



UNIVERSIDADE DO ESTADO DO AMAZONAS – UEA
ESCOLA SUPERIOR DE TECNOLOGIA – EST
BACHARELADO EM ENGENHARIA ELÉTRICA



APLICAÇÃO DE TÉCNICAS DE DEEP LEARNING PARA AUTENTICAÇÃO DE USUÁRIOS
ATRAVÉS DO RECONHECIMENTO FACIAL EM NOTEBOOKS

Monique Sousa Silva

MANAUS – AM
2023

Monique Sousa Silva

APLICAÇÃO DE TÉCNICAS DE DEEP LEARNING PARA AUTENTICAÇÃO DE USUÁRIOS
ATRAVÉS DO RECONHECIMENTO FACIAL EM NOTEBOOKS

Trabalho de Conclusão de Curso apresentado à banca avaliadora do curso de Bacharelado Engenharia Elétrica da Escola Superior de Tecnologia da Universidade do Estado do Amazonas, como pré-requisito para obtenção do título de bacharel em Engenharia Elétrica. Orientador: Prof. Dr. Jozias Parente de Oliveira.

MANAUS – AM

2023

Universidade do Estado do Amazonas – UEA
Escola Superior de Tecnologia – EST

Reitor:

André Luiz Nunes Zogahib

Vice-Reitor:

Kátia do Nascimento Couceiro

Diretor da Escola Superior de Tecnologia:

Ingrid Sammyne Gadelha Figueiredo

Coordenador do Curso de Engenharia Elétrica:

Israel Gondres Torné

Banca avaliadora composta por:

Prof. Jozias Parente de Oliveira (Orientador)

Prof. Rubens de Andrade Fernandes

Prof. Antônio Luiz Alencar Pantoja

CIP – Catalogação na Publicação

Silva, Monique S.,

Aplicação de técnicas de *deep learning* para autenticação de usuários através do reconhecimento facial em *notebooks* / Monique Sousa Silva; [orientada por] Jozias Parente de Oliveira. – Manaus: 2023.

56 p.: il.

Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica). Universidade do Estado do Amazonas, 2023.

1. Processamento e análise de imagens. 2. Reconhecimento de padrões. 3. *Deep learning*. 4. *Machine learning*.

Monique Sousa Silva

APLICAÇÃO DE TÉCNICAS DE DEEP LEARNING PARA AUTENTICAÇÃO DE USUÁRIOS
ATRAVÉS DO RECONHECIMENTO FACIAL EM NOTEBOOKS

Trabalho de Conclusão de Curso apresentado à banca avaliadora do curso de Bacharelado Engenharia Elétrica da Escola Superior de Tecnologia da Universidade do Estado do Amazonas, como pré-requisito para obtenção do título de bacharel em Engenharia Elétrica.

Nota obtida: 9,8 (nove vírgula oito)

Aprovada em: 01 / 09 / 2023

BANCA EXAMINADORA

JOZIAS PARENTE DE
OLIVEIRA:44545142249

Assinado de forma digital por JOZIAS PARENTE DE
OLIVEIRA:44545142249
Dados: 2023.09.06 12:01:36 -03'00'

Prof. Dr. Jozias Parente de Oliveira, Presidente

Documento assinado digitalmente



RUBENS DE ANDRADE FERNANDES
Data: 03/09/2023 22:42:58 -0300
Verifique em <https://validar.iti.gov.br>

Prof. M.Sc. Rubens de Andrade Fernandes, Membro

Antonio Luiz Alencar
Pantoja

Assinado de forma digital por Antonio
Luiz Alencar Pantoja
Dados: 2023.09.03 19:37:11 -04'00'

Prof. Antônio Luiz Alencar Pantoja, Membro

MANAUS – AM

2023

Este trabalho é dedicado aos que não precisam necessariamente de coisas novas para se sentirem ansiosos, não permita que seus medos sejam maiores que você.

AGRADECIMENTOS

Agradeço aos meus pais Maria Vilma Xavier de Sousa Silva e João Fábio Santos da Silva, que proporcionaram em estrutura, cobranças e incentivos, o melhor que foi possível para o meu desenvolvimento acadêmico. Sem dúvida, sem o suporte vindo de casa, tudo teria sido mais difícil.

Agradeço ao meu amigo RAFEL CASTRO MELO, que se dedicou implacavelmente para manter a minha atenção e progresso de maneira assertiva em prol da conclusão deste trabalho de conclusão de curso. Muito obrigada por ceder seu tempo, por manter-se impassível frente aos meus protestos para desaceleração do avanço do trabalho, e por relevar os momentos em que agi ou falei por estresse ou impaciência. Certamente você foi a pessoa que mais me oferece suporte e confiança na conclusão das últimas fases deste trabalho.

Agradeço aos meus amigos Esther de Carvalho Costa, Mauricio Souza Calheiro, Igor Benayon Cunha, Renan Castiel Coutinho e Sol Petrus Praia Walkey, por tornarem dias ruins mais suportáveis e dias bons em excelentes, pelo seu incentivo na disciplina e no aprendizado, por terem me dado tantas forças para não desistir ao longo do curso, por compartilharem tantos momentos comigo, e por permitir que pudéssemos cruzar nossos caminhos, mesmo que somente por algum tempo. Participar de um grupo com seus melhores amigos, foi o que me permitiu chegar tão longe e tão rápido.

Agradeço aos meus amigos Samuel Bruno Torres Rego, Roberto Conhago Tavares de Sousa, Rubens de Andrade Fernandes, Thalyta Reis de Brito, Joêmio Tavares Freire Junior, e muitos outros, que mesmo não precisando, sempre tiveram a paciência para explicar alguma questão de listas de exercícios quando eu não entendia. Agradeço também aos colegas que vieram do curso técnico para a faculdade, e a tantos outros colegas que participaram do meu aprendizado e da minha vida.

Agradeço aos professores Jozias Parente de Oliveira, Ingrid Sammyne Gadelha Figueiredo e Cleto Cavalcante de Souza Leal, que ministraram algumas das melhores aulas da minha vida, seus métodos de aula foram imprescindíveis para que eu pudesse absorver tanto conhecimento. Agradeço também por entenderem o lado humano dos alunos do curso, e por apenas existirem, e ocuparem o corpo docente com as suas presenças.

Agradeço também aos técnicos administrativos da instituição, que podem se tornar invisíveis a muitos olhos, mas que foram essenciais para conservar a faculdade em pleno

funcionamento. Agradeço especialmente a técnica administrativa Araci de Carvalho Freitas, que guiou muitos estudantes perdidos sobre os preceitos do curso.

RESUMO

O presente trabalho visa aplicar técnicas de processamento digital de imagens e redes neurais para realizar a autenticação de usuários e desbloqueio de *notebooks*. Para isto, utilizou-se de conhecimentos do algoritmo de Viola-Jones, que aplica as características retangulares de Haar para tornar imagens complexas mais simples para análise, e do Algoritmo de *Adaptive Boosting*, de forma a classificar o peso das características identificadas nas imagens. Também se realizou uma abordagem de *deep learning* associada a redes neurais artificiais para reforçar a identificação correta dos rostos utilizados para a verificação do sistema, de forma a compreender melhor o funcionamento por trás da manipulação do sinal digital de imagem para a sua análise. Para validação e testes do sistema, foram utilizados um banco com 700 imagens em domínio público adquiridas de 50 usuários distintos, utilizando métricas de avaliação como acurácia, precisão, sensibilidade, especificidade e *F-score*. De acordo com os resultados, obteve-se uma acurácia mínima igual a 98,6% e máxima igual a 100%, uma precisão mínima igual a 62,5% e máxima igual a 100%, uma sensibilidade mínima igual a 50% e a máxima igual a 100%, uma especificidade mínima igual a 98,5% e máxima igual a 100% e *F-score* mínima igual a 66,7% e a máxima igual a 100%.

Palavras-chave: Processamento e análise de imagens. Reconhecimento de padrões. *Deep learning*. *Machine learning*.

ABSTRACT

The present work aims to apply techniques of digital image processing and neural networks to authenticate and unlock notebooks. For this, knowledge of the Viola-Jones algorithm was used on this project, which applies Haar's rectangular features to make complex images simpler for analysis, and of the Adaptive Boosting Algorithm, to classify the weight of the features identified in the images. A deep learning approach associated with artificial neural networks was also conducted to reinforce the correct identification of the faces used to verify the system, and to better understand the operation behind the manipulation of the digital image signal for its analysis. For validation and testing of the system, a dataset of 700 publicly available images from 50 distinct users was employed, utilizing evaluation metrics such as accuracy, precision, sensitivity, specificity, and F-score. According to the results, a minimum accuracy of 98,6% and a maximum of 100% were achieved, along with a minimum precision of 62,5% and a maximum of 100%. Sensitivity ranged from a minimum of 50% to a maximum of 100%, specificity ranged from a minimum of 98,5% to a maximum of 100%, and F-score ranged from a minimum of 66,7% to a maximum of 100%.

Keywords: Image processing and analysis. Pattern recognition. Deep learning. Machine learning.

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1: Sequência do padrão de processamento e análise digital de imagens (PADI)..... | 9 |
| Figura 2: Representação gráfica de um sistema de captura..... | 10 |
| Figura 3: Extração de atributos e características das imagens..... | 14 |
| Figura 4: Subdivisão do sistema automático de reconhecimento em blocos..... | 14 |
| Figura 5: Aplicação das características de Haar em um rosto..... | 16 |
| Figura 6: Características retangulares de Haar em imagens integrais..... | 16 |
| Figura 7: Cascata de classificadores de Viola-Jones..... | 17 |
| Figura 8: Modelo de neurônio artificial..... | 18 |
| Figura 9: Uma arquitetura de CNN de dois estágios..... | 18 |
| Figura 10: Arquitetura geral de uma CNN..... | 19 |
| Figura 11: Arquitetura da rede neural AlexNet..... | 22 |
| Figura 12: Etapas da metodologia desenvolvida no projeto..... | 23 |
| Figura 13: Banco de dados fotográfico de faces..... | 23 |
| Figura 14: Divisão do banco de dados entre imagens de teste e base para o sistema..... | 24 |
| Figura 15: Arquitetura do projeto desenvolvido..... | 24 |
| Figura 16: Interface do sistema com o usuário..... | 25 |
| Figura 17: Código para treinamento da AlexNet..... | 30 |
| Figura 18: Diretório dos arquivos de teste..... | 30 |
| Figura 19: Código para registro de usuários..... | 31 |
| Figura 20: Fluxograma do processo de registro de novos usuários..... | 31 |
| Figura 21: Código para autenticação de usuário..... | 32 |
| Figura 22: Fluxograma do processo de autenticação de usuário..... | 32 |
| Figura 23: Imagens de teste para simulação do ambiente de autenticação facial..... | 33 |
| Figura 24: Imagens de teste cuja CNN não identificou o usuário corretamente..... | 34 |
| Figura 25: Gráfico da acurácia (ACC) X classe..... | 34 |
| Figura 26: Gráfico da precisão X classe..... | 35 |
| Figura 27: Gráfico da sensibilidade (TPR) X classe..... | 35 |
| Figura 28: Gráfico da especificidade (TNR) X classe..... | 35 |
| Figura 29: Gráfico F-score X classe..... | 36 |

Figura 30: Gráfico dos erros X classe, sendo os erros discriminados entre falso negativos e falso positivos..... 39

LISTA DE SIGLAS E ABREVIATURAS

| | |
|----------|--|
| AdaBoost | <i>Adaptative Boosting</i> |
| ADI | Análise Digital de Imagens |
| CCD | <i>Charged Couple Device</i> |
| CNN | <i>Convolutional Neural Network</i> |
| COVID-19 | <i>Corona Virus Disease 2019</i> |
| FN | Falso negativo |
| FNR | <i>False Negative Rate</i> |
| FP | Falso positivo |
| FPR | <i>False Positive Rate</i> |
| IDLE | <i>Integrated Development and Learning Environment</i> |
| MatLab | <i>Matrix Laboratory</i> |
| PADI | Processamento e Análise de Digital de Imagens |
| PDI | Processamento Digital de Imagens |
| PDS | Processamento Digital de Sinais |
| PIN | <i>Personal Identification Number</i> |
| ReLU | <i>Rectified Linear Unit</i> |
| RNA | Rede Neural Artificial |
| SCV | Sinal Composto de Vídeo |
| TNR | <i>True Negative Rate</i> |
| TPR | <i>True Positive Rate</i> |
| UEA | Universidade do Estado do Amazonas |
| VN | Verdadeiro positivo |
| VP | Verdadeiro negativo |

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO | 3 |
| 1.1 | Objetivos | 6 |
| 1.1.1 | Objetivo geral..... | 6 |
| 1.1.2 | Objetivos específicos | 6 |
| 2 | REFERENCIAL TEÓRICO | 8 |
| 2.1 | Visão computacional..... | 8 |
| 2.2 | Processamento e análise digital de imagens..... | 8 |
| 2.2.1 | Aquisição da imagem..... | 9 |
| 2.2.2 | Digitalização da imagem..... | 10 |
| 2.2.3 | Pré-processamento da imagem..... | 11 |
| 2.2.4 | Segmentação em regiões | 11 |
| 2.2.5 | Pós-processamento | 12 |
| 2.2.6 | Extração de atributos..... | 12 |
| 2.2.7 | Classificação e reconhecimento | 13 |
| 2.3 | Aplicação em autenticação biométrica por reconhecimento facial..... | 13 |
| 2.3.1 | O algoritmo de Viola-Jones | 15 |
| 2.3.1.1 | Características de Haar e integral de imagem | 15 |
| 2.3.1.2 | Classificador e cascata de características | 17 |
| 2.3.2 | Rede neural artificial | 17 |
| 3 | MATERIAIS E MÉTODOS..... | 21 |
| 3.1 | Ambiente de desenvolvimento | 21 |
| 3.2 | Ferramentas de desenvolvimento..... | 21 |
| 3.3 | Métodos..... | 22 |
| 3.3.1 | Preparação do conjunto de dados | 23 |
| 3.3.2 | Definição da arquitetura..... | 24 |
| 3.3.3 | Implementação da arquitetura | 25 |
| 3.3.4 | Avaliação da arquitetura | 26 |
| 3.3.4.1 | Matriz de confusão | 26 |
| 3.3.4.2 | Métricas que avaliam a qualidade do classificador | 27 |
| 3.3.5 | Análise dos resultados..... | 28 |

| | | |
|----------|--|-----------|
| 4 | IMPLEMENTAÇÃO | 29 |
| 4.1 | Arquitetura do sistema | 29 |
| 4.1.1 | Treinamento da rede neural convolucional | 29 |
| 4.1.2 | Registro de novo usuário..... | 30 |
| 4.1.3 | Autenticação de usuário | 32 |
| 5 | RESULTADOS E DISCUSSÕES | 33 |
| 5.1 | Avaliação qualitativa..... | 33 |
| 5.2 | Avaliação Quantitativa..... | 34 |
| 5.3 | Tempo de Processamento..... | 40 |
| 6 | CONCLUSÃO | 41 |
| | REFERÊNCIAS | 43 |

1 INTRODUÇÃO

Segurança da informação é uma área que tem evoluído e buscado incorporar diferentes técnicas e métodos para realizar o controle de acesso de dados. Em um mundo globalizado e cercado de tecnologia, onde a produção e circulação de dados é grande, a preocupação com este tema somente cresce, sempre buscando aperfeiçoar sistemas de autenticações mais seguros, rápidos e práticos. Neste cenário, originaram-se os sistemas biométricos, que revolucionaram a experiência de autenticação do usuário e estão presentes progressivamente nos mais diversos dispositivos, como *smartphones*, computadores, fechaduras eletrônicas, entre outros (Biesseck & Zacarkim, 2021).

