

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

Thiago Oliveira de Santana

**COMPARAÇÃO ENTRE TÉCNICAS DE  
APRENDIZADO DE MÁQUINA NO PROCESSO  
DE IDENTIFICAÇÃO BIOMÉTRICA ATRAVÉS  
DE IMAGENS DA ORELHA**

Ouro Preto, MG  
2019

Thiago Oliveira de Santana

COMPARAÇÃO ENTRE TÉCNICAS DE APRENDIZADO DE MÁQUINA NO PROCESSO  
DE IDENTIFICAÇÃO BIOMÉTRICA ATRAVÉS DE IMAGENS DA ORELHA

Monografia II apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

**Orientador:** Prof. Dr. Guillermo Cámara Chávez

Ouro Preto, MG  
2019

S591c

Santana, Thiago Oliveira de .

Comparação entre técnicas de aprendizado de máquina no processo de identificação biométrica através de imagens da orelha [manuscrito] / Thiago Oliveira de Santana. - 2019.

62f.: il.: color; tabs.

Orientador: Prof. Dr. Guillermo Cámara Chávez.

Monografia (Graduação). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Departamento de Computação.

1. Identificação biométrica. 2. Orelhas. 3. Redes neurais (Computação). I. Cámara Chávez, Guillermo . II. Universidade Federal de Ouro Preto. III. Título.

CDU: 004.85

Catálogo: [ficha.sisbin@ufop.edu.br](mailto:ficha.sisbin@ufop.edu.br)

Thiago Oliveira de Santana

**COMPARAÇÃO ENTRE TÉCNICAS DE APRENDIZADO DE  
MÁQUINA NO PROCESSO DE IDENTIFICAÇÃO BIOMÉTRICA  
ATRAVÉS DE IMAGENS DA ORELHA**

Monografia II apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau em Bacharel em Ciência da Computação.

Aprovada em Ouro Preto, 12 de dezembro de 2019.



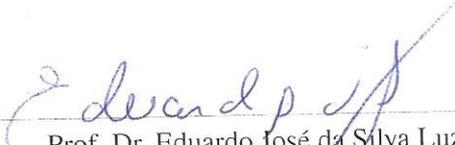
---

Prof. Dr. Guillermo Cámara Chávez  
Universidade Federal de Ouro Preto  
Orientador



---

Profa. Dra. Roberta Barbosa Oliveira  
Universidade de Brasília - UnB  
Examinador



---

Prof. Dr. Eduardo José da Silva Luz  
Universidade Federal de Ouro Preto - UFOP  
Examinador

# Agradecimentos

Primeiramente a Deus, pelo fim de mais essa etapa, pelos sonhos que se concretizam e por ter permitido que tudo isso acontecesse ao longo da minha vida, não somente nestes anos mas que todos os momentos é o maior mestre que alguém pode conhecer. Grato sou pelo dom da vida e pelo seu amor infinito. Porque Dele, e por meio Dele, e para Ele são todas as coisas.

Aos meus pais, meu abrigo seguro, de onde recebi apoio incondicional nessa empreitada. Os primeiros a sonhar tudo isso, apesar de todas as dificuldades e de todos os contratemplos sempre me incentivaram e despertaram em mim o interesse pelos estudos. Por eles terem me ensinado sempre os caminhos corretos para lutar pelos meus objetivos e alcançá-los com dignidade e humildade. Por eles vierem preparando o meu caminho, desde que eu nasci, para que esse dia enfim chegasse. Acredito que sem o apoio de vocês seria muito difícil vencer esse desafio, sendo assim agradeço por acreditarem sempre no meu potencial.

A minha amada esposa, por me tonar capaz de enfrentar novos desafios sabendo que estará sempre ao meu lado e que nos momentos da minha ausência dedicada aos estudos, sempre fez entender que o futuro é feitor a partir da constante dedicação no presente.

Aos meus irmãos, pela compreensão do tempo de convívio muitas vezes sacrificado durante esses anos.

Aos meus colegas, por ter compartilhado os prazeres e dificuldades desta jornada com os quais convivemos durante tantas horas e carregamos a marca de experiências comuns que tivemos. Partamos confiantes em busca de nossos ideais, no exercício de nossa profissão.

