadas (MLP) é um tipo de rede neural construída sobre os princípios do aprendizado supervisionado, utilizando como abordagem de aprendizado o método de retropropagação. A Figura 10 demonstra uma MLP que se constitui de uma estrutura de três camadas, incluindo a camada de entrada, a camada oculta e a camada de saída, na qual cada neurônio é conectado a todos os neurônios na próxima camada. É frequentemente relatado que a MLP tem uma grande função em problemas não lineares.

Figura 10: Exemplo de Arquitetura de uma Rede Perceptron de uma Saída.



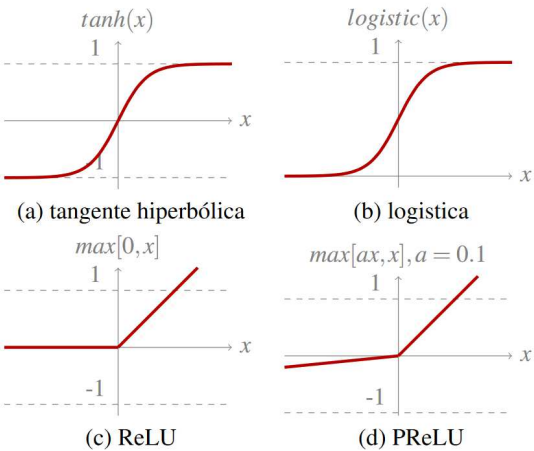
Fonte: (NOSRATABADI et al., 2021)

A estrutura da camada de entrada corresponde aos dados já tratados, cuja transição para a camada posterior é dada por meio de um vetor de pesos comumente denotado de W_i . Cada camada h_i , excluindo-se a camada de entrada, é atualizada, por meio do uso de b_i (o fator de convergência atual) e ativada da seguinte forma (PONTI; DA COSTA, 2018):

$$h_{i+1} = f(W_i h_i + b_i) \quad \text{Eq. 4}$$

A soma é submetida a uma função $f(\cdot)$, a qual trata de realizar a ativação. A função mais comum é a Perceptron tanto para redes rasas (uma única camada oculta) quanto para redes de multicamadas. Outras funções de ativação podem ser utilizadas, porém dependem do tipo de problema a ser tratado, por exemplo a ReLu, a qual é uma função linear que anula valores negativos e comprime os demais valores positivos em uma reta (PONTI; DA COSTA, 2018). A figura 11 aborda as funções de ativação mais comuns.

Figura 11: Funções de Ativação Comuns.

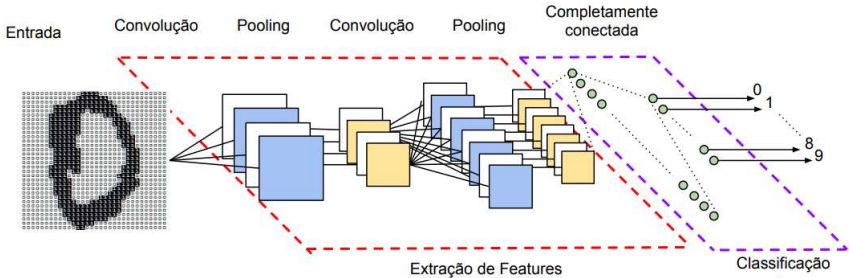


Fonte: (PONTI; DA COSTA, 2018)

2.3.1.2. Redes Neurais Convolucionais

Uma rede neural convolucional, CNN, é uma arquitetura de rede neural artificial, variante do perceptron de múltiplas camadas (multicamadas) que, assim como outras modalidades de redes, são baseadas em um processo biológico: o processamento de dados visuais. As redes neurais convolucionais são comumente usadas em aplicações de detecção, classificação e reconhecimento em imagens e vídeos (CHAVES, 2019). A RNC é dividida em alguns estágios principais, como pode ser visto na Figura 12.

Figura 12: Exemplo de Arquitetura de uma Rede Convolucional.



Fonte: (VARGAS; PAES; VASCONCELOS, 2016)

Redes neurais convolucionais, CNNs, também conhecidas como redes de convolução, são um tipo específico de rede neural que processa dados representados como dados matriciais, como séries temporais (1-D), imagens (2-D) e vídeo (3D). Como sugerido pelo nome, ao contrário das redes neurais que operam multiplicando matrizes, as redes de convolução usam um tipo particular de operação matemática linear conhecida como convolução (GOODFELLOW; BENGIO; COURVILLE, 2016).

Conforme o nome da rede, as primeiras camadas nas quais os dados de entrada passam são as de convolução, responsáveis por extrair características dessas imagens de entrada, em outras palavras, gerar o que é chamado de mapa de características.

A operação de convolução é obtida através da seguinte equação:

$$s(t) = \int x(a)w(\tau - a)da \quad \text{Eq. 5}$$

De tal maneira que, no contexto de redes convolucionais, o primeiro argumento (função x) é conhecido como a entrada da rede, o segundo argumento (função w) é chamado de kernel ou máscara convolucional e a saída, $s(t)$, é normalmente conhecida como mapa de atributo. Em sua forma discreta, a operação de convolução entre uma imagem bidimensional I e uma máscara K ($I * K$), onde o operador $*$ indica a operação de convolução, poderia ser definida pela equação:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad \text{Eq. 6}$$

Embora as equações as definidas na equação 6 não sejam fielmente utilizadas na maioria do algoritmos de convolução, uma operação semelhante é utilizada, a correlação, que difere da convolução apenas no não rebatimento da máscara, isto é, na operação de flip que então era característica da convolução. A maioria das bibliotecas de aprendizagem de máquina fazem uso da correlação para realizar o processo de convolução (GOODFELLOW; BENGIO; COURVILLE, 2016).

As camadas convolucionais são responsáveis por representar os mesmos processos tradicionais de visão computacional, a aplicação de filtros. Os neurônios pertencentes a essas camadas aplicam os mesmos filtros em janelas de pixels por toda a imagem através do processo de convolução descrito anteriormente.