Um destaque em sistemas de autenticação biométricos é o método via reconhecimento facial, que utiliza de conhecimentos em visão computacional para identificar as características presentes no rosto, como formato dos olhos, distância entre nariz e boca, identificação da pele, entre outros. Este método se diferencia dos demais por somente necessitar de uma câmera, elemento comum em vários equipamentos, e por não exigir contato direto do usuário com o sensor (BATISTA, *et al.*, 2017).

Apesar da distinção do reconhecimento facial em relação a outros métodos, a maioria das estações de trabalho com computador, mesmo que acompanhadas de *webcam*, não oferecem o serviço nativamente. Dependendo das informações contidas e da necessidade do usuário, os principais métodos de autenticação como senhas, *Personal Identification Number* (PIN), e outros padrões, não são suficientes para garantir um nível adequado de segurança em certos cenários. Nestes sistemas, existe o risco da utilização de chaves de segurança sem robustez, como sequência de números, padrões conhecidos, o próprio nome, e outros recursos que podem facilmente ser descobertos por terceiros mal-intencionados (Magalhães & Santos, 2003).

Os meios de controle de acesso tradicionais possuem o risco de transmissibilidade, mesmo que não voluntário, a terceiros, tendo em vista que a utilização de chaves considerados fortes acarretam padrões de difícil memorização, sendo salvos em diretórios ou arquivos de fácil acesso a outros, sem a criptografia adequada (Magalhães & Santos, 2003).

Além disso, os principais métodos de autenticação não são práticos e rápidos para a identificação e validação rápida de usuários autorizados, exigindo o contato direto do usuário com a interface. A preparação e instalação de algumas autenticações biométricas são mais complexas e com custo elevado, visto que podem exigir equipamentos específicos como os utilizados para reconhecimento de retina ou impressão digital. Desta forma, este trabalho

buscou implementar um sistema de reconhecimento facial em uma estação de trabalho com computador e com uma *webcam*, aplicando conhecimentos de processamento digital de imagens, e visando o aumento na segurança dessas estações de trabalho. Além disso, almeja-se o aumento da praticidade no acesso à informação por pessoas autorizadas.

Tendo em vista a problemática supracitada, partiu-se da hipótese de que seria possível o desenvolvimento de um sistema de reconhecimento biométrico facial automático para aplicações de controle de acesso em estações de trabalho utilizando técnicas de processamento digital de imagem e *deep learning* para aquisição, segmentação e comparação de imagens com um banco de dados por meio da aplicação de redes neurais convolucionais. Para validação dessa proposta, será utilizado a plataforma *Matrix Laboratory* (MatLab) instalada em um *notebook* com *webcam* incluso e um banco de imagens que será elaborado ao logo do projeto. O diferencial deste projeto consta na utilização de uma rede neural artificial pré-treinada, que se propõe a oferecer uma taxa de acertos elevada com menos treinamentos realizados para aprendizagem da rede neural.

O presente trabalho tem como justificativa o fato de que sistemas de autenticação biométrica estão cada vez mais presentes nos dispositivos utilizados cotidianamente, onde é comum a presença de dispositivos com recursos para tal, como, por exemplo, nos telefones móveis. Esta tendência se dá por conta da praticidade destes métodos, que facilitam o acesso de pessoas autorizadas sem comprometer com a experiência final do usuário e com a segurança das informações por eles protegidos.

Para o reconhecimento facial em estações de trabalho com os recursos necessários de *hardware*, o desenvolvimento e implementação do método proposto em sistemas que ainda não o possuem é plenamente viável. Dessa forma, contribui-se para uma robustez adicional e necessária na segurança da informação presente. O trabalho pode trazer benefícios para diversas situações. Por exemplo, em *notebooks* de trabalho que lidam com um tráfego diário de informações sensíveis, o projeto pode aprimorar a segurança e a eficiência na gestão dos dados. Para dispositivos pessoais, onde se busca maior segurança e praticidade no acesso às informações, a implementação do projeto pode oferecer soluções aprimoradas. Além disso, em máquinas industriais integradas a dispositivos computacionais, onde é crucial limitar o acesso apenas a profissionais habilitados para operá-las, o projeto também pode proporcionar uma valiosa contribuição.

Ao utilizar o método de validação biométrica em conjunto com os métodos de segurança pré-existentes para a segurança de sistemas computadorizados, adiciona-se uma barreira a mais para evitar o acesso de pessoas não autorizadas a conteúdos sensíveis presentes no sistema. Por outro lado, ao substituir os métodos de segurança existentes pelo método de validação biométrica para a segurança de sistemas computadorizados, consegue-se um nível de segurança satisfatório e mais prático.

O reconhecimento biométrico facial é realizado automaticamente, agilizando o processo de identificação e autenticação, além de não necessitar de contato direto com a interface, oferecendo maior acessibilidade para pessoas que não podem utilizar de sistemas como a verificação via impressão digital devido à perda dos membros superiores ou devido ao histórico de trabalho com frutas ácidas como a laranja, que com o tempo lixa a camada córnea da pele (Silva, 2023).

O desenvolvimento do projeto de reconhecimento facial aperfeiçoa as habilidades técnicas do projetista, que deve ponderar sobre a melhor técnica a ser utilizada dado as bases, ao ambiente e aos objetivos do projeto, deve se atualizar quanto às estratégias para a realização do processo de identificação e autenticação, deve realizar verificações ou testes para aperfeiçoar o projeto antes da disponibilização e implementação e garantir o desempenho correto do sistema, e deve trabalhar junto à comunidade para desenvolver um programa factível e longo.

A visão computacional é uma área aonde a aplicabilidade de conceitos de reconhecimento facial tende a crescer, assim como o aprimoramento das técnicas utilizadas para reconhecimento e identificação do usuário, que sempre buscam obter a melhor eficiência mesmo em situações desfavoráveis.

A partir desta necessidade, os cursos técnicos e superiores vêm incorporando cada vez mais disciplinas que ofereçam embasamento teórico relacionado a visão computacional para seus discentes, como por exemplo as disciplinas de Processamento Digital de Sinais (PDS) e Processamento Digital de Imagens (PDI), que são disponibilizadas no curso de Engenharia Elétrica da Universidade do Estado do Amazonas (UEA), e servirão de base para este projeto, fornecendo conhecimentos como técnicas de correlação, obtenção de imagens, aplicação de conhecimentos em processamento e análise em imagens, entre outros.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Desenvolver um sistema de reconhecimento facial para autenticação de usuário e desbloqueio em estações de trabalho com computador e com uma *webcam* visando o aumento da segurança ao acesso de informações por usuários cadastrados na base de dados.

1.1.2 Objetivos específicos

- Reunir referencial bibliográfico para abordagem e desenvolvimento do tema;
- Determinar a abordagem e delimitação do tema da pesquisa, considerando os conhecimentos básicos adquiridos durante o período do curso;
- Definir a metodologia de reconhecimento facial mais apropriada para estações de trabalho com computadores e *webcams*;
- Desenvolver uma plataforma de reconhecimento facial;
- Elaborar o banco de dados fotográfico da plataforma para autenticação de usuários;
- Implementar o sistema em uma estação de trabalho com computador e *webcam*;
- Realizar testes de verificação do desempenho da proposta;
- Analisar se o desempenho e retorno do projeto aplicado são satisfatórios quando comparado com outras soluções do mercado.

O presente trabalho está organizado em 5 capítulos, dispostos da seguinte maneira:

- a) Referencial teórico: Neste tópico serão abordados o ramo da visão computacional, destrinchando-a em etapas básicas de processamento como a aquisição, processamento, análise e o reconhecimento de padrões; a aplicação do sistema de visão computacional em cenários de autenticação biométrica; e a metodologia de *deep learning* no panorama da visão computacional, de forma a refinar o reconhecimento de usuários e ajustar os padrões de conhecimentos incorretos.
- b) Materiais e métodos: Neste tópico, serão abordadas as ferramentas que serão utilizadas no desenvolvimento deste projeto, descrevendo o ambiente e as aplicações empregadas, considerando o conjunto de dados, ferramentas, bem como métricas de análise de performance do projeto.

- c) Implementação: Neste tópico, serão abordadas a arquitetura para a implementação do treinamento da rede neural convolucional, registro e autenticação de usuários, discriminando as abordagens lógicas para a efetivação do projeto.
- d) Resultados e discussões: Neste tópico, será feita uma análise quantitativa e qualitativa do projeto desenvolvido, aferindo as métricas de avaliação de desempenho do projeto e o tempo de processamento para as etapas de treinamento, registro e autenticação de usuários.
- e) Conclusão: Neste tópico, serão consolidados os objetivos, métodos e resultados obtidos do projeto, identificando as aplicações práticas do projeto e realizando uma análise do cenário em pesquisa, enumerando algumas implicações práticas ou futuras com base em pesquisas relacionadas a área de visão computacional, correlatas a este projeto.

2 REFERENCIAL TEÓRICO

2.1 VISÃO COMPUTACIONAL

Visão computacional é um ramo da ciência que caracteriza a forma com que a tecnologia, seja computador ou dispositivo, interpreta o cenário em que analisa, para então decompor a imagem, obter as partes significantes e manejar o conhecimento com base em sua aplicação (Milano & Honorato, 2014). Um dos objetivos da visão computacional é aproximar o seu desempenho com o sistema de visão humana, sendo uma motivação para isto a capacidade da visão humana, que foi resultado de um grande processo evolutivo biológico (Helene & Helene, 2011), interpretando os sinais visuais com relativa eficiência e rapidez.

Outra motivação provém da característica operacional dos sistemas computacionais, que devem traduzir e comunicar com os humanos, como por exemplo a verificação e classificação de cores presentes em uma imagem, onde a máquina deve adquirir a imagem em questão, tratá-la para evitar distorções e aumentar a eficiência do sistema, verificar e classificar a presença de cores da imagem (azul, verde, vermelho, ...), e entregar esta informação para o usuário, sistema que a visão humana já realiza.

Entretanto, em alguns aspectos o sistema computacional se desenvolveu tanto que por vezes supera o sistema visual humano, como a recepção e análise de espectros de luz fora da faixa visível, em situações de iluminação desfavorável como à noite, determinação da exata faixa de frequência correspondente a cor visível, entre outros. Desta forma, pode-se desenvolver sistemas baseados em visão computacional, que podem funcionar ininterruptamente ou de forma inteligente, e obterem e condensarem informações para as mais diversas aplicabilidades.

Para alcançar o objetivo de aproximar, ou até mesmo superar, o sistema visual humano, a visão computacional constantemente desenvolve e integra áreas como física, matemática, processamento digital, inteligência artificial, entre outros (Autores, 2012). Para isto, primeiro se deve levar em consideração a aplicabilidade e o resultado esperado para o projeto.

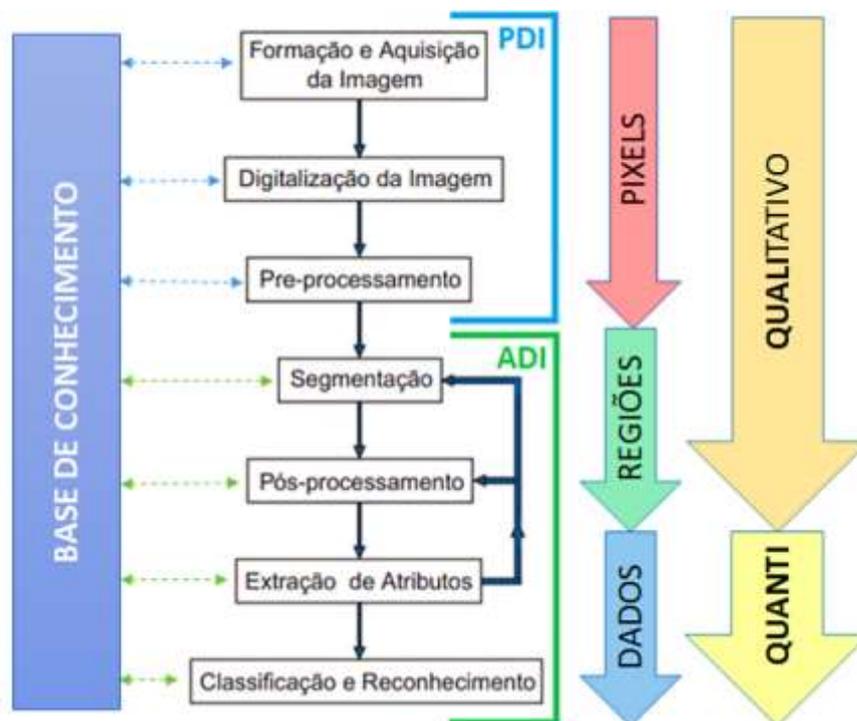
2.2 PROCESSAMENTO E ANÁLISE DIGITAL DE IMAGENS

O Processamento e Análise Digital de Imagens (PADI) é uma parte específica do Processamento Digital de Sinais (PDS), que lida com a obtenção, remodelagem e interpretação dos sinais visuais e da informação que eles contêm (Ignacio, 2013), e pode ser separada em

duas etapas, sendo elas o processamento e a análise digital de imagens, conforme demonstrado na Figura 1.

A área de Processamento Digital de Imagens (PDI) trata da manipulação ou tratamento da imagem adquirida para melhor operacionalização nas etapas seguintes (Queiroz & Gomes, 2001), podendo ser dividida em 3 principais etapas: formação e aquisição da imagem, digitalização da imagem e pré-processamento, que respectivamente lidam com a definição de parâmetros e captura da imagem, a execução de sua correspondência numérica para permitir a sua manipulação, o refinamento e tratamento da imagem para as próximas etapas.