Ao meu orientador Guillermo Cámara Chávez, pelo suporte dado durante o desenvolvimento do trabalho, contribuindo muito para que esse projeto fosse realizado.

À todos professores, pelo estímulo acadêmico e pela valorização cultural que atribuem ao processo pedagógico. Pela amizade que se construiu, pelos ensinamentos e orientações prestados que ao longo desses anos ajudou a construir parte da minha carreira profissional.

# Resumo

Os métodos tradicionais de identificação pessoal baseados em conhecimento (por exemplo, senhas) e baseados em cartões (por exemplo, cartões de identidade) podem ser facilmente esquecidos, clonados, roubados ou perdidos. Com isso, a biometria que é uma ciência que estuda a mensuração dos seres vivos, tem sido amplamente utilizada como sendo um mecanismo de segurança para identificação. Os recentes avanços na tecnologia e a vulnerabilidade dos sistemas de autenticação tradicional têm fomentado diversas pesquisas nesta área, dando origem à sistemas inovadores e eficientes para identificação biométrica. Nos últimos anos, a orelha humana obteve um significativo interesse pelo motivo desta apresentar algumas vantagens quando comparadas a outras medidas biométricas. A forma da orelha, por exemplo, é menor do que a face o que permite menor custo de computação e, portanto, o torna adequada para aplicações em tempo real. Além disso, diferente de outros, a orelha não muda de acordo com a vida humana e a captura de imagens à distância destas pode ser feita facilmente com uma câmera comum. Portanto, neste trabalho são propostos métodos para identificação através de imagens de orelhas, sendo que alguns métodos utilizam o extrator de características locais *Histograma de Gradientes Orientados (HOG)* combinado com classificadores convencionais encontrados na literatura (K-Vizinhos Mais Próximos (K-NN), Análise Discriminante Linear (LDA) e Máquinas de Vetores de Suporte (SVM)), e outros utilizam diferentes arquiteturas de *Redes Neurais Convolucionais* também presentes na literatura (Alexnet, Squeezenet e Googlenet), com o intuito de realizar uma comparação entre os mesmos. Para a realização dos experimentos foi utilizada a base de dados IIT Delhi Ear disponibilizada pelo *Indian Institute of Technology Delhi* e a base AMI Ear disponibilizada pela *Universidad de Las Palmas de Gran Canaria (ULPGC)*. A primeira base contém dois conjuntos de imagens de orelha que são formados por 125 pessoas que estão na faixa etária de 14 a 58 anos na qual cada pessoa tem pelo menos três imagens de orelha. A segunda é constituída de 700 imagens que envolve 100 pessoas diferentes com idade entre 19 e 65 anos, sendo seis imagens da orelha direita e uma da orelha esquerda. Os resultados mostram que o HOG combinado com o SVM para a base de dados IIT Delhi Ear supera as demais abordagens, perfazendo uma taxa média de acurácia de 93,78%, ao contrário do que acontece para a base de dados AMI Ear, sendo a melhor abordagem o método que utiliza a arquitetura de Rede Neural Convolucional *Alexnet*, alcançando uma taxa média de acurácia de 95,40%.

**Palavras-chave:** Biometria. Identificação Biométrica. Orelhas. Histograma de Gradientes Orientados. Classificadores Convencionais. Aprendizado Profundo. Redes Neurais Convolucionais.