As CNNs de última geração utilizam a normalização do lote, *batch normalization* (BN), que acontece após a cadência de convolução. Seu principal objetivo é normalizar os mapas de características, mantendo a ativação média próxima 0 e o desvio padrão perto de 1. Essa técnica diminui a sensibilidade à inicialização da rede e acelera o treinamento em rede. Os filtros de cada camada não são definidos; em vez disso, o processo de aprendizagem ajusta os pesos dos neurônios, caracterizando os filtros como resultado. É importante notar que, de acordo com observações feitas após o treinamento, os filtros na camuflagem inicial são mais básicos do que aqueles na camuflagem mais profunda, que têm características mais complexas (CHAVES, 2019).

2.3.1.3. Redes Neurais Siamesas

Uma rede siamesa compreende uma rede composta por duas ou mais sub-redes idênticas que são conectadas por suas saídas, sendo seu conceito introduzido inicialmente para procurar correspondências em imagens em problemas de comparação de assinaturas. Como resultado, cada sub-rede recebe uma entrada exclusiva que é atribuída a um descritor de características. Dois descritores de características são, portanto, obtidos e comparados para ver se são semelhantes, o que causa a redundância da rede. É importante lembrar que, para tornar a saída de cada sub-rede comparável, é necessário que estas redes compartilhem entre si os mesmos parâmetros e pesos (VARGAS; ZHANG; IZQUIERDO, 2020).

Em aplicações cujo objetivo é direcionado para comparação entre dois elementos, como validação de assinaturas ou íris, a arquitetura das redes siamesas fazem uso de duas entradas, ou seja, uma amostra de treinamento compreende um par de dois ou mais elementos, $x^{(1)}, x^{(2)}$ que alimentam as redes gêmeas SN_1, SN_2 , cuja saída é direcionada à rede que as comporta. Nessas aplicações em que as entradas costumam ser imagens, as redes siamesas são redes convolucionais, onde a rede que as incorpora compreende uma rede densa, conectadas por uma camada de comparação entre os dois vetores de características $f(x^{(1)}), f(x^{(2)})$. Isso garante que duas imagens que são extremamente similares não sejam mapeadas em suas respectivas redes para espaços vetoriais distintos, preservando capacidade de representação da semelhança no espaço vetorial (KOCH; ZEMEL; SALAKHUTDINOV, 2015).

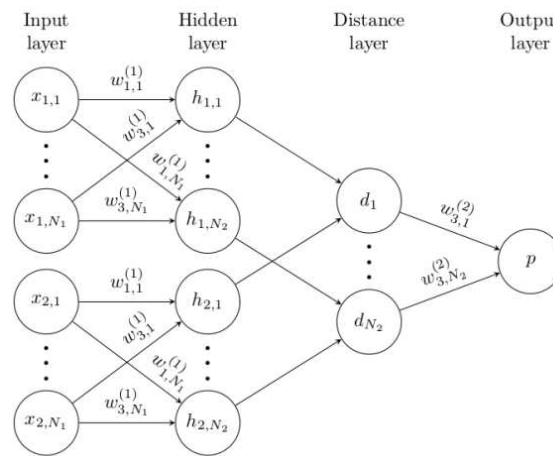
Essa abordagem utiliza os vetores obtidos das redes irmãs para definir uma métrica de similaridade D . Essa métrica é calculada sobre os vetores $f(x^{(1)}), f(x^{(2)})$ por uma camada G que computa diferença absoluta ou outra função de comparação, como a similaridade de cossenos,

entres os dois vetores (VARGAS; ZHANG; IZQUIERDO, 2020). Desse modo, o fluxo de uma rede siamesa até a camada de cálculo de similaridade pode ser compreendida através da equação abaixo.

$$D = G(|f(x^{(1)}) - f(x^{(2)})|) \quad \text{Eq. 7}$$

Ao generalizar o caso de 2 entradas para N entradas, obtém-se uma rede semelhante à da figura 13.

Figura 13: Exemplo de Arquitetura de uma Rede Siamesa.



Fonte: (KOCH; ZEMEL; SALAKHUTDINOV, 2015)

2.3.2. Métricas

2.3.2.1. Contrastive Loss

Nas aplicações cujo objetivo é direcionado para comparação entre dois elementos, baseado na diferença entre dois elementos, encorajando exemplos semelhantes a serem mais próximos e exemplos divergentes a serem distintos por uma distância de pelo menos uma margem m , a função de perda por contraste pode ser aplicada (HADSELL; CHOPRA; LECUN, 2006) e é definida através da seguinte equação:

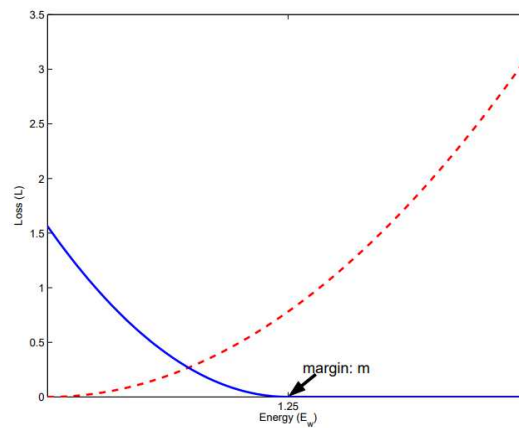
$$L = \frac{1}{2}lD^2 + \frac{1}{2}(1 - l)\{\max(0, m - D)\}^2 \quad \text{Eq. 8}$$

Onde l é um rótulo binário que seleciona se a combinação de entrada, que neste trabalho compreendem as imagem de íris $x^{(1)}$ e $x^{(2)}$, é um par positivo ($l = 1$) ou negativo ($l = 0$),

($m > 0$) é a margem para pares diferentes e $D = |f(x^{(1)}) - f(x^{(2)})|$ é a distância euclidiana entre vetores de características $f(x^{(1)})$ e $f(x^{(2)})$ das imagens de entrada $x^{(1)}$ e $x^{(2)}$.