Figura 1: Sequência do padrão de processamento e análise digital de imagens (PADI).



Fonte: Ignacio (p.25, 2013).

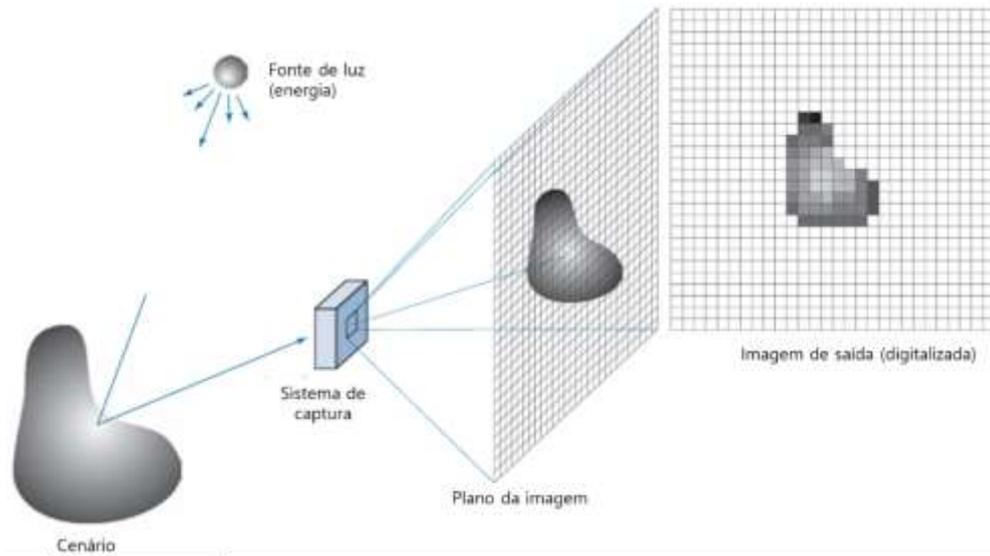
Após a realização da etapa de PDI, realiza-se a Análise Digital de Imagens (ADI), que recebe a imagem tratada e operacionaliza medições e o reconhecimento da imagem de acordo com a aplicação em que está situada, dividindo-se em 4 principais etapas: segmentação, pós-processamento, extração de atributos e o reconhecimento de padrões e classificação.

2.2.1 Aquisição da imagem

O processamento digital de amostras é iniciado com a aquisição da imagem, que depende necessariamente de um equipamento físico sensível a uma faixa de energia refletida pelo cenário a ser capturado (Filho & Neto, 1999). Então, os sinais analógicos obtidos são

codificados em sinais digitais, representados por uma matriz binária, e armazenados por meio do sistema de digitalização que será descrito no tópico 2.2.2. Um resultado do processo de aquisição das imagens é a transformação de um cenário tridimensional para um sistema dimensional.

Figura 2: Representação gráfica de um sistema de captura.



Fonte: Gonzales & Woods, 2018.

O sistema de aquisição de imagens mais utilizado é a câmera *Charged Couple Device* (CCD), composto por uma matriz de células semicondutoras fotossensíveis, que captam a energia refletida pela fonte de luz e produzem um Sinal Composto de Vídeo (SCV), analógico e monocromático (Filho & Neto, 1999).

2.2.2 Digitalização da imagem

No processo de transformação do cenário real para uma imagem digital, ocorre o processo de digitalização, em que uma imagem tridimensional do mundo real é remodelada para a forma de imagem digital. Ao longo deste procedimento, ocorre um processo de perda de dados por conta da discretização do sinal analógico captado pelo receptor para o sistema binário, utilizado nos equipamentos digitais. O resultado desta quantização é uma matriz $M \times N$, onde cada valor corresponde a uma intensidade de no nível de cinza estabelecido entre um conjunto de dados estabelecido.

$$f(x, y) = \begin{bmatrix} f(0,0) & \cdots & f(0, N - 1) \\ \vdots & \ddots & \vdots \\ f(M - 1, 0) & \cdots & f(M - 1, N - 1) \end{bmatrix} \quad (1)$$

A discretização do sinal analógico é realizada por duas formas, via amostragem e via quantização, onde a amostragem realiza a discretização das coordenadas espaciais, ou seja, da densidade de unidade de dados, também chamada posteriormente de *pixels*, e realiza também a discretização dos valores de intensidade de brilho, relacionados à sensibilidade do sistema de captura em relação às diferentes faixas de frequência refletidas pelo cenário (Filho & Neto, 1999).

2.2.3 Pré-processamento da imagem

Considerando que, as etapas anteriores de PADI não sejam perfeitas na captação e conversão dos sinais, pressupõe-se que falhas como ruído, distorção de contraste e brilho e perda de sinais possam acontecer. Portanto, aplica-se um pré-processamento na imagem, de forma a melhorar a sua qualidade para os passos seguintes, tendo em vista a aplicação em que este sistema está inserido.

Para o desenvolvimento deste projeto, realizou-se a conversão da imagem captada em cores para preto e branco, tendo em vista que o projeto realiza uma verificação das características, formas e contraste geral em comparação com diferentes regiões da imagem, também se ajustou o tamanho da imagem de forma a eliminar o cenário e manter somente o rosto identificado para as etapas seguintes.

Até a etapa de processamento, tendo em vista que utilizam o *pixel* como unidade de processamento e manipulação, são consideradas como operações de baixo nível (Filho & Neto, 1999). Até esta etapa, a interface digital não identifica regiões como olhos, plano de fundo e óculos, somente identifica os níveis de cinza resultantes do processo de discretização. A partir da próxima etapa, podemos separar regiões para a análise, iniciando-se a Análise Digital de Imagens.

2.2.4 Segmentação em regiões

Quando observamos uma imagem digital através dos nossos próprios olhos, normalmente realizamos uma rápida análise das características que compõe a imagem observada. Entretanto, nota-se que as características que podem se apresentar são infinitas; ao analisar um retrato de um rosto por exemplo, podemos notar a forma dos olhos, nariz, boca e mandíbula do rosto, também podemos analisar o caimento do cabelo e sobrancelha, bem como

as expressões que o rosto demonstra, entre muitas outras observações. Uma maneira de aprimorar a análise do sistema via PADI, é implementada via segmentação da imagem em regiões ou partes de interesse presentes na imagem, de forma a realizar uma análise de cada região separadamente, e eliminando as regiões que não serão utilizadas.

Na realização de segmentação de imagens digitais monocromáticas, utiliza-se de dois arranjos, o arranjo de descontinuidade é caracterizado pela verificação de mudanças bruscas nos níveis de cinza, enquanto o arranjo de similaridade é caracterizado pelo agrupamento de regiões com níveis de cinza equivalentes entre si (Filho & Neto, 1999).

2.2.5 Pós-processamento

A etapa de pós-processamento é dividida em basicamente dois métodos, sendo eles efetuados por meio de operações lógicas ou morfológicas (Filho & Neto, 1999). As operações lógicas são constituídas por três funções base, por meio de interseção (E), de união (OU) ou por meio de complemento (NÃO), aonde a combinação destas funções lógicas pode produzir os mais diversos resultados. As operações morfológicas atuam de maneira diferente, varrendo a imagem com uma matriz de números binários, também chamada de máscara, que pode produzir dois efeitos, o efeito de erosão e o de dilatação, quem também podem ser combinadas para produzir diferentes efeitos.

2.2.6 Extração de atributos

A extração de atributos busca discriminar a imagem em regiões de interesse, como área dos olhos, nariz, entre outros. Desta forma, ao invés de ser necessário varrer toda a região do rosto para a realização da rotulação, busca-se por atributos específicos de grande importância presente na imagem, de forma a evitar uma maior necessidade de processamento e com pouco ganho efetivo na identificação.

Para isto, diferentes atributos podem ser utilizados tanto individualmente como em conjunto, sendo estes definidos através de aprendizagem de máquina, que será descrito posteriormente.

2.2.7 Classificação e reconhecimento

Em posse somente dos atributos de maior importância dentro da imagem, é realizado um cruzamento de atributos com as previamente imputadas no sistema. Esse cruzamento de características forma em si, uma rede de conhecimento que estabelece conexões entre características relacionadas, dispensando a necessidade de averiguação de todos os atributos para a classificação da imagem.

2.3 APLICAÇÃO EM AUTENTICAÇÃO BIOMÉTRICA POR RECONHECIMENTO FACIAL

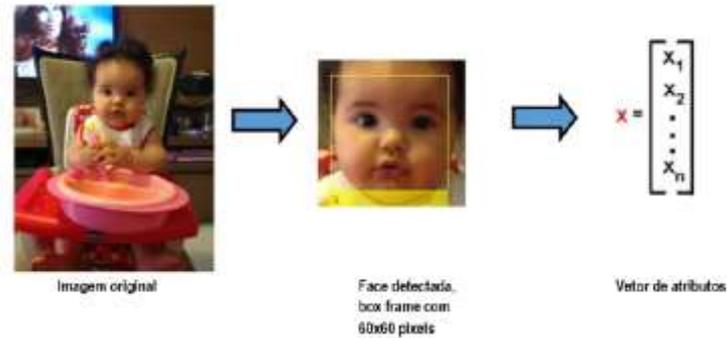
No contexto de controle de acesso, existem três principais conceitos na validação do usuário: a identificação, a autenticação ou verificação e a autorização (Magalhães & Santos, 2003). A identificação é o ato de identificação do usuário, por exemplo ao apresentar a sua carteira de identidade, carteira de trabalho, documento com foto, entre outros para a realização de abertura de conta bancária. Autenticação é o processo de cruzamento de informações, que é realizado por exemplo ao verificar se a carteira de identidade apresentada é válida, se o portador do documento e o documento dizem respeito a mesma pessoa, entre outros. A autorização diz respeito ao perfil de acesso às informações, por exemplo ao criar a conta bancária, o usuário poderá fazer transações bancárias livremente na sua conta, porém não terá acesso para movimentar o capital de outras pessoas.

Existem diversos métodos para a realização do controle de acessos, sendo o mais tradicional via compartilhamento das chaves. Entretanto, este método possui alguns problemas, como a possibilidade de transmissibilidade, seja voluntária ou não, a terceiros, a necessidade de arquivar ou a sua memorização pelos usuários, a possibilidade de uso de chaves de segurança fracas, como sequência de números, padrões conhecidos, o próprio nome, entre outros (Magalhães & Santos, 2003).

Uma alternativa ao método de compartilhamento de chaves de segurança é a biometria, que se refere a utilização dos traços físicos e/ou comportamentais de cada indivíduo, que são intransponíveis, para utilizá-los como método de autenticação e/ou identificação (BATISTA, *et al.*, 2017). Neste contexto, vários métodos biométricos podem ser utilizados, como impressões digitais, reconhecimento de voz, padrões de escrita, caligrafia, forma de locomover-se, reconhecimento facial, entre outros.

Na Figura 3 é ilustrado o exemplo de um sistema de extração de atributos e características da imagem visando o reconhecimento facial. Basicamente, quando a imagem for capturada, os atributos e as características são extraídos e comparados com a base dados previamente armazenada para tomada de decisão, isto é, reconhecimento facial.

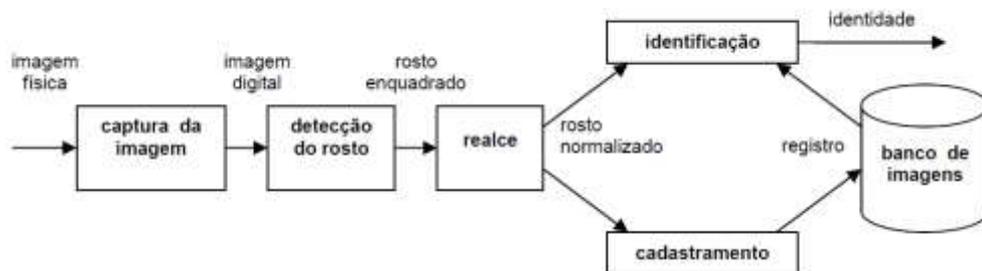
Figura 3: Extração de atributos e características das imagens.



Fonte: Maia (p.15, 2016).

Um dos sistemas biométricos de controle de acesso é o *Automatic Face Recognition* (AFR) ou sistema automático de reconhecimento, que identifica as características faciais do usuário e as autentica com base em um banco de dados. Este método vem sendo amplamente utilizado por conta de sua praticidade em campo, visto que não existe um contato direto do usuário com o sensor como em sistemas baseados em impressões digitais, são discretos e baratos visto que a presença de câmeras já existe nos mais diversos dispositivos, o registro de faces para a base de dados pode ser reduzida e ainda sim oferecer um rendimento satisfatório, entre outros.

Figura 4: Subdivisão do sistema automático de reconhecimento em blocos.



Fonte: Feitosa e Pereira (p.2, 2003).

Para o desenvolvimento do projeto, serão utilizados os moldes da Figura 4, em que será realizada a identificação do usuário, por meio da plataforma MatLab, que realizará contato indireto para a captura da imagem. O algoritmo irá realizar a detecção do rosto e o realce das características para a sua identificação, em seguida, será verificando-se o rosto identificado se encontra na base de dados, caso sim, será realizada a autenticação do usuário e autorização das

permissões para acesso aos dados conforme o registro, e caso não esteja na base de dados, será disponibilizado a realização do cadastro de usuário, que será aprovada pelo administrador da ferramenta.

2.3.1 O algoritmo de Viola-Jones

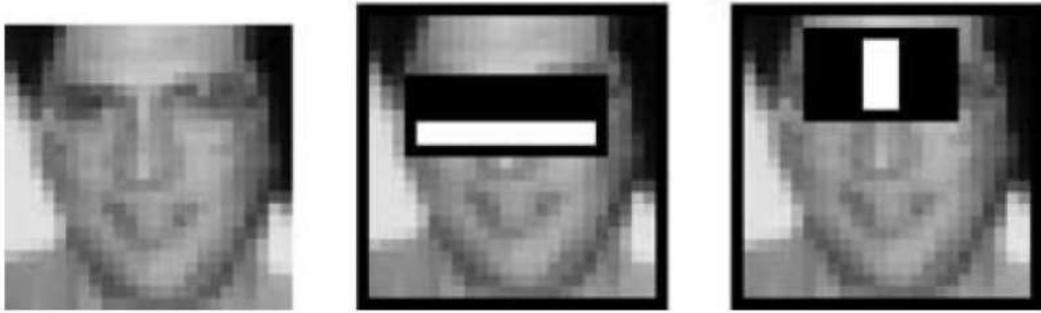
Existem diferentes métodos para a detecção e validação de faces, os baseados em conhecimentos, os baseados em características invariantes, os baseados em modelos, e os baseados na aparência (Costa, 2018). Os métodos baseados em conhecimento fundamentam-se na definição de atributos, que normalmente são geométricos, e que definem o complexo do rosto; os baseados em características invariantes fundamentam-se principalmente na segmentação da pele, o que facilita a detecção do rosto quando em diferentes posicionamentos, como em perfil e em diferentes inclinações; os métodos baseados em modelos fundamentam-se após uma descrição genérica do que é um rosto, utilizando-se de contornos para a definição do rosto; e os baseados em aparência, que utilizam de dados pré-estabelecidos para identificação da face.