# Abstract

Traditional knowledge-based (e.g. passwords) and card-based (e.g. identity cards) personal identification methods can be easily forgotten, cloned, stolen or lost. Therefore, biometrics, which is a science that studies the measurement of living beings, has been widely used as a security mechanism for identification. Recent advances in technology and the vulnerability of traditional authentication systems have fostered much research in this area, leading to innovative and efficient biometric identification systems. In recent years, the human ear has gained significant interest because it has some advantages compared to other biometric measurements. The shape of the ear, for example, is smaller than the face which allows for lower computing cost and therefore makes it suitable for real time applications. In addition, unlike others, the ear does not change with human life, and distant images can be easily captured with an ordinary camera. Therefore, this work proposes methods for identification through ear images, and some methods use the local characteristic extractor *Histogram of Oriented Gradients (HOG)* combined with conventional classifiers found in the literature (K-Nearest Neighbor, Linear Discriminant Analysis e Support Vector Machine), and others use different *Convolutional Neural Networks* architectures also present in the literature (Alexnet, Squeezenet e Googlenet), in order to make a comparison between them. To perform the experiments, the IIT Delhi Ear database provided by the Indian Institute of Technology Delhi and the AMI Ear database provided by the University of Las Palmas of Gran Canaria (ULPGC) were used. The first base contains two sets of ear images that are made up of 125 people who are in the 14 to 58 age group in which each person has at least three ear images. The second one consists of 700 images involving 100 different people aged 19 to 65 years, six images of the right ear and one of the left ear. The results show that the HOG combined with SVM for the IIT Delhi Ear database outperforms the other approaches, yielding an average accuracy rate of 93.78 %, as opposed to the AMI Ear database, the best approach being the method using the Convolutional Neural Network architecture *Alexnet*, achieving an average accuracy rate 95.40 %.

**Keywords:** Biometrics, Biometric Identification, Ears, Histogram of Oriented Gradients, Conventional Classifiers, Deep Learning, Convolutional Neural Networks.

# Lista de Ilustrações

Figura 1.1 – Exemplos de modalidades biométricas (ROSS, 2007). . . . .	2
Figura 1.2 – Componentes da Orelha (UOL, 2015). . . . .	3
Figura 3.1 – Exemplo da segmentação de uma imagem da orelha (NOSRATI; FAEZ; FARADJI, 2007) - a) Imagem original e b) Segmentação com correspondência de modelos. . . . .	16
Figura 3.2 – Exemplo de aplicação do HOG (HYPRAPTIVE, 2017). a) Imagem de entrada e b) Imagem resultante. . . . .	16
Figura 3.3 – Etapas do HOG (CERNA, 2014). . . . .	17
Figura 3.4 – A Hierarquia do aprendizado (MONTEIRO, 2016). . . . .	18
Figura 3.5 – Exemplo de classificação do KNN com dois rótulos de classe e $k = 7$ (PACHECO, 2019). . . . .	21
Figura 3.6 – Exemplo de escolha de fronteiras de separação (MENOTTI, 2011). . . . .	21
Figura 3.7 – Exemplo de mapeamento bidimensional para tridimensional (MENOTTI, 2011). . . . .	22
Figura 3.8 – Arquitetura LeNet-5, destacando as operações aplicadas ao longo da rede (JUNIOR, 2018). . . . .	23
Figura 3.9 – Camada convolucional parcial, utilizando um filtro $3 \times 3$ (PAVLOVSKY, 2017). . . . .	24
Figura 3.10–Exemplo de aplicação da função ReLU sobre uma mapa de características (SINGHAL, 2017). . . . .	25
Figura 3.11–Exemplo de aplicação do <i>max pooling</i> e <i>average pooling</i> de tamanho $2 \times 2$ com passo 2 sobre uma mapa de características (SINGHAL, 2017). . . . .	26
Figura 3.12–Arquitetura da Rede Neural Alexnet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). . . . .	27
Figura 3.13–Módulo <i>Inception</i> (SZEGEDY et al., 2015). . . . .	28
Figura 3.14–Módulo <i>Incêndio</i> (IANDOLA et al., 2016). . . . .	28
Figura 4.1 – Metodologia proposta para os métodos. . . . .	30
Figura 4.2 – Validação cruzada 4-fold (SOUZA, 2016). . . . .	31
Figura 4.3 – Exemplo de extração de características - a) Imagem da orelha e b) Descrição resultante. . . . .	32
Figura 4.4 – Exemplo do aumento dos dados realizados - a) Imagem original, b) Espelhamento, c) Rotação em 30 graus e d) Aumento de escala (1.2x). . . . .	33
Figura 5.1 – Amostras da base IITD Ear - a) Imagem do conjunto I e b) Imagem do conjunto II. . . . .	35
Figura 5.2 – Exemplos de imagens da AMI Ear. . . . .	36
Figura 5.3 – <b>IITD Ear</b> - Exemplo de confusão na classificação. . . . .	38
Figura 5.4 – <b>AMI Ear</b> - Exemplo de oclusão parcial por brinco. . . . .	38