É possível notar que, para amostras semelhantes, a função de perda aumenta de acordo com a distância entre as duas entradas e, para amostras divergentes, diminui com o inverso da distância entre as duas entradas. A figura 14 expressa os valores da função de perda de acordo com a distância dos dados de entrada.

Figura 14: Perda por Energia para valores semelhantes (vermelho) e divergentes (azul).



Fonte: (HADSELL; CHOPRA; LECUN, 2006)

2.3.2.2. Binary Crossentropy

Outra abordagem para a função de custo é baseada em probabilidades ou estimativas para a adesão à uma classe. Para problemas de caráter binário ou de duas classes, em especial, é possível fragmentar o problema em questões de pertencer ou não a uma das duas classes, calculando a probabilidade de pertencimento e encontrando a de não pertencimento. A função de custo de entropia cruzada binária considera esses aspectos (GUO et al., 2022) e é definida da seguinte forma:

$$H(q) = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)) \quad \text{Eq. 9}$$

Onde y_i indica o rótulo da instância i , $p(y_i)$ a probabilidade da instância pertencer à determinada classe que, para o caso binário, compreende a adesão à classe 1 ou 0.

3. PROJETO

Objetivando a validação de identidade de indivíduos através da biometria baseada em íris, este projeto propõe o desenvolvimento de um modelo de rede neural profunda capaz de rotular dados de imagens oculares.

Os arquivos do projeto se encontram hospedados no gitub do autor e estão acessíveis através do apêndice A.

3.1. FRAMEWORK EXPERIMENTAL

Para o desenvolvimento deste trabalho na direção dos objetivos gerais e específicos, a princípio, buscou-se na literatura bases de dados de íris que simulassem ambientes menos restritos, isto é, ambientes que mais se aproximam das condições reais de captura de imagem, estando sujeitas a ruídos de diversos tipos, como alteração de iluminação, sobreposições de cílios, etc.

Em seguida, executou-se uma busca, também na literatura, a respeito dos métodos utilizados para a extração da íris, a área de interesse. Estes métodos foram aplicados no conjunto de dados e então gerado um novo conjunto contendo apenas dados com a região de interesse destacada. Este conjunto será então fracionado em treinamento, validação e teste nas proporções de 70%, 10% e 20%, respectivamente.

Uma vez possuída a nova base de dados, construiu-se um modelo de classificação baseado em pares de imagem e seus rótulos, identificando se pertencem ou não ao mesmo indivíduo. Este modelo combina as abordagens das redes convolucionais e densas numa arquitetura de rede siamesa.

O treinamento dessa rede é realizado com a fração do conjunto de dados destinado a esse fim, isto é, as frações de treinamento e validação. Concluindo-se o treinamento, o modelo é avaliado no conjunto de teste, onde a análise de seu desempenho será abordada.

3.1.1. Base de Dados

A base de dados UBIRISV2 composta por 11.101 imagens, idealizada por Proença (2010), foi aderida ao projeto,. As imagens foram capturadas em uma sala dedicada, sob iluminação natural e fontes de iluminação artificial.

Figura 15: Exemplos de Imagens Ruidosas.



Fonte: (PROENCA et al., 2010)

Cada imagem faz parte de uma configuração específica, como a seção na qual a imagem foi tirada (primeira ou segunda seção), distância de captura (entre três e dez metros de distância do dispositivo de aquisição) e posições do voluntário sob diferentes ângulos. O intervalo de captura de duas imagens distintas de um mesmo participante compreende durante duas semanas, sendo a segunda captura realizada uma semana após a primeira. Da primeira para a segunda sessão, o local e orientação do dispositivo de aquisição e fontes de luz artificial foi alterada, a fim de aumentar a heterogeneidade. Cerca de 60% dos voluntários realizaram ambas as sessões de imagem, enquanto 40% realizaram exclusivamente um, seja durante o primeiro ou segundo período de aquisição (PROENCA et al., 2010).

Tabela 1: Detalhes do Dataset.

Estrutura e configuração de aquisição de imagem	
Câmera = Canon EOS 5D	Representação de cores = sRGB
Velocidade do Obturador = 1/197 sec.	Abertura da Lente = F/6.4 - F/7
Distância Focal = 400 mm	Número F = F/6.3 - F/7.1
Período de exposição = 1/200 sec.	Velocidade ISO = ISO-1600
Modo de medição = Padrão	
Detalhes das imagens resultantes cortadas manualmente	
Largura = 400 pixels	Altura = 300 pixels
Formato = tiff	Resolução Horizontal = 72 dpi
Resolução Vertical = 72 dpi	Profundidade de Bit = 24 bit
Voluntários	

Total = 261 Voluntários; 522 Íris; 11.102 Imagens	Gêneros = Masculino: 54.4%; Feminino: 45.6%
Idades = [0,20]: 6.6% [21,25]: 32.9% [26,30]: 23.8% [31,35]: 21.0% [36,99]: 15.7%	Pigmentação da Íris = Clara : 18.3% Média : 42.6% Escura: 39.1%

Fonte: Adaptado de (PROENCA et al., 2010)

Cada elemento presente no dataset segue à nomenclatura padronizada da seguinte forma: "C"+ID_EYE+"_S"+ID_SESSION+"_I"+ID_IMAGE+".tiff". A tabela a seguir detalha o significado de cada termo, bem como exemplifica seus possíveis valores.

Tabela 2: Nomenclatura das Imagens.

IDENTIFICADOR	DESCRIÇÃO
"C"+ID_EYE	Identificador do indivíduo. ID_EYE ímpar indica a região do olho direito. Similarmente, seu valor par indica a região esquerda. $ID_EYE=2k$ e $2k + 1$ ($k=1, \dots, \#Indivíduos$) pertencem à mesma pessoa.
"_S"+ID_SESSION	Indica a seção em que o indivíduo participou.
"_I"+ID_IMAGE	Indica a distância em que a imagem foi registrada e a posição do olho. [1,3]: registrada a 8m. [4,6]: registrada a 7m. [7,9]: registrada a 6m. [10,12]: registrada a 5m. [13,15]: registrada a 4m. {1,4,7,10,13}: frontal. {2,5,8,11,14}: deslocado (à esquerda). {3,6,9,12,15}: deslocado (à direita).