O modelo baseado em aparência, em específico, utiliza técnicas de aprendizado de máquina para ajustar a identificação do objeto na imagem, o que o torna esta técnica um diferencial para evitar o ajuste manual do reconhecimento, e por conta disso, este foi o método utilizado neste trabalho. Um algoritmo desenvolvido por Paul Viola e Michael Jones em 2001, que se baseia em 3 principais aspectos, utiliza desta técnica, onde se baseia em um filtro de características, e um classificador simples para ordenar a relevância das características identificadas, podendo-se aplicar diferentes combinações de classificadores para ajuste de performance (Fonseca, 2016).

2.3.1.1 Características de Haar e integral de imagem

O algoritmo de Viola-Jones utiliza de identificações de características, baseando-se no valor dos *pixels* resultante de sucessivas aplicações da transformada *wavelet* com a base de Haar. Desta forma, decompõe imagens complexas em resultantes simples, possibilitando por exemplo, a informação que a região dos olhos e extremidades do rosto são bem mais escuras que a do nariz e das maçãs do rosto, de forma a interpretar não somente a posição do rosto, mas também as características individuais de cada rosto (Costa, 2018).

Figura 5: Aplicação das características de Haar em um rosto.



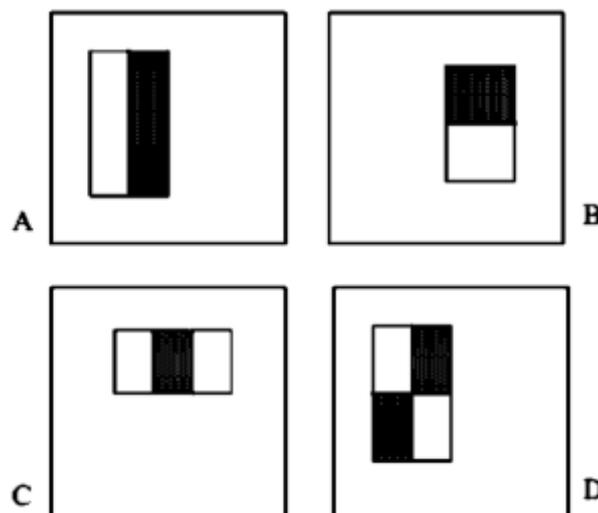
Fonte: (Fonseca, 2016)

De forma a obter as características de imagem, utiliza-se a integral de imagem, também conhecida como tabela de soma de áreas, que representa uma região da imagem com base na soma de *pixels* da região, representado pela equação abaixo, aonde $ii(x, y)$ é a imagem integral, $i(x', y')$ é a imagem original e (x, y) são as coordenadas do *pixel* (Costa, 2018).

$$ii(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y') \quad (2)$$

O resultado da integral de imagem em uma região então, resulta em uma das quatro características de Haar, demonstradas na Figura 6, sendo cada característica de forma retangular, e resultante da soma dos *pixels* da região, que podem ser separados em 3 características, sendo elas a de dois retângulos (a) e (b), de três retângulos (c) e de quatro retângulos (d).

Figura 6: Características retangulares de Haar em imagens integrais.



Fonte: (Costa, 2018)

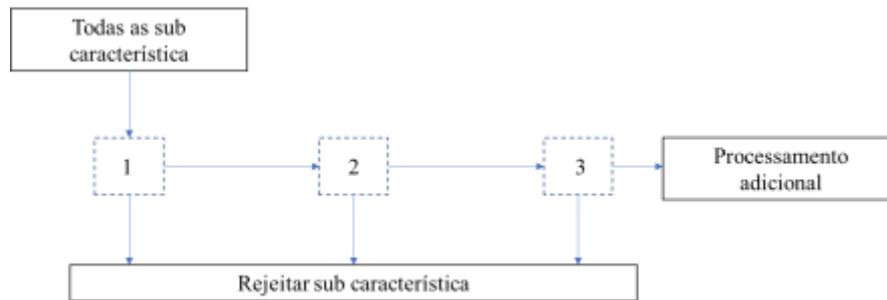
2.3.1.2 Classificador e cascata de características

Outra abordagem utilizada pelo algoritmo de Viola-Jones, utiliza o algoritmo de *Adaptive Boosting* (AdaBoost), que estrutura um classificador eficiente com base em vários classificadores menos eficientes. Este algoritmo pode ser demonstrado pela equação abaixo, onde a função $h_t(x)$ representa os classificadores menos eficientes.

$$f(x) = \sum_{t=1}^T a_t h_t(x) \quad (3)$$

Tendo em vista que o algoritmo de Viola-Jones utiliza poucas, mas fortes características, a execução deste algoritmo torna-se mais rápida que os demais, onde evita-se o travamento em identificação de características muito detalhadas. Desta forma, combinando-se múltiplos classificadores fortes, obtêm-se um algoritmo que permite uma otimização de tempo para a sua conclusão.

Figura 7: Cascata de classificadores de Viola-Jones.



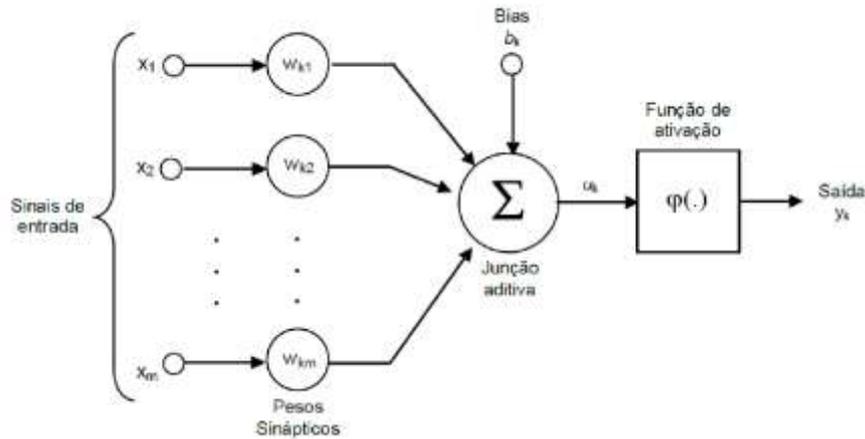
Fonte: (Fonseca, 2016)

2.3.2 Rede neural artificial

A Rede Neural Artificial (RNA) é um método baseado em aparências, que conforme descrito no item 2.3.1, utiliza de dados pré-estabelecidos para compreender as características que devem ser identificadas para classificar uma imagem, e o funcionamento de uma RNA pode ser comparada analogamente a uma rede de neurônios conectados, que percorrem diferentes níveis de camadas para realizar o processamento do conhecimento (Costa, 2018).

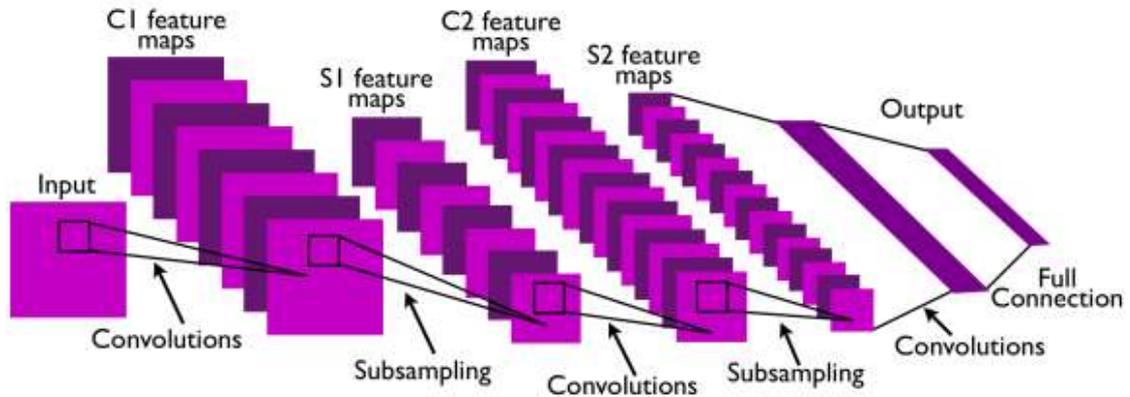
Assim como o funcionamento de um neurônio real, o neurônio artificial recebe muitas entradas de dados (x_1, x_2, \dots, x_m) , onde cada conexão possui um peso em relevância $(w_{k1}, w_{k2}, \dots, w_{km})$ que é integralizada e enviada a saída (y_k) de acordo com uma função de ativação $(\varphi())$, esta ideia está demonstrada na Figura 8 (Fonseca, 2016).

Figura 8: Modelo de neurônio artificial.



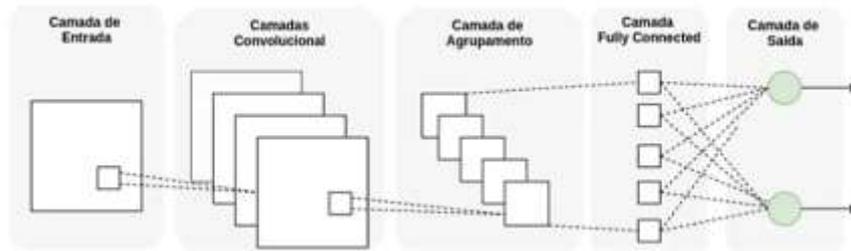
A classificação de uma RNA é baseada na quantidade e arranjo das camadas presentes na rede neural, onde a RNA utilizada para este trabalho é a rede neural convolucional, do inglês *Convolutional Neural Network* (CNN), que utiliza métodos de aprendizagem profunda para aperfeiçoar a identificação (LeCun, Kavukcuoglu, & Farabet, 2010).

Figura 9: Uma arquitetura de CNN de dois estágios.



Em CNNs, cada *pixel* da imagem corresponde a uma informação que será entregue a um neurônio, ou seja, um *pixel* por neurônio. Entretanto, o mesmo não acontece para a camada seguinte, onde primeiramente, ocorre um agrupamento de uma região de neurônios, que então é recebido por neurônios da camada seguinte, utilizando dos mesmos pesos, também chamados de filtros, ou *kernels* (Chaves, 2018).

Figura 10: Arquitetura geral de uma CNN.



Fonte: (Rede neural convolucional, 2022)

A partir deste momento, denominamos esta etapa de camada de convolução, cujo resultado do neurônio oculto é composto pela convolução do filtro $w_{l,m}$, uma matriz de tamanho $t \times r$, pela imagem $a_{j+l,k+m}$ na posição $a_{x,y}$, acrescentado do peso de classificação, ou viés b , e ativados pela função de ativação σ (Chaves, 2018). Este processo é repetido em toda a imagem, com o deslizamento da janela de filtro e agrupamento de *pixels*, até percorrer todos os *pixels* da imagem, e o resultado desta convolução é chamado de mapa de recursos, ou *feature maps*.

$$\sigma \left(b + \sum_{l=0}^t \sum_{m=0}^r w_{l,m} a_{j+l,k+m} \right) \quad (4)$$

Supondo que o mapa de recursos seja baseado numa imagem de entrada de tamanho $(H \times H)$, e de um agrupamento de *pixels* de tamanho $(F \times F)$, com um deslocamento de passo S , define-se o tamanho do mapa de recursos (Chaves, 2018).

$$W = \frac{H - F}{S} + 1 \quad (5)$$

A próxima camada é denominada de camada de *polling*, ou *down-sampling*, é resultado de uma sub amostragem da camada anterior, de forma a reduzir a resolução dos mapas de recursos. Esta camada realiza uma substituição estatística dos valores dos *pixels* de uma região, de forma a diminuir a quantidade de processamento necessário para varrer a imagem. O método de *polling* utilizado para este trabalho foi o *max-polling*, que substitui os *pixels* de uma região pelo valor correspondente ao maior *pixel*. Invariavelmente, o processo de *polling* produz uma distorção com o objetivo de forma a aumentar a velocidade no processamento da informação (Chaves, 2018).

Após a camada de *polling*, ocorre o processo de conversão dos dados para um vetor unidimensional, que é processo da camada de reformatação, também chamada de *flattening*. Este processo é realizado através da aplicação de uma função não linear em cada elemento do mapa de características, onde um exemplo de unidade linear retificada, ou do inglês *Rectified*

Linear Unit (ReLU), onde valores de entrada negativos são convertidos em 0, eliminando a ativação de neurônios artificiais não necessários. (Juraszek, 2014).

$$ReLU(z) = \max(0, z) \quad (6)$$

Então, o vetor unidimensional é estruturado de forma a realizar conexões de parâmetros relacionados, onde este processo é resultado da camada densa, ou do inglês *fully connected*. Nesta camada, utiliza-se deste vetor unidimensional para realizar o treinamento de um classificador totalmente conectado (Juraszek, 2014).

$$Softmax(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (7)$$

3 MATERIAIS E MÉTODOS

A seguir são abordados os aspectos metodológicos da pesquisa realizada, escrevendo-se os procedimentos necessários para desenvolver um sistema de reconhecimento facial para autenticação de usuário e desbloqueio em estações de trabalho com computador e com uma *webcam* visando o aumento da segurança ao acesso de informações por usuários cadastrados na base de dados. Esse estudo teve por finalidade realizar uma pesquisa de natureza aplicada. Para alcançar os objetivos propostos e melhor apreciação deste trabalho, foi utilizada uma abordagem quali-quantitativa. Com intuito de conhecer a problemática sobre a área de estudo foi realizada uma pesquisa exploratória e bibliográfica (GIL, 2019). A seguir são descritos os materiais utilizados ao longo da pesquisa, bem como os procedimentos metodológicos.

3.1 AMBIENTE DE DESENVOLVIMENTO

O *hardware* utilizado nesta pesquisa possui a seguinte configuração: Processador Intel® Core™ i7 – 7500U CPU 2.70GHz 2.90GHz ; Placa de vídeo Nvidia® GeForce® 940MX 2GB; RAM instalada 8,00GB (utilizável 7,88GB); Sistema operacional de 64 *bits*, edição Windows 10 *Home Single Language* versão 22H2.