# Lista de Tabelas

Tabela 2.1 – Comparação das técnicas recentes em reconhecimento biométrico de orelha.	10
Tabela 2.2 – Comparação das técnicas recentes em reconhecimento biométrico de orelha com o uso de aprendizado profundo. . . . .	14
Tabela 4.1 – Parâmetros usados para treinamento dos modelos de CNN. . . . .	34
Tabela 5.1 – <b>IITD Ear</b> - Média dos resultados em relação à Precisão, Revocação, Acurácia e Medida-F1. . . . .	37
Tabela 5.2 – <b>AMI Ear</b> - Média dos resultados em relação à Precisão, Revocação, Acurácia e Medida-F1. . . . .	38
Tabela 5.3 – Comparação ao estado da arte com base na porcentagem de acurácia. . . . .	39

# Lista de Abreviaturas e Siglas

ABC	Artificial Bee Colony
AWE	Annotated Web Ears
BPNN	Back-Propagation Neural Network
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
DWT	Discrete Wavelet Transform
FNCC	Fast Normalized Cross Correlation
HE	Histogram Equalization
HOG	Histogram Of Oriented Gradients
ICA	Independent Component Analysis
ICP	Iterative Closest Point
ITTD	Indian Institute of Technology Delhi
ITWE	In-The-Wild Ear
K-NN	K-Nearest Neighbors
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LSBP	Local Similarity Binary Pattern
MFFNN	Multilayer Feed Forward Neural Network
MSHOG	Multi-Scale Histogram Of Oriented Gradients
PCA	Principal Component Analysis
RBF	Radial Basis Functions
SIFT	Scale Invariant Feature Transform
SURF	Speeded-Up Robust Features
SVM	Support Vector Machine

UERC Unconstrained Ear Recognition Challenge

UND University of Notre Dame

USTB University of Science Technology Beijing

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Justificativa	4
1.2	Objetivos Geral e Específicos	4
1.3	Organização do Trabalho	5
<b>2</b>	<b>Revisão de Literatura</b>	<b>6</b>
2.1	Descritores Manuais	6
2.2	Aprendizagem em Profundidade	11
<b>3</b>	<b>Fundamentação Teórica</b>	<b>15</b>
3.1	Segmentação	15
3.2	Extração de Características	16
3.3	Classificação	18
3.3.1	Aprendizado de Máquina	18
3.3.2	Classificadores Convencionais	19
3.3.2.1	Análise Discriminante Linear	19
3.3.2.2	K-Vizinhos Mais Próximos	20
3.3.2.3	Máquinas de Vetores de Suporte	21
3.3.3	Redes Neurais Convolucionais	23
3.3.3.1	Camada Convolucional	24
3.3.3.2	Camada de Subamostragem	25
3.3.3.3	Camada Totalmente Conectada	26
3.3.3.4	AlexNet	26
3.3.3.5	GoogleNet	27
3.3.3.6	SqueezeNet	28
<b>4</b>	<b>Metodologia</b>	<b>30</b>
4.1	Separação dos Dados	31
4.2	Classificação Convencional	31
4.2.1	Extração de Características	31
4.2.2	Classificação	32
4.3	Classificação com Redes Neurais Convolucionais	32
4.3.1	Aumento dos Dados	32
4.3.2	Classificação	33
<b>5</b>	<b>Experimentos Computacionais</b>	<b>35</b>
5.1	Base de Dados	35
5.1.1	IITD Ear	35
5.1.2	AMI Ear	36
5.2	Métricas de Avaliação	36

5.3	Resultados Alcançados . . . . .	37
5.3.1	Resultados em Comparação ao Estado da Arte . . . . .	39
<b>6</b>	<b>Conclusão . . . . .</b>	<b>40</b>
6.1	Trabalhos Futuros . . . . .	40
	<b>Referências . . . . .</b>	<b>42</b>