Fonte: (PROENCA et al., 2010)

3.1.2. Pré-processamento

Uma vez que a imagem com a região de interesse já em disco, faz-se necessário processar a imagem para que esta seja compatível com a entrada da rede.

O primeiro processo a ser aplicado refere-se à normalização das dimensões das imagens recortadas. Cada imagem que comporta a íris possui valores de largura e altura não padronizados, portanto, foi aplicado o redimensionamento da imagem para 150x150, no formato RGB.

A remoção de cílios, no entanto, mostrou-se uma operação delicada, uma vez que as imagens obtidas sofrem com perda de informação, principalmente em íris cuja cor se aproxima de tonalidades mais escuras, por esse motivo permitiu-se a presença dos cílios na imagem.

A remoção de reflexos luminosos foi realizada através de um filtro passa baixa na imagem em escala cinza, com um valor de limiar igual a 250.

3.1.3. Localização da Íris

Nesta etapa, restringiu-se o número de imagens do dataset para comportar apenas as imagens que foram registradas à distância de 4 metros, isto é, apenas as imagens cujo ID_IMAGE pertence ao intervalo [13,15]. Esta filtragem resultou em imagens mais próximas do indivíduo e, por consequência, mais próximas da íris.

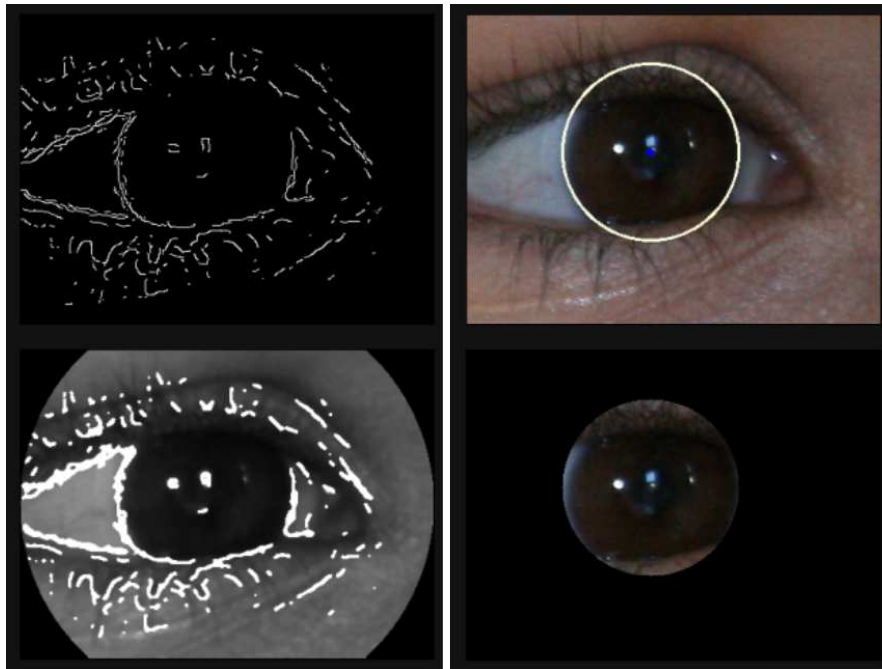
Para a localização da íris, utilizou-se o filtro de Canny para a detecção de bordas. A saída do filtro foi então aprimorada utilizando as transformações morfológicas de abertura, fechamento, erosão e abertura, respectivamente. Posteriormente, aplicou-se a transformada de Hough para a detecção de círculos.

Devido à heterogeneidade dos dados e principalmente pela variedade de posições da íris, abrangeu-se o tamanho do raio da transformada, resultando assim em mais círculos. A escolha do círculo que mais se adequa ao indivíduo baseia-se no fato de que o olho está centralizado na imagem e, portanto, os círculos de interesse também estarão.

Dessa forma, para reduzir os círculos encontrados pela aplicação da transformada ocorre pelo cálculo da distância euclidiana entre o centro da imagem e os centros dos círculos. Dessa forma, garante-se que os círculos com menores valores de distâncias são os que detêm relevância para o estudo. Porém a abordagem falha para as íris que estão levemente deslocadas do centro, para casos como esse é escolhido k círculos mais próximos do centro e então realizada a média aritmética das coordenadas de seus centros, resultando em uma nova coordenada (x_{m3}, y_{m3}) , onde $k = 3$. A coordenada encontrada em relação ao centro da imagem indica a direção do

deslocamento da íris, a distância entre as duas coordenadas é calculada para ajustar a coordenada do centro da imagem, agora tendenciada ao deslocamento da íris. Se a nova busca com o centro ajustado anteriormente resultar em um círculo cuja distância euclidiana for menor que a encontrada pelo centro original da imagem, este círculo é selecionado. A figura 16 aborda uma amostra do dataset durante a aplicação dos procedimentos de localização da íris.

Figura 16: Localização da Íris



Fonte: Autor (2022)