3.2 FERRAMENTAS DE DESENVOLVIMENTO

O ambiente de programação utilizado para a elaboração do projeto foi o MatLab, um *software* desenvolvido pela empresa MathWorks®, caracterizado pela sua própria linguagem de programação, bibliotecas, *toolboxes*, e ambiente de desenvolvimento integrado, ou do inglês *Integrated Development and Learning Environment* (IDLE), permitindo uma maior visualização de dados e capacidade de simulação de cenários complexos.

A ferramenta utilizada para aplicar as CNNs nas imagens adquiridas é a AlexNet, uma rede neural densa pré-treinada, que utiliza três camadas de convolução e três camadas densas da sua arquitetura, conforme mostrado na Figura 11 (AlexNet, 2023). Esta rede neural foi desenvolvida como tese de doutorado por Alex Krizhevsky, Ilya Sutskever, e Geoffrey Hinton, ganhando em 2012 uma competição de reconhecimento de imagens chamada de *ImageNet large scale visual recognition challenge*.

Figura 11: Arquitetura da rede neural AlexNet.

| net.layers | | |
|-------------------------------|----------|---|
| ans = | | |
| 25x1 Layer array with layers: | | |
| 1 | 'data' | Image Input: 227x227x3 images with 'zerocenter' normalization |
| 2 | 'conv1' | Convolution: 96 11x11x3 convolutions with stride [4 4] and padding [0 0 0 0] |
| 3 | 'relu1' | ReLU |
| 4 | 'norm1' | Cross Channel Normalization: cross channel normalization with 5 channels per element |
| 5 | 'pool1' | Max Pooling: 3x3 max pooling with stride [2 2] and padding [0 0 0 0] |
| 6 | 'conv2' | Grouped Convolution: 2 groups of 128 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2] |
| 7 | 'relu2' | ReLU |
| 8 | 'norm2' | Cross Channel Normalization: cross channel normalization with 5 channels per element |
| 9 | 'pool2' | Max Pooling: 3x3 max pooling with stride [2 2] and padding [0 0 0 0] |
| 10 | 'conv3' | Convolution: 384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1] |
| 11 | 'relu3' | ReLU |
| 12 | 'conv4' | Grouped Convolution: 2 groups of 192 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1] |
| 13 | 'relu4' | ReLU |
| 14 | 'conv5' | Grouped Convolution: 2 groups of 128 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1] |
| 15 | 'relu5' | ReLU |
| 16 | 'pool5' | Max Pooling: 3x3 max pooling with stride [2 2] and padding [0 0 0 0] |
| 17 | 'fc6' | Fully Connected: 4096 fully connected layer |
| 18 | 'relu6' | ReLU |
| 19 | 'drop6' | Dropout: 50% dropout |
| 20 | 'fc7' | Fully Connected: 4096 fully connected layer |
| 21 | 'relu7' | ReLU |
| 22 | 'drop7' | Dropout: 50% dropout |
| 23 | 'fc8' | Fully Connected: 1000 fully connected layer |
| 24 | 'prob' | Softmax |
| 25 | 'output' | Classification Output: crossentropyx with 'tench' and 999 other classes |

Fonte: (AlexNet, 2023)

AlexNet utiliza no seu processamento um total de 25 camadas, aonde a imagem analisada é inserida na primeira camada, em seguida são realizadas duas execuções das camadas de convolução, ReLU, normalização e *pooling*; então são realizadas duas execuções das camadas de convolução e ReLU ; depois é realizada uma execução das camadas de convolução ReLU e *pooling*; posteriormente, são executadas duas execuções das camadas densa, ReLU e *dropout*; em seguida, é realizada uma execução da camada densa, *SoftMax* e então de classificação.

3.3 MÉTODOS

De forma a elucidar a metodologia aplicada para o desenvolvimento deste projeto, foi realizada a decomposição da metodologia em etapas, sendo elas a de preparação do conjunto de dados, definição da arquitetura, implementação da arquitetura, avaliação da arquitetura, e análise dos resultados; desta forma, procura-se justificar as resoluções adotadas em cada etapa do projeto, a fim de proporcionar clareza e transparência no processo. A partir da decomposição metodológica, pretende-se realizar uma abordagem sistemática para as etapas seguintes de implementação e resultados. A metodologia aplicada neste projeto se encontra exemplificada no diagrama de blocos da Figura 12.

Figura 12: Etapas da metodologia desenvolvida no projeto.



Fonte: Autoria própria.

3.3.1 Preparação do conjunto de dados

De forma a evitar enviesamento na coleta de imagens de pessoas de similar etnia, gênero ou localização, buscou-se um banco de dados público, (Chandra, 2019), sendo o banco dividido em 4 partes, composto por 14 fotos de cada pessoa, de um total de 50 pessoas diferentes em distintos posicionamentos de rosto e condições de iluminação, totalizando 700 fotos em cada parte. Para a utilização do sistema foram utilizados somente um quarto da quantidade total do banco, ou seja, 700 fotos.

Figura 13: Banco de dados fotográfico de faces.

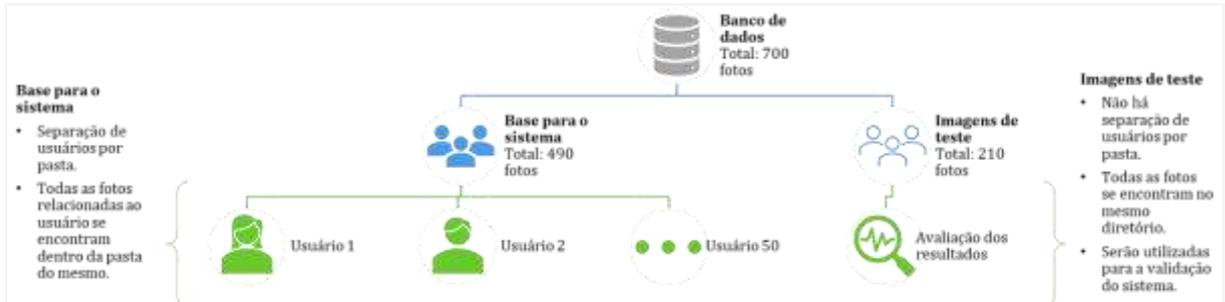


Fonte: Autoria própria.

Após a aquisição do banco de dados, dividiu-se aleatoriamente o diretório de imagens de acordo com o seu propósito, resultando em dois grandes grupos, as imagens base utilizadas para o treinamento do sistema e sendo imagens de validação para avaliação da performance do sistema projetado, em uma proporção de 70% para a base do sistema e 30% para o teste do sistema. A seleção aleatória adotada para divisão do banco de dados possui grande relevância principalmente por conta da nomenclatura adotada originalmente no banco de dados, que possui uma estrutura fixa da numeração associada ao usuário, seguida da característica da posição do rosto e iluminação adotados, conforme demonstrado na Figura 13.

Posteriormente, segmentou-se as imagens base do sistema em diretórios individuais correspondentes a cada usuário, que servirão para informar as classificações, no caso os usuários, para treinamento da rede neural convolucional utilizada. Por último, agrupou-se as imagens de teste do sistema em um mesmo diretório, não realizando distinção de usuários por pastas da mesma maneira que realizado com as imagens base para treinamento. Esta divisão das imagens do banco de dados se encontra ilustrada na Figura 14.

Figura 14: Divisão do banco de dados entre imagens de teste e base para o sistema.

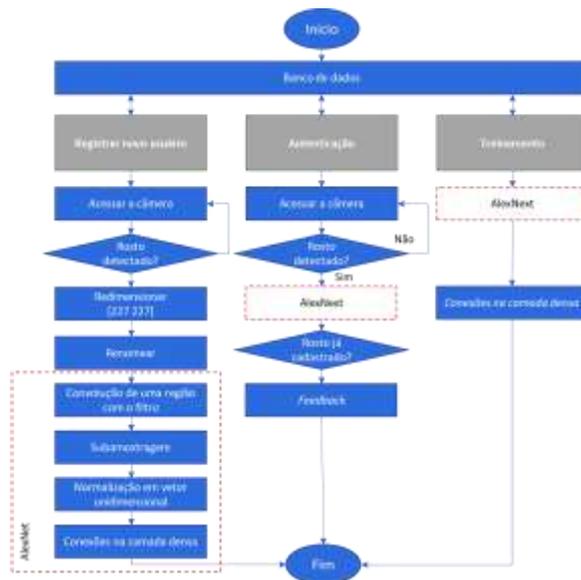


Fonte: Autoria própria.

3.3.2 Definição da arquitetura

Na etapa de definição de arquitetura, foi realizada a seleção da base lógica que será adotada para a implementação do projeto na prática onde o projeto foi diferenciado em três principais funcionalidades, a de registro de novos usuários utilizando a câmera embutida no *notebook* de trabalho, a autenticação de usuários e o treinamento da rede neural convolucional.

Figura 15: Arquitetura do projeto desenvolvido.



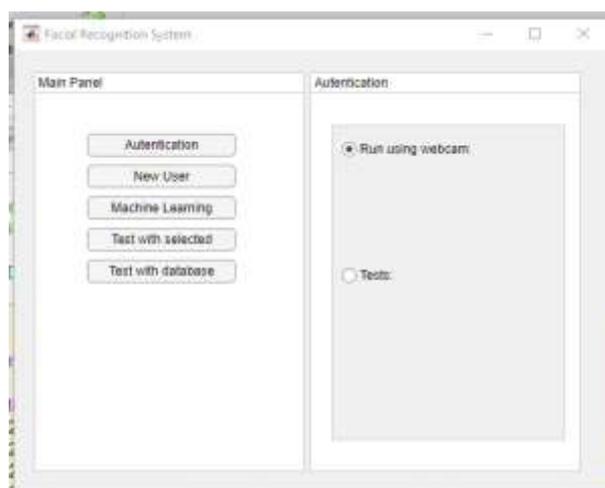
Fonte: Autoria própria.

Inicialmente, foram consideradas apenas as funcionalidades de treinamento e autenticação, entretanto o sistema ficaria retido aos usuários presentes no banco, impossibilitado de atualização para o cadastro de novos usuários ou utilização em cenários de diferentes posições do rosto e iluminação dos cenários capturados pelo banco de dados original, portanto, posteriormente definiu-se o sistema como o demonstrado na Figura 15. A partir das subdivisões principais, considerou-se a necessidade do acesso ao *hardware* disponível, como a *webcam*, à manipulação dos arquivos de imagem, e a incorporação da CNN.

3.3.3 Implementação da arquitetura

Nesta etapa, foi realizada a elaboração da aplicação no MatLab, descrito no tópico 3.2, e ambiente de trabalho descrito no tópico 3.1, e para a implementação deste projeto, utilizou-se as diretrizes estabelecidas no tópico 3.3.3.

Figura 16: Interface do sistema com o usuário.



Fonte: Autoria própria.

Primeiramente foi estabelecido o acesso da ferramenta com os diretórios e acesso a *webcam* presente no ambiente de desenvolvimento, para isso, foram utilizados os aplicativos de *image processing and computer vision* e *machine learning and deep learning*, presentes gratuitamente para a ferramenta MatLab. Após a elaboração do funcionamento do sistema, também se otimizou a interação com o sistema por meio de uma interface gráfica, atribuindo mais receptividade e acesso para o usuário final, levando em consideração que ele pode não estar familiarizado com as ferramentas de desenvolvimento para informar seu comando para a ferramenta. A interface desenvolvida para a comunicação com o usuário se encontra representada na Figura 16.

3.3.4 Avaliação da arquitetura

O sistema de reconhecimento facial proposto atua, de maneira simplificada, como um sistema de classificação. Para avaliar a eficiência do sistema desenvolvido, utiliza-se de métricas de avaliação para identificar a acurácia que o sistema pode apresentar e o grau de erros em relação ao valor esperado, para comparação com outros sistemas e avaliação de performance.

3.3.4.1 Matriz de confusão

Uma matriz de confusão é disposta em forma de matriz bidimensional, onde seus eixos são compostos pelo resultado esperado em contrapartida ao resultado obtido, e os valores de cada posição da matriz compostos por classes de dados preditivos, sendo eles verdadeiro/falso, positivo/negativo, entre outros. (Matos, Ciferri, & Vieira, 2009). Com base nos dados referentes a matriz de confusão, obtém-se informações chave para a avaliação do sistema, conforme descritos abaixo.

- a) Verdadeiros positivos: a informação avaliada possui o valor lógico verdadeiro/positivo, e o sistema trouxe o resultado esperado.
- b) Verdadeiros negativos: a informação avaliada possui valor lógico falso/negativo, e o sistema trouxe o resultado esperado.
- c) Falsos positivos (erro tipo 1): a informação avaliada possui valor lógico falso/negativo, mas o sistema a classificou como verdadeiro/positivo.
- d) Falsos negativos (erro tipo 2): a informação avaliada possui valor lógico verdadeiro/positivo, mas o sistema a classificou como falso/negativo.

Utilizando as informações obtidas pela matriz de confusão, dispõe-se ao projetista e ao consumidor mais ferramentas para analisar se o projeto atende aos propósitos mínimos dos quais foi designado e para atribuir confiabilidade aos resultados que o sistema informa. De posse dos elementos da matriz de confusão, pode-se obter o total de predições, o total de valores reais, o total de acertos e o total de erros, que depois servirão para o cálculo de métricas que podem informar a qualidade do sistema, conforme estabelecido nas Equações 8 a 11.

$$real_{positivo} = VP + FN \quad (8)$$

$$real_{negativo} = VN + FP \quad (9)$$

$$acertos = VP + VN \quad (10)$$

$$erros = FP + FN \quad (11)$$

3.3.4.2 Métricas de avaliação de qualidade do classificador

Ao realizar a análise das redes neurais convolucionais, algumas métricas estabelecem um panorama geral sobre o desempenho dos classificadores para a identificação e atribuição de rostos de diferentes usuários.

- a) Acurácia (ACC): A acurácia, do inglês *accuracy*, indica o percentual de acertos em relação à quantidade total de imagens. Esta é uma das métricas mais básicas para obter-se um panorama geral da capacidade de classificação correta do sistema;
- b) Precisão: A precisão informa a proporção de predições positivas corretas em relação a todas as predições positivas feitas pelo sistema, ou seja, ela indica a capacidade do modelo de não classificar erroneamente exemplos negativos como positivos;
- c) Sensibilidade: A sensibilidade, ou revocação, ou do inglês *recall*, também é chamada de taxa positiva verdadeira, do inglês *True Positive Rate* (TPR), e indica a capacidade de detecção de classificadores de resultado positivo;
- d) Especificidade: a especificidade, também chamada de taxa negativa verdadeira, ou *True Negative Rate* (TNR), é o inverso da sensibilidade, indicando a capacidade de detecção de classificadores de resultado negativo, e é relevante sobretudo em cenários aonde a detecção correta de cenários negativos é fundamental;
- e) Taxa de falsos positivos: A taxa de falsos positivos, do inglês *False Positive Rate* (FPR), avalia as detecções errôneas de exemplos negativos como positivos;
- f) Taxa de falsos negativos: A taxa de falsos negativos, do inglês *False Negative Rate* (FNR), avalia as detecções errôneas de exemplos positivos como negativos;
- g) *F-score*: A *F-score*, ou *F-measure*, é uma média harmônica calculada com base na precisão e na revocação, e é mais relevante em cenários em que as categorias ou classes estão desbalanceadas, seja em quantidade ou importância, pois realiza uma avaliação individual do desempenho do sistema em cada categoria de um banco de dados.