# 1 Introdução

Desde os primórdios da civilização humana, características consideradas peculiares como, por exemplo, o rosto e a voz, têm sido utilizadas pelos seres humanos para se reconhecerem mutuamente. Devido a essas características que são únicas em cada ser humano, houve o surgimento do termo identificação biométrica que basicamente tem por objetivo efetuar o reconhecimento de uma pessoa utilizando a tecnologia. Os métodos tradicionais de identificação pessoal baseados em conhecimento (por exemplo, senhas) e baseados em cartões (por exemplo, cartões de identidade) podem ser facilmente esquecidos, clonados, roubados ou perdidos. Com isso, a biometria que é uma ciência que estuda a mensuração dos seres vivos, tem sido amplamente utilizada nos dias atuais como sendo um mecanismo de segurança para identificação, devido a sua eficiência (ALVA; SRINIVASARAGHAVAN; SONAWANE, 2019). Porém, é válido ressaltar que ela existe há séculos, já que no decorrer da história, indícios mostram que mesmo na antiguidade os egípcios já usavam a biometria visando à identificação de pessoas. Estima-se também que em meados do século XIX, o chefe da divisão de identificação criminal da polícia de Paris conhecido como Alphonse Bertillon, utilizava de várias medidas corporais para identificar criminosos (BERTILLON, 1890).

Com o surgimento dos computadores em meados do século XX, foi possível alcançar melhores resultados na identificação humana ocasionados pela automatização dos sistemas (ALMEIDA, 2006). Dessa forma, a exploração de outras características por sistemas computacionais foram viabilizadas. Por conseguinte, as características humanas que podem ser utilizadas para a identificação biométrica pode ser divididas em dois grupos: fisiológicas e comportamentais. As fisiológicas referem-se às características físicas das pessoas na qual devem ser relativamente estáveis como, por exemplo, a orelha. Já as comportamentais diz a respeito às características que está ligada ao comportamento da pessoa como, por exemplo, a forma de andar.

Para efetuar a escolha de uma característica entre as várias específicas do ser humano, deve-se ter uma breve cautela, já que existem inúmeras delas e o custo para utilizá-la em sistema de identificação biométrica varia muito. Por esse motivo é necessário averiguar a relação custo benefício de cada uma e, atentar para alguns aspectos práticos (proteção, desempenho e aceitabilidade) bem como questões teóricas fundamentais (universalidade, singularidade, permanência e mensuralidade). A Figura 1.1 exemplifica algumas características biométricas que podem ser usadas para identificar um indivíduo.