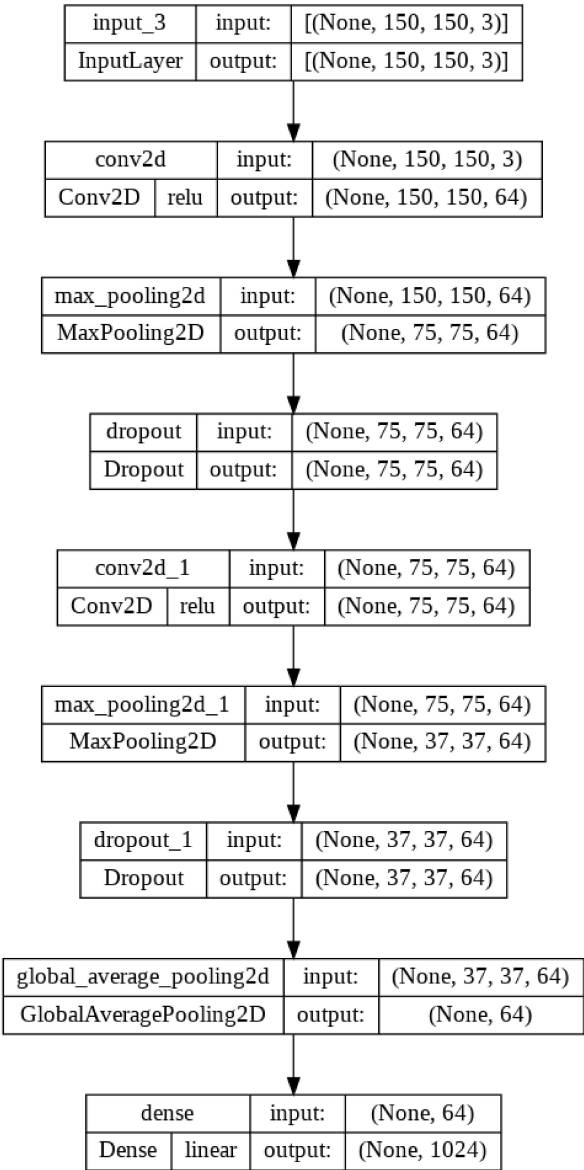
Esta etapa foi realizada de maneira semiautomática, uma vez que a aplicação do processo nem sempre retornou uma segmentação aceitável para as amostras, confirmando as segmentações que contornavam a íris com uma tolerância de 5% de oclusão. As imagens que estavam de acordo com esse critério foram recortadas em um quadrado cujo círculo identificado pelo algoritmo se encontra inscrito nesta região quadrática e as que não se encaixaram ao critério de inclusão foram selecionadas para aplicação posterior do mesmo processo, porém com ajuste nos parâmetros da transformada de Hough e do filtro de Canny. O recorte da imagem é então salvo em uma pasta nomeada com a concatenação da letra “C” com o ID_EYE.

3.1.4. Arquitetura da Proposta

3.1.4.1. Extrator de Características

Para compor a fração da rede que corresponde à extração de características, compondo o conjunto das redes irmãs, ou siamesas, propõe-se uma arquitetura baseada em dois pares de convoluções, redução de dimensionalidade e dropout, seguida uma camada de compressão e uma última de mapeamento para neurônios completamente conectados.

Figura 17: Estrutura do Extrator de Características.

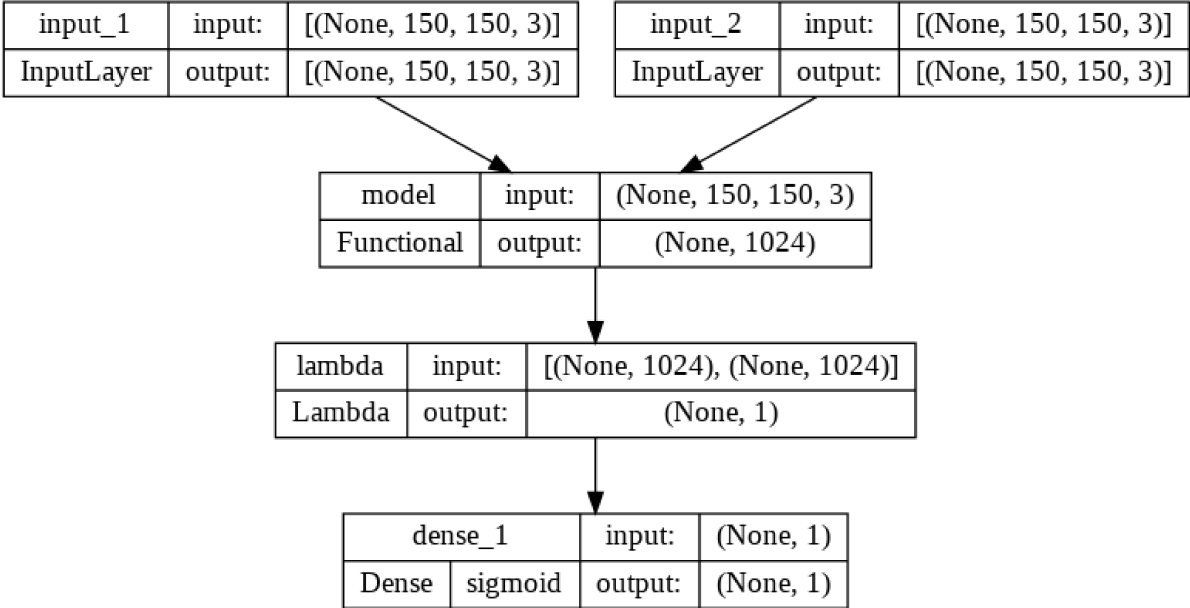


Fonte: Autor (2022)

3.1.4.2. Rede Simples

Para o processamento das características obtidas do modelo extrator, uma rede não muito complexa é utilizada para a validação da relação das duas entradas. A camada funcional computa a distância euclidiana entre os dois tensores de entrada, sendo direcionado a um único neurônio na camada imediatamente posterior.

Figura 18: Estrutura da Rede Simples.