As métricas supracitadas auxiliam a avaliação da eficácia e do desempenho do sistema de autenticação desenvolvido neste projeto, aonde as métricas estão descritas nas equações 12 a 18.

$$ACC = \frac{acertos}{total} = \frac{VP + VN}{VP + VN + FP + FN} \quad (12)$$

$$precisão = \frac{VP}{VP + FP} \quad (13)$$

$$sensibilidade = TPR = \frac{VP}{real_{positivo}} = \frac{VP}{VP + FN} \quad (14)$$

$$especificidade = TNR = \frac{VN}{real_{negativo}} = \frac{VN}{VN + FP} \quad (15)$$

$$FPR = \frac{FP}{real_{negativo}} = \frac{FP}{FP + VN} = 1 - TNR \quad (16)$$

$$FNR = \frac{FN}{real_{positivo}} = \frac{FN}{FN + VP} = 1 - TPR \quad (17)$$

$$f1 = 2 \frac{precisão \cdot sensibilidade}{precisão + sensibilidade} \quad (18)$$

3.3.5 Análise dos resultados

Após o cálculo das métricas de avaliação descritas no tópico 3.3.4, uma análise quantitativa será realizada, considerando não somente os números adquiridos, mas também o contexto de autenticação facial em que o sistema está inserido. Por exemplo, resultados falso positivos possuem um impacto prejudicial muito maior para este projeto que resultados falso negativos, isto ocorre, pois, resultados falso positivos podem viabilizar o acesso aos dados a usuários que não possuem registro na base de dados registrada, propiciando uma falha de segurança, enquanto resultados falso negativos, que podem inviabilizar o acesso dos dados a um usuário com registro na base de dados, podem ser revertidos por uma atualização das fotos do usuário que serão utilizadas para o treinamento da CNN.

4 IMPLEMENTAÇÃO

4.1 ARQUITETURA DO SISTEMA

No desenvolvimento do projeto, separou-se o código em três principais funcionalidades, a de registro de novo usuário, a de autenticação de usuário e a de treinamento, sendo que todos realizam conexão com o banco de dados e a AlexNet. Por meio da AlexNet, aplica-se o método de redes neurais convolucionais para extrair as características e atributos das imagens obtidas. A camada densa por sua vez, utiliza os neurônios interconectados e as funções de ativação para efetuar a autenticação de rostos. Desta forma utiliza-se as informações pré-treinadas para realizar a identificação geral se existe um rosto ou não, permitindo a identificação de atributos distintos entre os usuários cadastrados.

A arquitetura do sistema consiste, primeiramente, no banco de dados que serve de apoio para as outras funcionalidades do sistema, então ao selecionar entre as funcionalidades elaboradas no sistema, registro de novo usuário, autenticação ou treinamento, uma diferente rota lógica é percorrida. Ao percorrer a rota de registro de usuário, as variáveis e a CNN são inicializadas e o sistema acessa a *webcam* do sistema, armazenando os arquivos com rostos identificados e adicionando-os ao banco de dados. Ao percorrer a rota de autenticação do usuário, as variáveis são inicializadas e o sistema realiza uma captura da imagem adquirida pela *webcam* do sistema, então a CNN realiza uma varredura para indicar se o rosto identificado pertence a um usuário cuja biometria facial consta no banco de dados, e informa o resultado da análise. Por último, ao percorrer a rota de treinamento da CNN, o diretório das imagens do banco de dados é acessado e as classificações de usuários são reforçados na camada densa da CNN, de forma a extrair rapidamente as principais características das imagens adquiridas.

4.1.1 Treinamento da rede neural convolucional

A rede convolucional pré-treinada AlexNet utiliza de um banco de dados pré-estabelecido para adquirir conhecimento de algumas classificações básicas, e pode ser treinado localmente, utilizando o banco de dados utilizado para este projeto, conforme demonstrado na Figura 17.

Figura 17: Código para treinamento da AlexNet.

```

net = alexnet;
analyzeNetwork(net);

layers = net.Layers;
layers(23) = fullyConnectedLayer(50);
layers(25) = classificationLayer;

allImages=imageDatastore('DataBase_resize_2','IncludeSubfolders',true,'LabelSource','foldernames');

opts = trainingOptions('sgdm','InitialLearnRate',0.001,'MaxEpochs',20,'MiniBatchSize',64);
myNet = trainNetwork(allImages,layers,opts);
save myNet;

```

Fonte: Autoria própria.

O treinamento ocorre primeiramente, carregando a CNN AlexNet, em seguida é disponibilizado para visualização as camadas da rede neural, com suas respectivas funcionalidades, através da função *analyze network*. Após a disponibilização anterior, é informado para a camada 23, ou seja, a camada densa, quantas categorias diferentes o sistema possui, onde o sistema possui 50 diferentes rostos para classificação.

Figura 18: Diretório dos arquivos de teste.



Fonte: Autoria própria.

Então é informado o caminho que contém os arquivos de imagem a serem testados, conforme comentado no tópico 3.3.1, indicando também para considerar as subpastas do caminho e considerar o nome das subpastas como rótulo para as diferentes classificações do rosto, ou seja, as fotos do rosto de cada pessoa se encontrarão em uma subpasta cujo rótulo pode ser o nome do usuário em questão, ou número único de referência para consultas e averiguação dos resultados.

4.1.2 Registro de novo usuário

Outra categoria de desenvolvimento utiliza a *webcam* presente no *notebook* para capturar imagens do usuário atual e agregá-las ao banco de dados. O código utilizado para realizar esta operação está demonstrado na Figura 19.

Figura 19: Código para registro de usuários.

```

vid = videoinput('winvideo'); % Access an image acquisition device
% preview(vid); % TEST - Show webcam capture
faceDetector = vision.CascadeObjectDetector; % Detect objects using the Viola-Jones algorithm

counting_limit=14;
counting=1;

while true
    img = getsnapshot(vid);
    bboxes = step(faceDetector, img); % Detect faces on screenshot

    if(sum(sum(bboxes))~=0) % Check if face is detected
        if(counting==counting_limit) % Stop when limit of images per person is reached
            return;
        else
            OnlyFace = imcrop(img, bboxes(1,:)); % Crop to maintain only the face area
            OnlyFace = imresize(OnlyFace, [277 277]);
            filename = strcat(num2str(counting), ' bmp');
            imwrite(OnlyFace, filename); % Save
            counting=counting+1

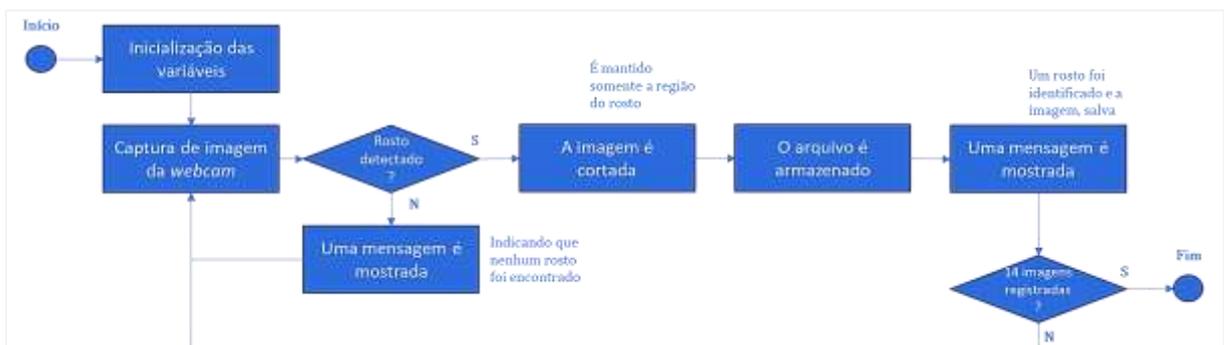
            imshow(OnlyFace);
            title('Face detected. Registering image ');
            drawnow;
        end
    else
        imshow(img);
        title('Face not detected, please change your position ');
        drawnow;
    end
end
end

```

Fonte: Autoria própria.

Primeiramente, são inicializadas as variáveis do sistema, e em seguida é inicializado um laço para capturar as imagens do rosto do novo usuário. Dentro deste laço, é realizada uma captura de imagem da *webcam* e então é inicializada a rede neural para identificar se a imagem adquirida possui um rosto, caso não seja identificada nenhum rosto, é realizado o processo de captura da imagem na *webcam* novamente, e caso um rosto seja identificado, a imagem é cortada de forma a manter-se somente a região identificada com rosto, e o arquivo é armazenado no sistema. Todo este processo dentro do loop ocorre até serem armazenadas 14 fotos do usuário a ser registrado. Todo o processo do algoritmo para registro de usuário se encontra em forma de fluxograma na Figura 20.

Figura 20: Fluxograma do processo de registro de novos usuários.



Fonte: Autoria própria.

4.1.3 Autenticação de usuário

De forma a realizar a autenticação do usuário, inicializa-se a CNN AlexNet, que recebe uma imagem capturada da *webcam*, utilizando o algoritmo de Viola-Jones para classificá-la de acordo com os rostos no banco de dados previamente estabelecido. De forma a visualizar o resultado, é realizada a disposição da figura capturada com o título correspondente à classificação atribuída pela CNN, caso nenhum rosto seja identificado na imagem obtida pela *webcam*, uma mensagem de retorno é informada ao usuário.

Figura 21: Código para autenticação de usuário.

```

vid = videoinput('winvideo');
load myNet;
faceDetector = vision.CascadeObjectDetector;

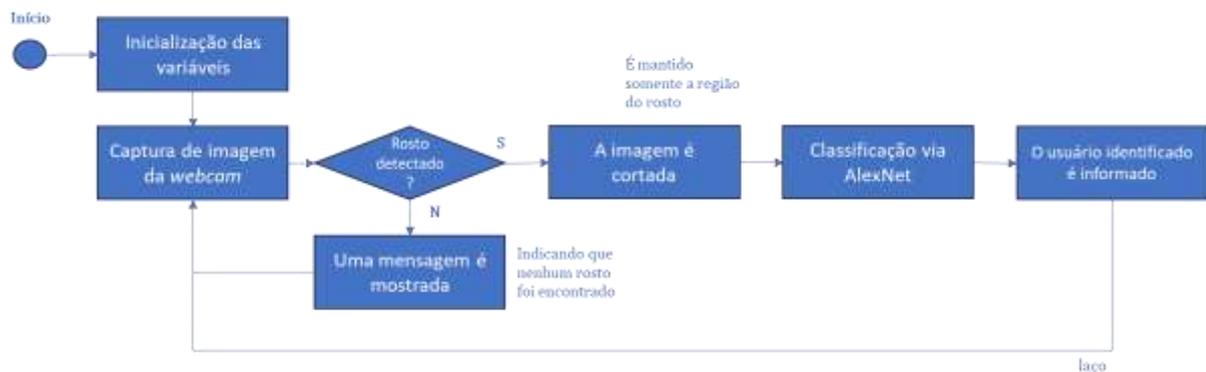
while true
    img = getsnapshot(vid);
    bboxes = step(faceDetector, img);
    if (sum(sum(bboxes)) ~= 0)
        es = imcrop(img, bboxes(1, :));
        es = imresize(es, [227 227]);
        label = classify(myNet, es);
        image(img);
        title(char(label));
        drawnow;
    else
        image(img);
        title('No Face Detected');
    end
end

```

Fonte: Autoria própria.

Tendo em vista que o usuário pode tentar realizar a validação não somente uma vez, mas várias vezes, criou-se um laço que continua capturando e classificando o rosto com base nos usuários presentes no banco de dados. Esta estratégia para o laço se encontra disposta em forma de fluxograma na Figura 22.

Figura 22: Fluxograma do processo de autenticação de usuário.



Fonte: Autoria própria.

5 RESULTADOS E DISCUSSÕES

Com o intuito de investigar de maneira abrangente e elucidativa, este capítulo explora os resultados obtidos a partir da análise tanto qualitativa quanto quantitativa realizada. Serão apresentados os desdobramentos das investigações conduzidas, oferecendo uma visão detalhada e abrangente dos dados coletados e processados.

5.1 AVALIAÇÃO QUALITATIVA

Utilizando as fotos separadas para teste da ferramenta, conforme descrito no tópico 3.3.1, foi verificado se a atribuição de usuário para cada foto, informada pelo sistema de autenticação facial, foi correta ou incorreta. Desta forma, investigou-se os cenários de posição de rosto e iluminação da foto que podem interferir na avaliação correta do sistema, onde as fotos utilizadas para teste se encontram na Figura 23, e as imagens marcadas com contorno vermelho indicam as fotos cuja atribuição feita pelo sistema foi incorreta.

Figura 23: Imagens de teste para simulação do ambiente de autenticação facial.



Fonte: Autoria própria.

Examinando a resolução determinada pelo sistema para as imagens de teste, nota-se que as mudanças angulares do rosto no plano sagital, que divide o corpo humano em metades direita e esquerda, não exercem influência na identificação correta do usuário apresentado. Também se observa que, tanto em situações em que a CNN possuiu mais fotos do usuário para o seu treinamento e menos fotos de teste, quando em situações em que a CNN possuiu menos fotos de usuário e mais fotos para teste, a identificação do usuário não foi impactada, sendo esta impactada principalmente pela iluminação da foto apresentada. Nota-se que, apenas em situações de fotos capturadas em cenários de baixíssima luminosidade, a CNN falhou na identificação do usuário presente na foto, como demonstrado na Figura 24.

Figura 24: Imagens de teste cuja CNN não identificou o usuário corretamente.