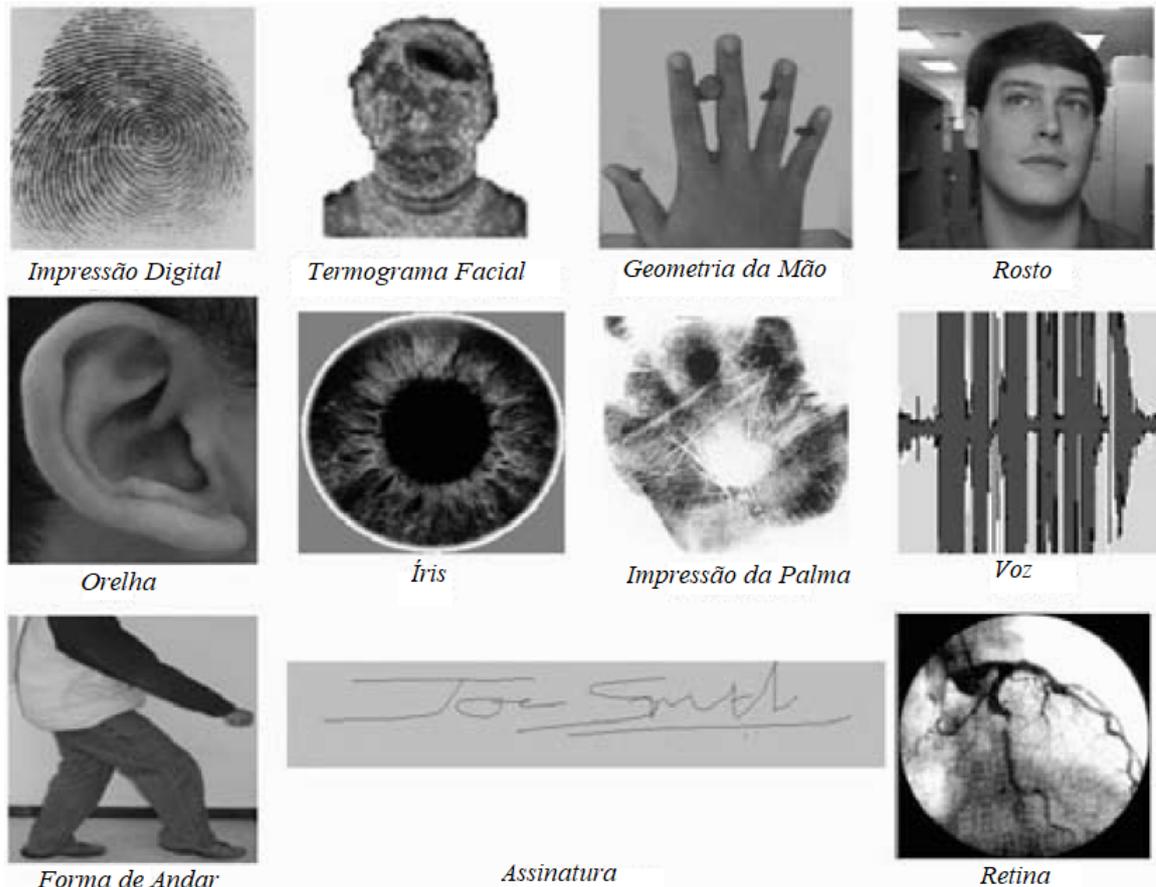


Figura 1.1 – Exemplos de modalidades biométricas (ROSS, 2007).

Além disso, pode-se classificar como sendo intrusivos e não intrusivos, os diferentes métodos de identificação biométrica:

1. **Intrusivo:** são aqueles que só podem ser elaborados para identificação através da autorização das pessoas em análise. Temos como exemplos a íris, a mão, a impressão digital, etc;
2. **Não intrusivo:** são aqueles que podem ser elaborados para identificação independente da autorização das pessoas em análise. Temos como exemplo o reconhecimento da orelha, o reconhecimento facial, o reconhecimento através da forma de andar, etc.

Os recentes avanços na tecnologia e a vulnerabilidade do sistema de autenticação tradicional tem fomentado diversas pesquisas no campo da biometria, dando origem à sistemas inovadores e eficientes para identificação biométrica (ALMEIDA, 2006). Portanto, neste trabalho iremos realizar um estudo sobre um tipo de biometria fisiológico e não intrusivo para a identificação humana, utilizando imagens da orelha.

Nos últimos anos, a orelha humana obteve um significativo interesse da área da biometria pelo motivo desta apresentar algumas vantagens quando comparadas a outras medidas biométri-

cas (KURNIAWAN; SHAFRY; RAHIM, 2012). O tamanho da orelha, por exemplo, é grande comparável à íris, retina e impressão digital e, portanto, pode ser capturado mais facilmente à distância. Além disso, sua forma é menor do que a face, o que permite menor custo de computação e, portanto, o torna adequado para aplicações em tempo real (ALVA; SRINIVASARAGHAVAN; SONAWANE, 2019).

Estudos médicos mostram que o crescimento da orelha nos primeiros quatro meses após o nascimento é proporcional, ou seja, ela cresce sem alterar sua forma, sendo assim uma ótima maneira de identificar pessoas (RAPOSO, 2012). Com isso, um grande interesse dos pesquisadores sobre a mesma tem sido despertado pelo fato desta apresentar características únicas. Na Figura 1.2 temos um exemplo da parte externa de uma orelha juntamente com o nome dos componentes que a constituem.



Figura 1.2 – Componentes da Orelha (UOL, 2015).