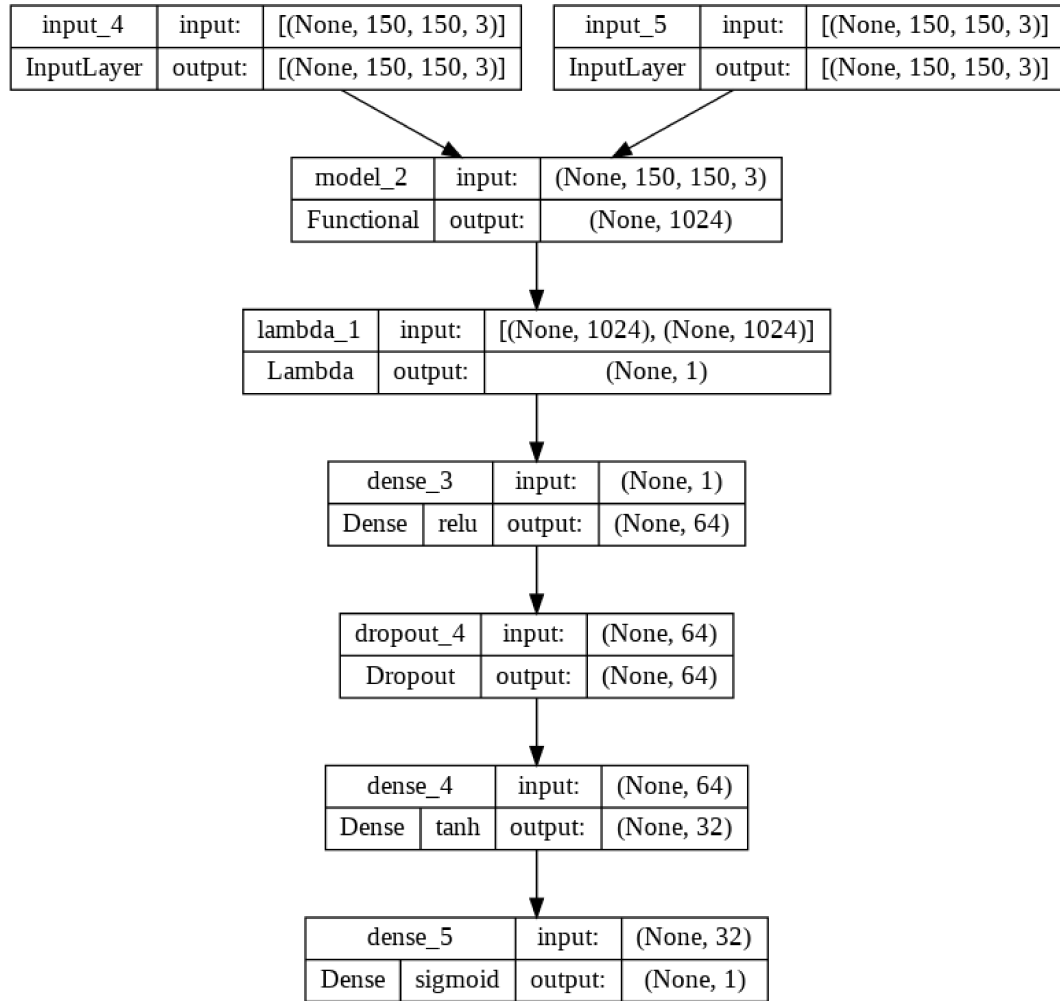


Fonte: Autor (2022)

3.1.4.3. Rede Densa

A efeito de comparação, também para o processamento das características obtidas do modelo extrator, uma rede mais complexa foi projetada. A camada funcional computa a distância euclidiana entre os dois tensores de entrada, sendo direcionado a uma camada densa, seguida de uma camada de dropout e mais duas camadas densas seguintes..

Figura 19: Estrutura da Rede Densa.



Fonte: Autor (2022)

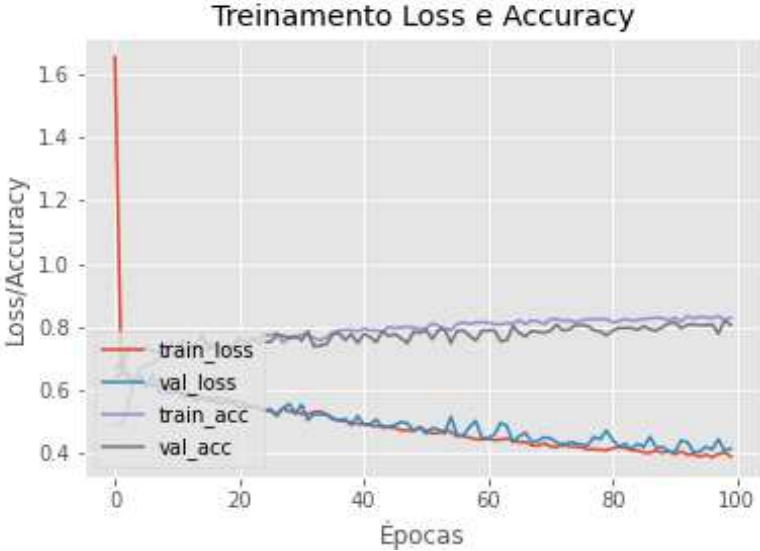
3.2. RESULTADOS

De acordo com os modelos construídos, algumas observações são levantadas quanto ao número de canais da imagem. Ao definir apenas um canal para o conjunto de dados, isto é, converter as amostras para a escala cinza, modelos de mesma arquitetura apresentaram desempenhos inferiores se comparados com sua aplicação na escala BGR, portanto, aderiu-se a escala BGR para a representação dos pixels.

Alguns modelos, principalmente nos modelos baseados na arquitetura mais simples, apresentaram uma maior consistência em seu treinamento, mantendo a precisão dos valores de

treinamento e validação muito próximos. Esses modelos são destacados por sua coesão de classificação que tendem a se traduzir em uma classificação mais confiável em dados de teste.

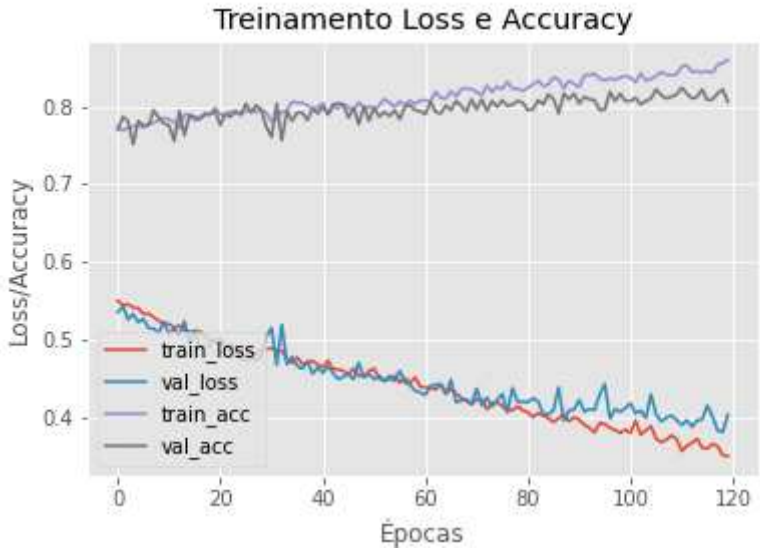
Figura 20: Um dos Treinamentos Consistentes Encontrados.