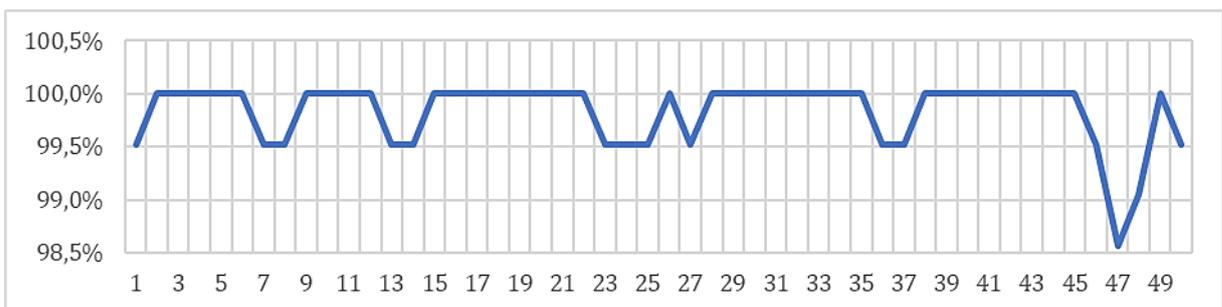
Fonte: Autoria própria.

Considerando os casos em que houve atribuição equivocada do usuário, nota-se que situações de baixa iluminação dificultam a identificação precisa das características faciais, resultando em identificações errôneas. Isso ocorre devido à diminuição do contraste nas imagens, o que impacta negativamente a correta identificação do usuário apresentado na imagem.

5.2 AVALIAÇÃO QUANTITATIVA

Utilizando as métricas de avaliação descritas no tópico 3.3.4, realizou-se o cálculo da acurácia, precisão, sensibilidade, especificidade, TFP, TFN e *F-score* para cada classe, ou seja, para cada usuário. Analisando o gráfico obtido da acurácia em relação à classe, disposto na Figura 25, observa-se que a acurácia mínima foi igual a 98,6% e a acurácia máxima foi igual a 100%, apontando que a maioria das avaliações de usuário realizada pela CNN foram performadas corretamente, identificando corretamente o usuário apresentado. Este dado pode estar relacionado ao fato de a maioria das fotos presentes no banco apresentar uma boa iluminação para a identificação de características, utilizadas para a categorização do usuário.

Figura 25: Gráfico da acurácia (ACC) X classe.

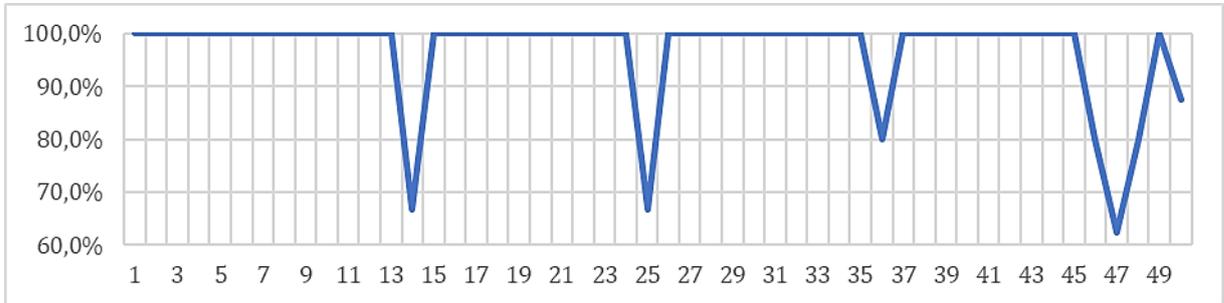


Fonte: Autoria própria.

Analisando o gráfico obtido da precisão, ou do inglês *recall*, em relação a classe, disposto na Figura 26, observa-se que a precisão mínima foi igual a 62,5% e a precisão máxima foi igual a 100%, informando a proporção das classificações positivas cujo valor esperado era realmente positivo. Essa métrica é especialmente significativa na avaliação de classificadores

aonde os resultados falso positivos impactam muito o sistema, como é o caso do sistema de autenticação desenvolvido neste projeto.

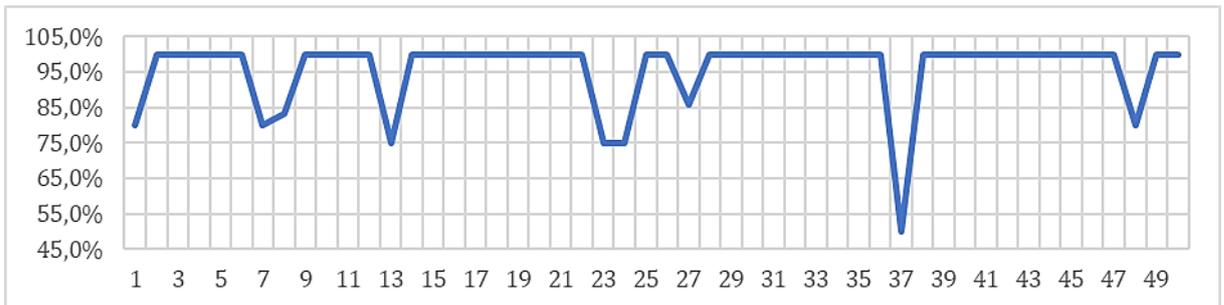
Figura 26: Gráfico da precisão X classe.



Fonte: Autoria própria.

Analisando o gráfico obtido da sensibilidade em relação a classe, disposto na Figura 27, observa-se que a sensibilidade mínima foi igual a 50% e a sensibilidade máxima foi igual a 100%, informando das instâncias verdadeiramente positivas no conjunto de dados, quantas o modelo conseguiu identificar corretamente.

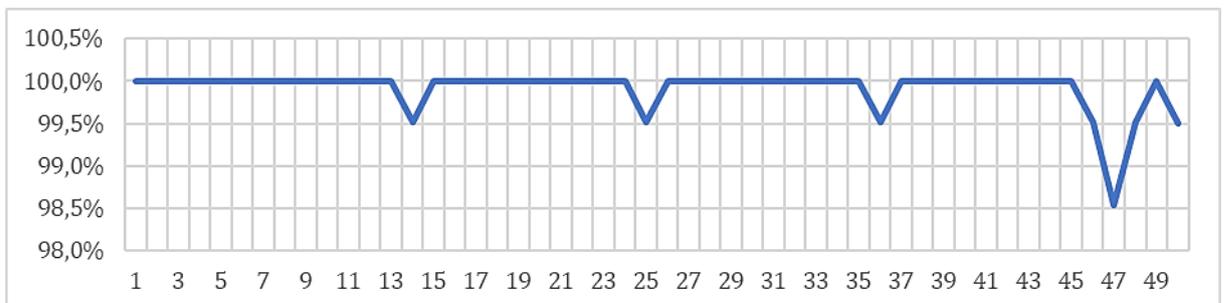
Figura 27: Gráfico da sensibilidade (TPR) X classe.



Fonte: Autoria própria.

Analisando o gráfico obtido da especificidade em relação a classe, disposto na Figura 28, observa-se que a especificidade mínima foi igual a 98,5% e a especificidade máxima foi igual a 100%, informando nas instâncias verdadeiramente negativas no conjunto de dados, quantas o modelo identificou corretamente como negativas.

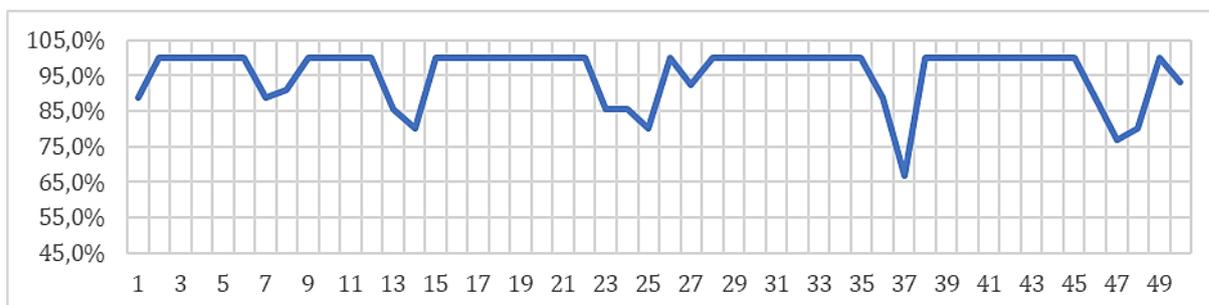
Figura 28: Gráfico da especificidade (TNR) X classe.



Fonte: Autoria própria.

Analisando o gráfico obtido da *F-score* em relação a classe, disposto na Figura 29, observa-se que a pontuação mínima foi igual a 66,7% e a pontuação máxima foi igual a 100%, informando uma medida mais equilibrada do desempenho do modelo em relação aos resultados falso positivos e falso negativos, quando ambos impactam no funcionamento do projeto.

Figura 29: Gráfico *F-score* X classe.



Fonte: Autoria própria.

Após a realização das avaliações dos usuários utilizando a CNN, os resultados obtidos foram compilados para criar uma matriz de confusão multinível, apresentada de forma detalhada na Tabela 1. Essa matriz de confusão fornece informações sobre a quantidade de usuários previstos pela CNN em relação aos usuários que foram identificados corretamente, onde a representação visual da matriz utiliza uma escala de cores graduada, em que as cores mais frias representam valores mais baixos e as cores mais quentes representam valores mais altos.

Um aspecto interessante revelado pela análise da matriz de confusão é a variação na quantidade de fotos de teste utilizadas para cada usuário, com um total de 1 a 7 fotos de teste atribuídas a cada usuário. Isso corresponde a uma gama de variação que abrange de 7% a 50% do total de fotos do usuário sendo empregadas nos testes de autenticação.

Além disso, um ponto de destaque que emerge da análise da matriz de confusão é que a CNN demonstrou uma capacidade notável de identificar corretamente os usuários na maioria das fotos apresentadas. Isso é evidenciado pelo maior agrupamento de valores na diagonal principal da matriz de confusão, que representa os casos em que as previsões da CNN coincidem com as identificações corretas dos usuários. Esse padrão reflete um bom desempenho da CNN na tarefa de autenticação dos usuários com base nas fotos fornecidas. Ao considerar esses achados da matriz de confusão, podemos concluir que a CNN mostrou um nível geral de acurácia promissor na identificação dos usuários, validando sua eficácia em uma variedade de situações de teste.

métricas fornecem uma compreensão mais detalhada do desempenho do modelo para cada classe específica. Métricas como precisão, sensibilidade (*recall*), especificidade e *F-score* podem ser calculados para cada classe, permitindo uma análise mais precisa de como o modelo está se comportando em diferentes cenários.

Tabela 2: Resultados obtidos e métricas de avaliação do projeto.

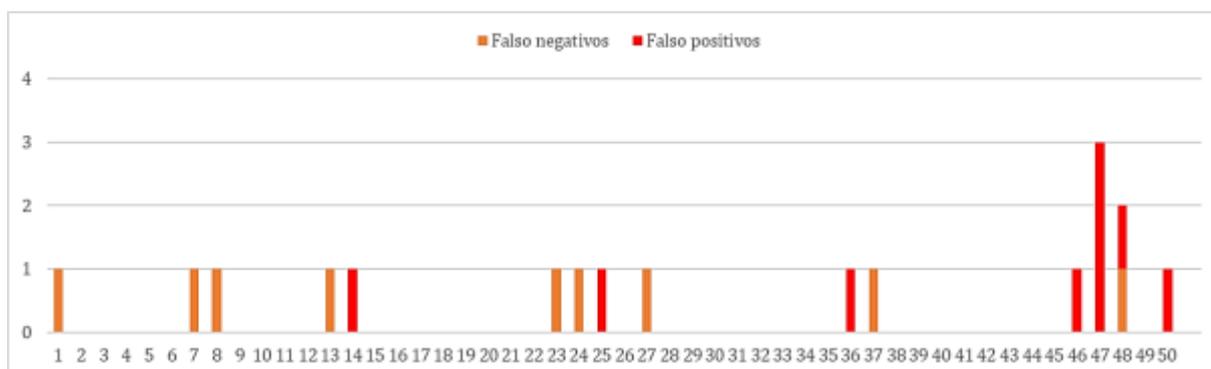
| Classe | Matriz de confusão (Quantidade) | | | | | | | | Métricas de avaliação (%) | | | | | | |
|--------|---------------------------------|----|----|-----|-------|-------|---------|-------|---------------------------|----------|-------|-------|-------|-------|---------|
| | VP | FN | FP | VN | Real+ | Real- | Acertos | Erros | ACC | Precisão | TPR | TNR | FPR | FNR | F-score |
| 1 | 4 | 1 | 0 | 205 | 5 | 205 | 209 | 1 | 0,995 | 1 | 0,800 | 1 | 0 | 0,200 | 0,889 |
| 2 | 5 | 0 | 0 | 205 | 5 | 205 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 3 | 1 | 0 | 0 | 209 | 1 | 209 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 4 | 6 | 0 | 0 | 204 | 6 | 204 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 5 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 6 | 5 | 0 | 0 | 205 | 5 | 205 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 7 | 4 | 1 | 0 | 205 | 5 | 205 | 209 | 1 | 0,995 | 1 | 0,800 | 1 | 0 | 0,200 | 0,889 |
| 8 | 5 | 1 | 0 | 204 | 6 | 204 | 209 | 1 | 0,995 | 1 | 0,833 | 1 | 0 | 0,167 | 0,909 |
| 9 | 6 | 0 | 0 | 204 | 6 | 204 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 10 | 2 | 0 | 0 | 208 | 2 | 208 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 11 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 12 | 6 | 0 | 0 | 204 | 6 | 204 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 13 | 3 | 1 | 0 | 206 | 4 | 206 | 209 | 1 | 0,995 | 1 | 0,750 | 1 | 0 | 0,250 | 0,857 |
| 14 | 2 | 0 | 1 | 207 | 2 | 208 | 209 | 1 | 0,995 | 0,667 | 1 | 0,995 | 0,005 | 0 | 0,800 |
| 15 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 16 | 6 | 0 | 0 | 204 | 6 | 204 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 17 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 18 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 19 | 7 | 0 | 0 | 203 | 7 | 203 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 20 | 6 | 0 | 0 | 204 | 6 | 204 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 21 | 3 | 0 | 0 | 207 | 3 | 207 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 22 | 3 | 0 | 0 | 207 | 3 | 207 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 23 | 3 | 1 | 0 | 206 | 4 | 206 | 209 | 1 | 0,995 | 1 | 0,750 | 1 | 0 | 0,250 | 0,857 |
| 24 | 3 | 1 | 0 | 206 | 4 | 206 | 209 | 1 | 0,995 | 1 | 0,750 | 1 | 0 | 0,250 | 0,857 |
| 25 | 2 | 0 | 1 | 207 | 2 | 208 | 209 | 1 | 0,995 | 0,667 | 1 | 0,995 | 0,005 | 0 | 0,800 |
| 26 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 27 | 6 | 1 | 0 | 203 | 7 | 203 | 209 | 1 | 0,995 | 1 | 0,857 | 1,000 | 0 | 0,143 | 0,923 |
| 28 | 6 | 0 | 0 | 204 | 6 | 204 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 29 | 1 | 0 | 0 | 209 | 1 | 209 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 30 | 5 | 0 | 0 | 205 | 5 | 205 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 31 | 2 | 0 | 0 | 208 | 2 | 208 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 32 | 5 | 0 | 0 | 205 | 5 | 205 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 33 | 3 | 0 | 0 | 207 | 3 | 207 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 34 | 2 | 0 | 0 | 208 | 2 | 208 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 35 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 36 | 4 | 0 | 1 | 205 | 4 | 206 | 209 | 1 | 0,995 | 0,800 | 1 | 0,995 | 0,005 | 0 | 0,889 |
| 37 | 1 | 1 | 0 | 208 | 2 | 208 | 209 | 1 | 0,995 | 1 | 0,500 | 1 | 0 | 0,500 | 0,667 |
| 38 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 39 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 40 | 5 | 0 | 0 | 205 | 5 | 205 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 41 | 5 | 0 | 0 | 205 | 5 | 205 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 42 | 3 | 0 | 0 | 207 | 3 | 207 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 43 | 5 | 0 | 0 | 205 | 5 | 205 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 44 | 4 | 0 | 0 | 206 | 4 | 206 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 45 | 3 | 0 | 0 | 207 | 3 | 207 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 46 | 4 | 0 | 1 | 205 | 4 | 206 | 209 | 1 | 0,995 | 0,800 | 1 | 0,995 | 0,005 | 0 | 0,889 |
| 47 | 5 | 0 | 3 | 202 | 5 | 205 | 207 | 3 | 0,986 | 0,625 | 1 | 0,985 | 0,015 | 0 | 0,769 |
| 48 | 4 | 1 | 1 | 204 | 5 | 205 | 208 | 2 | 0,990 | 0,800 | 0,800 | 0,995 | 0,005 | 0,200 | 0,800 |
| 49 | 3 | 0 | 0 | 207 | 3 | 207 | 210 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 50 | 7 | 0 | 1 | 202 | 7 | 203 | 209 | 1 | 0,995 | 0,875 | 1 | 0,995 | 0,005 | 0 | 0,933 |

Fonte: Autoria própria.