Nas últimas décadas, muitas pesquisas no campo de identificação biométrica mostraram o uso de imagens de orelha 2D ou modelos de orelha 3D para sistemas de reconhecimento de orelha (JAIN; KUMAR, 2012). Yaqubi, Faez e Motamed (2008) propuseram um sistema empregando recursos de borda assumidos em várias posições e orientações, onde as características extraídas foram classificadas usando Máquinas de Vetores de Suporte (*Support Vector Machine, SVM*) e K-Vizinhos Mais Próximos (*K-Nearest Neighbors, K-NN*). Utilizando uma versão modificada do Ponto Iterativo Mais Próximo (*Iterative Closest Point, ICP*), Yan e Bowyer (2007) desenvolveram um método capaz de realizar o reconhecimento de orelha em alguns cenários de oclusão por brinco e cabelo, alcançando uma precisão de 95.7%. Wang, Xia e Wang (2010) empregaram momentos invariantes para extração de características, na qual uma rede neural de retropropagação foi treinada para classificar os conjuntos de características.

Mediante à essas informações, é proposto neste trabalho métodos de reconhecimento biométrico de orelha no qual envolve a utilização de técnicas baseadas em *handcrafted features* e

em aprendizado profundo (*deep learning*) com o intuito de realizar uma comparação entre os mesmos. Desta forma, propõe-se utilizar técnicas de classificação convencional de imagens e arquiteturas de Redes Neurais Convolucionais disponíveis na literatura. Para a realização dos experimentos será utilizada a base de dados IIT Delhi Ear disponibilizada pelo *Indian Institute of Technology Delhi* e a base AMI Ear disponibilizada pela *Universidad de Las Palmas de Gran Canaria (ULPGC)*.

## 1.1 Justificativa

Devido a necessidade de proteger grandes volumes de informações geradas pelos sistemas computacionais bem como locais altamente sigilosos, procura-se desenvolver sistemas de segurança confiáveis já que informações pessoais e/ou industriais extremamente importantes não podem ser violadas. Para evitar os problemas enfrentados por métodos tradicionais, técnicas de biometria têm sido agregadas a estes sistemas, por estas serem em geral menos suscetíveis a falhas. Além disso, há um grande interesse em identificar criminosos com o máximo de precisão possível e com isso sistemas biométricos têm se mostrado muito eficientes.

É válido ressaltar que esta forma de identificação está cada vez mais inserida na sociedade, através de urnas eleitorais, *smartphones* e *notebooks* bem como caixas eletrônicos, que utilizam reconhecimento facial e impressão digital. Dessa forma, este trabalho visa realizar a identificação biométrica por meio de imagens da orelha, que por sua vez não apresenta problemas comparados a outros métodos de identificação biométrica, pelo fato da mesma sofrer alterações somente até os 4 primeiros meses de vida. Outro ponto positivo é que a orelha apresenta uma estrutura que não muda ao longo dos anos e não sofre das consequências do envelhecimento como rugas, desgaste e expressões faciais sendo que a captura de imagens de orelhas a distância pode ser feita facilmente com uma câmera comum, diferentemente de outras fontes biométricas como íris e a retina.

## 1.2 Objetivos Geral e Específicos

O objetivo principal deste trabalho é desenvolver métodos que sejam capazes de realizar a identificação humana através de imagens de orelha, realizando uma comparação entre *handcrafted features* e *Redes Neurais Convolucionais*. Desta forma, busca-se averiguar qual dos métodos possui o melhor desempenho para a identificação, apresentando as técnicas utilizadas para a elaboração dos mesmos bem como os resultados alcançados e as conclusões. Nesta perspectiva, os objetivos específicos para este trabalho são:

1. Implementar métodos *Handcrafted Features*;
2. Implementar métodos utilizando diferentes arquiteturas de *Redes Neurais Convolucionais*;

## **1.3 Organização do Trabalho**

O restante desta monografia está organizada da seguinte forma: no Capítulo 2 é apresentada uma sucinta revisão de literatura contendo os trabalhos diretamente relacionados ao tema de estudo, seguida pela fundamentação teórica descrita no Capítulo 3, ambos utilizados para a elaboração deste trabalho. O Capítulo 4 descreve detalhadamente a metodologia de trabalho dos métodos desenvolvidos, juntamente com a explicação das técnicas utilizadas. O Capítulo 5 apresenta os resultados alcançados pelo trabalho, seguido pelas conclusões e as expectativas de trabalhos futuros apresentadas no Capítulo 6.