Fonte: Autor (2022)

O modelo mais bem desempenhado da rede mais simples contou com um treinamento de 7 horas e 30 minutos, aproximadamente. Este modelo, embora tenha apresentado uma precisão em torno dos 80% em dados de validação, apresentou desempenho similar nos dados de teste, apresentando uma precisão de 81,19%.

Figura 21: Treinamento (Rede Simples).

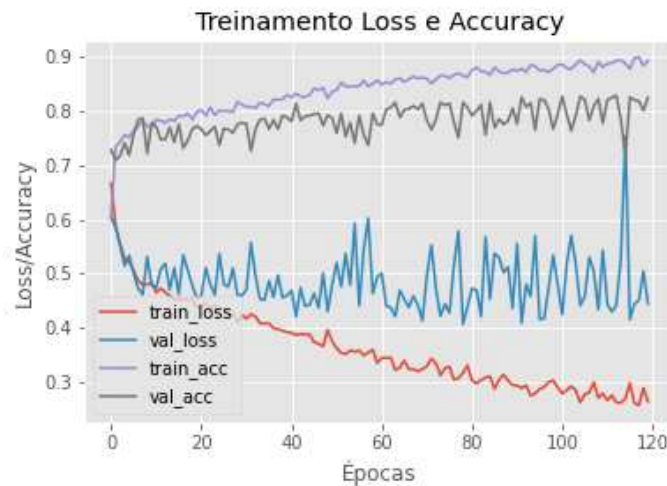


Fonte: Autor (2022)

Ademais, o modelo apresentou uma melhor adaptação ao utilizar a função de entropia cruzada como função de aprendizado ou perda.

O modelo mais bem desempenhado da rede densa, no entanto, contou com um treinamento de 8 horas e 50 minutos, aproximadamente. Este modelo, não apresentou a mesma consistência de treinamento e validação apresentado no modelo anterior, porém atingiu maiores valores de precisão, chegando a 90% em dados de treinamento, 84,51% em dados de validação e 84,18% em dados de teste.

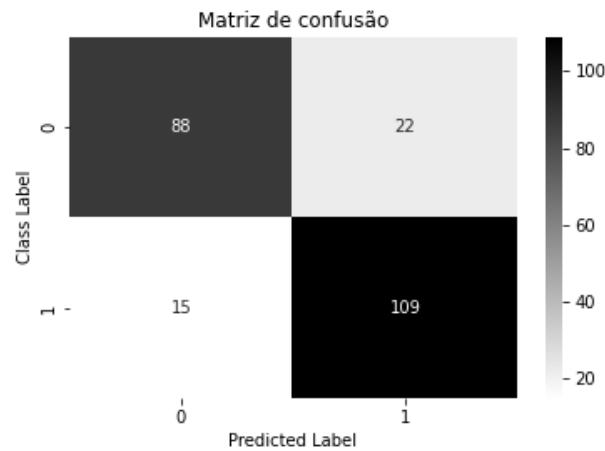
Figura 22: Treinamento (Rede Densa).



Fonte: Autor (2022)

Ao detalhar o desempenho do modelo mais bem adaptado, descrito na figura 22, é possível observar onde ocorre a predominância de seus acertos e erros. A figura 23 descreve a matriz de confusão deste modelo. No geral, o modelo melhor performou na relação verdadeira entre os rótulos, porém, dentre a proporção em que classificou equivocadamente, pior performou na relação entre uma entrada verdadeira e uma saída falsa, isto é, em falsos positivos.

Figura 23: Matriz de Confusão (Rede Densa).



Fonte: Autor (2022)

A tabela a 3 trata de um comparativo entre os dois modelos e explicita valores para treino, teste e validação dos modelos.

Tabela 3: Comparação da Acurácia dos Modelos.

Modelo	Loss	Treino	Validação	Teste
Dense Model	crossentropy	0.90	0.84	0.84
	contrastive	0.44	0.37	0.50
Simple Model	crossentropy	0.86	0.82	0.81
	contrastive	0.20	0.23	0.46

Fonte: Autor (2022)

É possível observar, comparando com a rede simples, que o aumento no número de iterações, as quais estudou-se a convergência até 120 épocas, e da complexidade propuseram um desempenho mais adaptado se comparado à rede mais simples, porém também demonstra um potencial de crescimento à rede mais simples.

Assim como a rede simples, a densa também apresentou queda de desempenho quando a função de perda utilizada era a contrastive loss, que proporcionou a precisão de ambos os modelos beirando os 50%, indicando que o aprendizado por esta função ocorre insatisfatoriamente. Contudo, a performance da função crossentropy indica bom desempenhos, considerando a quantidade de ruído presente, permanecendo em torno dos 80%.

4. CONSIDERAÇÕES FINAIS

Neste trabalho foi abordado o desenvolvimento de um modelo de redes neurais siamesas para a solução da problemática da validação biométrica de um indivíduo baseado em sua íris, capturada em momentos distintos.

Métodos de processamentos baseados em filtros passa baixa ou alta, no geral podem ser aplicados de maneira mais automática, isto é, sem requerer uma validação imediata para cada amostra, diferentemente dos métodos que necessitam de ajuste de parâmetros, como a transformada de Hough.

Embora a base de dados utilizada seja muito próxima de uma aquisição da íris rotineira em função da grande diversidade de ruídos como oclusões, baixa iluminação, interferência por reflexo, etc, o modelo criado apresentou uma performance satisfatória, apresentando 84,18% de precisão quando submetido aos dados de teste.