Ao analisar os dados apresentados na Tabela 2, é possível identificar algumas tendências significativas. No que se refere às 9 fotos nas quais o usuário não foi identificado corretamente, é notável que não houve nenhum caso em que mais de uma foto do mesmo usuário tenha sido corretamente identificada. Em outras palavras, em cada uma dessas situações

de identificação errônea, apenas uma foto foi classificada erroneamente para um determinado usuário. Isso implica que não houve repetição de erros de identificação para esses 9 usuários em particular. No entanto, ao direcionar a atenção para os casos de falsos positivos, foi observado que, no contexto específico do usuário identificado como número 47, três fotos foram erroneamente atribuídas a esse mesmo usuário. Isso indica que o modelo cometeu o erro de associar três fotos distintas a um único usuário, resultando em múltiplos falsos positivos para essa classe específica, como demonstrado na Figura 30.

Figura 30: Gráfico dos erros X classe, sendo os erros discriminados entre falso negativos e falso positivos.



Fonte: Autoria própria.

É de extrema importância considerar que, em um contexto de autenticação facial, os resultados de falsos positivos possuem um impacto significativamente mais prejudicial do que os resultados de falsos negativos. Essa conclusão assume um papel particularmente crítico pois, mesmo com a apresentação de 9 fotos distintas do usuário identificado como número 47, o sistema não conseguiu evitar a ocorrência de 3 falsos positivos detectados.

Essa observação salienta a necessidade de atenção especial aos casos em que o sistema erroneamente reconhece um indivíduo como autêntico quando, na verdade, não é. Em um cenário de autenticação, um falso positivo pode levar a sérias implicações, incluindo acesso não autorizado a sistemas protegidos e potencialmente dados sensíveis. Portanto, a detecção de múltiplos falsos positivos, como exemplificado pelo usuário 47, enfatiza a necessidade de estratégias e ajustes para reduzir essa ocorrência indesejada. Essa análise destaca o equilíbrio delicado entre a precisão e a sensibilidade do modelo, ressaltando que a minimização de falsos positivos é uma prioridade crucial em cenários de segurança. Os esforços para aprimorar o desempenho do sistema devem ser direcionados a identificar e corrigir as situações em que ocorrem os falsos positivos, mesmo quando diversas instâncias de um usuário específico são apresentadas ao modelo.

Portanto, reconhecer a diferença entre a quantidade de fotos fornecidas e a ocorrência de falsos positivos não deve ser considerado como uma justificativa para as falhas na identificação do usuário. No entanto, a qualidade das fotos de teste utilizadas para verificar o sistema deve ser considerada como ponto de partida para melhorar a otimização do sistema de autenticação facial. O objetivo é assegurar a confiabilidade e integridade do processo de identificação, e uma alternativa para alcançar um cenário de autenticação mais eficaz poderia ser estabelecer um critério mínimo de iluminação para iniciar a autenticação.

5.3 TEMPO DE PROCESSAMENTO

Após concluir as fases cruciais de treinamento e autenticação de usuários, conforme delineado nos pontos 4.1.1 e 4.1.3, respectivamente, os resultados obtidos informam que o processo de treinamento do conjunto de 50 usuários culminou em um intervalo total de 11 minutos e 56 segundos, e como resultado, uma média aproximada de 14 segundos foi aferida para o treinamento individual de um usuário. Vale ressaltar que esse valor aproximado é atingido considerando a variabilidade no número de fotografias empregadas no treinamento de cada usuário, tendo em vista que a discrepância na quantidade de fotos para treinamento entre os usuários pode influenciar esse tempo médio, uma vez que a complexidade do aprendizado pode ser diversificada de acordo com os indivíduos.

Além disso, em relação ao processo de autenticação, uma análise semelhante foi conduzida. Durante o teste do sistema, um conjunto de 210 fotos foi empregado para a avaliação da autenticidade, onde o tempo necessário para concluir a autenticação dessas 210 imagens totalizou aproximadamente 13 segundos.

Esses dados revelam informações importantes sobre o tempo necessário para as etapas principais de treinamento e autenticação, onde o treinamento de um único usuário pode ser afetado pela quantidade de fotos usadas, e a autenticação demonstrou uma notável rapidez mesmo diante de um conjunto substancial de imagens a serem avaliadas. Com base nesses dados, observa-se que o intervalo de tempo necessário para a execução do treinamento e autenticação não exercem impacto na experiência do usuário e na eficiência operacional do projeto.

6 CONCLUSÃO

A visão computacional busca traduzir para a linguagem de máquina, modelos matemáticos robustos, uma observação óptica da realidade a que nos é apresentada. Desta forma, busca-se utilizar de ferramentas, como a computação, para otimizar grandes categorizações de informação (Milano & Honorato, 2014). Em posse das características de Haar, por exemplo, pode-se sumarizar imagens complexas em unidades de tratamento mais simples para manipulação, e por meio do algoritmo de AdaBoost, pode-se estruturar melhor os classificadores para depois, classificá-los.

A utilização de redes neurais artificiais também se utiliza de mais processamento, para submeter a imagem a múltiplas camadas de dados, e realizar uma análise profunda das características que a compõe. E ao tratar os dados não como níveis individuais, mas sim como um conjunto de características que se relacionam entre si, nos aproximamos cada vez mais do sistema neural biológico com os quais percebemos o mundo (Costa, 2018). Entretanto, ainda há muito caminho a percorrer, tendo em vista as diversas diferenças entre o sistema biológico e o neural, como a não sintetização de mais de um tipo de sinal por vez, a velocidade de processamento, entre outros.

Entretanto, a aplicação das camadas de convolução são uma grande ferramenta para manter informações espaciais relevantes, e que se perderiam na camada de reformatação; isso porque os níveis de cinza da imagem não são pontos isolados no espaço, mas se relacionam com a vizinhança. Principalmente referente a autenticação facial, aonde pequenos detalhes como a distância entre os olhos e posição do nariz influenciam bastante para a classificação dos atributos de cada rosto, utiliza-se ao máximo da ferramenta de convolução para preservar dessas características.

Ao analisar a imagem como um sinal que carrega informação, e que quanto mais otimizar a percepção destas informações no conjunto da imagem, pode-se treinar e desenvolver diferentes redes neurais para melhorar o nível de eficiência da ferramenta. A rede neural utilizada neste projeto realiza de mais de uma etapa de convolução com o objetivo de reter o máximo de informação com o mínimo de espaço. Contudo sua aplicação ainda é limitada pela qualidade e quantidade de imagens presentes no banco de dados utilizado, o que restringe a sua aplicação para condições em que não se tem muita informação do objeto ou característica a ser identificada.

Considerando os pontos levantados, conclui-se que a utilização de redes neurais convolucionais são uma boa estratégia para identificação estática de pessoas em condições de pouca variação da posição do rosto. Por isso, aplicações em catracas ou portas de laboratórios, ou em filas de atendimento, na confirmação de identidade por oficiais de segurança, de atendimento, ou de governo, aumentando a segurança e velocidade na realização das tarefas diárias do ser humano. Este sistema também pode oferecer a preservação de documentos sensíveis ou oferecer assistência para deficientes visuais nas mais diversas situações, e em conjunto com as inteligências artificiais, podem revolucionar o mundo em que vivemos. Dada a demanda por alto processamento, e o crescente desenvolvimento de abordagens para a visão computacional, avalia-se que métodos de redes neurais artificiais possuem uma vantagem inerente de poder se moldar para as mais diferentes finalidades, estando aplicada em diferentes contextos e, portanto, oferecendo um conhecimento sólido e moldável para ambientes e circunstâncias diferentes.

Após a análise dos resultados deste projeto, fica evidente que a qualidade da imagem utilizada para validação desempenhou um papel muito mais significativo na precisão do reconhecimento do usuário do que as imagens utilizadas para o treinamento da rede neural. Essa constatação sublinha a importância crucial de direcionar esforços para melhorar a captura e seleção das imagens utilizadas na fase de validação do sistema de reconhecimento facial. Desta forma, para futuros estudos e projetos nesta área, seria altamente benéfico focar na otimização do processo de captura de imagem, especialmente para a fase de validação. Melhorar a qualidade, resolução e representatividade das imagens usadas para validar a eficácia da rede neural teria um impacto extremamente positivo no desempenho geral do sistema de reconhecimento facial.

Além disso, a constatação de que a qualidade da imagem de validação é mais influente do que as imagens de treinamento sugerem que a coleta e preparação adequadas dos dados de validação são cruciais para garantir um modelo confiável e preciso. Portanto, abordagens que considerem a qualidade das imagens de validação, como técnicas de pré-processamento de imagem ou seleção cuidadosa de conjuntos de validação, podem ser áreas de pesquisa promissoras para aprimorar ainda mais o desempenho dos sistemas de reconhecimento facial baseados em redes neurais.

REFERÊNCIAS

- [1] *AlexNet*. (2023). Fonte: MathWorks Help: <https://www.mathworks.com/help/deeplearning/ref/alexnet.html>
- [2] Autores, V. (2012). *Avanços em Visão Computacional* (4 ed., Vol. 1). (L. A. Neves, H. V. Neto, & A. Gonzaga, Eds.) Curitiba: Omnipax.
- [3] Batista, G. d., Neves, F. d., Santos, I. F., Schinaid, A. A., & Sousa, W. L. (2017). *Sistema de Identificação e Autenticação Biométrica*. Atividade Prática (Ciência da Computação - UNIP), São Paulo.
- [4] Biesseck, B. J., & Zacarkim, V. L. (01 de Março de 2021). Avaliação da CNN FaceNet para reconhecimento facial de estudantes em sala de aula. *Brazilian Journal of Development*, 7, pp. 27558-27563. doi:10.34117/bjdv7n3-455
- [5] Chandra, A. (20 de Fevereiro de 2019). *FEI face data*. Fonte: kaggle: <https://www.kaggle.com/datasets/whizzkid/fei-face-data>
- [6] Chaves, R. R. (2018). Redes neurais deep learning aplicadas ao reconhecimento facial. *Monografia (Graduação em Engenharia de Computação)*.
- [7] Costa, T. F. (2018). Reconhecimento facial com Haar Cascade e KNN. *Monografia (Graduação em Engenharia de Computação)*, 53.
- [8] Feitosa, R. Q., & Pereira, C. V. (2003). FACEID - Sistema de Reconhecimento de Faces. 5. Filho, O. M., & Neto, H. V. (1999). *Processamento Digital de Imagens*. Rio de Janeiro: Brasport.
- [9] Fonseca, F. O. (2016). Detector de faces utilizando filtro de características. *Dissertação (Mestrado em Engenharia de Telecomunicações)*, 111.
- [10] GIL, Antônio Carlos. Como Elaborar Projetos de Pesquisa. 6. ed. São Paulo: Atlas, 2019. Gonzales, R. C., & Woods, R. E. (2018). *Digital Image Processing*. Nova York: Pearson.
- [11] Helene, O., & Helene, A. F. (Julho de 2011). Alguns aspectos da óptica do olho humano. *Revista Brasileira de Ensino de Física*.
- [12] Ignacio, J. d. (2013). *Processamento e Análise Digital de Imagens em estudos da cinética de recristalização de ligas Al-Mg-X*. Dissertação (Mestrado em Ciência de Tencologia Nuclear - Materiais) - Instituto de Pesquisas Energéticas Nucleares, São Paulo.
- [13] Juraszek, G. D. (2014). *Reconhecimento de produtos por imagen utilizando palavras visuais e redes neurais*. Joinville: Dissertação (Mestrado em Computação Aplicada).

- [14] LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). Convolutional Networks and Applications in Vision. *IEEE International Symposium on Circuits and Systems*. doi:10.1109/iscas.2010.5537907
- [15] Magalhães, P. S., & Santos, H. D. (2003). Biometria e autenticação. *Actas da 4ª Conferência da Associação Portuguesa de Sistemas de Informação*. Porto., 17.
- [16] Maia, H. L. (2016). *Detecção e Reconhecimento Facial por meio de Aprendizado de Máquina*. Monografia (Graduação em Engenharia de Redes de Comunicação) - Universidade de Brasília, Brasília.
- [17] Milano, D. d., & Honorato, L. B. (2014). Visão Computacional. p. 7.
- [18] Queiroz, J. E., & Gomes, H. M. (2001). Introdução ao Processamento Digital de Imagens. *Revista RITA*, p. 31.
- [19] *Rede neural convolucional*. (26 de Novembro de 2022). Fonte: Medium: <https://medium.com/@leticia.slopes/rede-neural-convolucional-f96699679e6d>
- [20] Silva, P. V. (30 de Março de 2023). Métodos estatísticos adequados ao reconhecimento facial. *Monografia (Graduação em Engenharia de Computação)*, 59.