Foi possível, também, observar a influência da quantidade de canais presentes na imagem sob os modelos de classificação, apresentando, no geral, modelos com desempenho confinando aos 50% para os baseados em apenas um canal, bem como modelos com precisão entre 80% e 84% para aqueles baseados em 3 canais.

Outro aspecto a se destacar consiste no tipo de função de aprendizado utilizada que, contraintuitivamente, revelou modelos mais bem adaptados aqueles que incorporaram a binary crossentropy, a qual demonstrou treinamentos mais consistentes e melhores evoluções do aprendizado.

SUGESTÕES PARA TRABALHOS FUTUROS

A extração da íris é um processo primordial para a identificação biométrica, portanto, estudos em abordagens de localização automática da íris são fundamentais para a eficiência e robustez do modelo desenvolvido, que, neste trabalho, abordou a estratégia semiautomática.

Neste trabalho, a rede modelada para a classificação não possui uma alta complexidade se comparada com redes de Deep Learning em função dos recursos computacionais disponíveis, portanto, abordagens em redes profundas podem apresentar resultados promissores.

A remoção de cílios é um processo delicado que requer uma verificação adicional para garantir que informações relevantes não sejam perdidas, fato esse que dificulta a generalização de métodos para tratar dessa finalidade, portanto, reforços no pré-processamento podem agregar um melhor aprendizado em função da retenção das informações relevantes.

REFERÊNCIAS

- [1] BERGUES, G. J. et al. Sub-pixel Gray-scale Hough Transform For An Electronic Visual Interface. **IEEE Latin America Transactions**, v. 13, n. 9, p. 3135–3141, 2015.
- [2] CHAVES, E. DE L. Detecção de câncer de mama por meio de imagens infravermelhas utilizando Redes Neurais Convolucionais. 2019.
- [3] DE SOUSA, M. H. S. et al. Contagem de objetos utilizando técnicas de processamento digital de imagem: Morfologia Matemática. **Brazilian Journal of Development**, v. 6, n. 8, p. 57432–57440, 2020.
- [4] DUARTE, G. D. Uso da Transformada de Hough na Detecção de Círculos em Imagens Digitais. **Thema-Revista Científica do Centro Federal de Educação Tecnológica**, v. 4, n. 1, p. 51–58, 2003.
- [5] DUARTE, R. et al. Aplicação dos Sistemas Biométricos de Reconhecimento Facial na Segurança Pública. **Brazilian Journal of Forensic Sciences, Medical Law and Bioethics**, v. 11, n. 1, p. 1–21, 2021.
- [6] GHALI, A. A. et al. A review of iris recognition algorithms. **JOIV: International Journal on Informatics Visualization**, v. 1, n. 4–2, p. 175–178, 2017.
- [7] GONZALEZ, R. C.; WOODS, R. C. **Processamento digital de imagens**. 3. ed. [s.l.] Pearson Educación, 2009.
- [8] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Disponível em: <<https://www.deeplearningbook.org/>>. Acesso em: 26 out. 2022.
- [9] GUO, C. et al. Multi-Stage Attentive Network for Motion Deblurring via Binary Cross-Entropy Loss. **Entropy**, v. 24, n. 10, p. 1414, 2022.
- [10] HADSELL, R.; CHOPRA, S.; LECUN, Y. **Dimensionality reduction by learning an invariant mapping**. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Anais. IEEE, 2006.
- [11] JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An Introduction to Biometric Recognition. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 14, n. 1, p. 4–20, jan. 2004.
- [12] KOCH, G.; ZEMEL, R.; SALAKHUTDINOV, R. **Siamese Neural Networks for One-shot Image Recognition**. v. 2, 2015.
- [13] MALTEZ, J.; DO VALE, L. Processamento de Imagem para Reconhecimento de Padrões na Íris. [s.d.].
- [14] NETO, D.; NETO, R. **Estudo comparativo entre dois métodos de localização da fronteira externa da íris: um estudo de caso**. Anais da XIX Escola Regional de Computação Bahia, Alagoas e Sergipe. **Anais...SBC**, 2019.

- [15] NOSRATABADI, S. et al. **Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS**. Agriculture, v. 11, n. 5, p. 408, 2021.
- [16] **OpenCV: Morphological Transformations**. Disponível em: <https://docs.opencv.org/4.x/d9/d61/tutorial_py_morphological_ops.html>. Acesso em: 20 dez. 2022.
- [17] PATIL, P. M. Iris recognition in less constrained environment. **International Journal of Emerging Technology and Advanced Engineering**, v. 3, n. 7, p. 196–200, 2013.
- [18] PONTI, M. A.; DA COSTA, G. B. P. **Como funciona o deep learning**. arXiv preprint arXiv:1806.07908, 2018.
- [19] PROENCA, 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**, v. 32, n. 8, p. 1529–1535, ago. 2010.
- [20] RATTANI, A.; DERAKHSHANI, R. Ocular biometrics in the visible spectrum: A survey. **Image and Vision Computing**, v. 59, p. 1–16, 2017.
- [21] VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. **Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres**. Proceedings of the xxix conference on graphics, patterns and images. **Anais...sn**, 2016.
- [22] VARGAS, C.; ZHANG, Q.; IZQUIERDO, E. **One shot logo recognition based on siamese neural networks**. Proceedings of the 2020 International Conference on Multimedia Retrieval. **Anais.2020**.
- [23] VIZONI, M. V. Reconhecimento da região ocular para a identificação biométrica de pessoas utilizando aprendizado em profundidade. 2019

GLOSSÁRIO

dataset	Também chamado de banco de dados, é o conjunto de arquivos ou um arquivo que indexa outros arquivos, utilizado para treinamento e teste de modelos de aprendizado de máquina.
função de perda	Função a qual compara as classes dos valores reais e as classes preditas pelo modelo
rede densa	Rede completamente conectada

APÊNDICE A - Repositório Github

[ThyagoFRTS/Iris-Classifier: Modelo de classificação baseado em pares de imagem e seus rótulos, identificando se pertencem ou não ao mesmo indivíduo. Este modelo combina as abordagens das redes convolucionais e densas numa arquitetura de rede siamesa. \(github.com\)](#)