## 2 Revisão de Literatura

O uso da orelha como uma medida biométrica pode fornecer uma medida muito boa para a identificação e verificação de pessoas. Bertillon (1890) apontou que as orelhas podem ser utilizadas para identificação criminal, pelo fato de possuírem uma estrutura rica e estável que é preservada desde o nascimento até a velhice, tendo certas vantagens sobre as demais conhecidas. Mais tarde, Lannarelli (1989) originou alguns trabalhos sobre biometria de orelha, proporcionando evidências suficientes para mostrar que as orelhas possuem características biométricas únicas.

O estudo feito por Burge e Burger (2000) demonstra que a variação da orelha ao longo do tempo é mais perceptível durante o período de quatro meses a oito anos e acima de 70 anos. Hurley, Carter e Nixon (2005) concluíram que a orelha possui menor variabilidade de expressões e tem uma distribuição de cor mais uniforme que as faces.

Nos últimos anos, grande quantidade de pesquisas foi conduzida para detecção e reconhecimento de orelhas em imagens 2D e 3D, resultando em promissoras técnicas de identificação biométrica. Portanto neste capítulo é apresentada uma breve revisão da literatura sobre reconhecimento biométrico através de imagens de orelha, utilizando descritores manuais e aprendizagem em profundidade.

### 2.1 Descritores Manuais

O problema deste trabalho, que é basicamente um processo de reconhecimento de imagem, é um problema clássico em áreas como processamento de imagens, visão computacional e reconhecimento de padrões. De maneira geral, esse processo consiste em identificar características, padrões ou estruturas relevantes em uma imagem, utilizando as mesmas para atribuir uma determinada classe para uma imagem (SOLOMON; BRECKON, 2011). Pode-se dizer que esta tarefa depende consideravelmente da escolha da técnica de extração de características quando se faz o uso de classificadores convencionais. Portanto, vários estudos foram feitos utilizando diferentes abordagens para extrair características relevantes das imagens de orelha.

Tariq, Anjum e Akram (2011) desenvolveram um sistema automatizado para o reconhecimento biométrico de orelha, utilizando-se técnicas de processamento de imagens digitais. A ideia abordada no trabalho é fazer primeiramente um pré-processamento das imagens, onde ocorre o recorte da orelha da imagem, redimensionamento para  $160 \times 120$  pixels, conversão para escala de cinza e aprimoramento do contraste da imagem em tons de cinza usando a equalização do histograma adaptativo limitado ao contraste. Em seguida, são extraídas as características através da transformada discreta de Wavelets de Haar de nível 2. Por fim, a classificação das imagens é feita usando correlação cruzada normalizada rápida (FNCC). Os resultados obtidos foram

satisfatórios, alcançando uma taxa de acurácia de 98,33% para a base de dados I da *University of Science Technology Beijing* (USTB), 96,10% para a base de dados II da USTB e 95,20% para a base de dados do *Indian Institute of Technology Delhi* (IITD).

Um conjunto de técnicas da área de Visão Computacional como pré-processamento, segmentação, detecção de bordas, extração de características e classificação foram utilizadas por [Anwara, Ghanyb e Elmahdy \(2015\)](#) para o desenvolvimento de um algoritmo capaz de fazer o reconhecimento da orelha. A princípio o algoritmo desenvolvido suaviza as imagens utilizando o filtro Gaussiano e faz o redimensionamento das mesmas. Em seguida, segmenta a orelha nas imagens utilizando um modelo de contorno ativo tradicional, também conhecido como *Snake* e aplica-se o filtro mediano para remover os ruídos. Posteriormente, as imagens são transformadas em binária usando um limiar global e as bordas são detectadas com o filtro de Canny, onde estas são conectadas e fechadas com algumas operações morfológicas. Para finalizar, sete valores são extraídos para compor o vetor de características, em que as características extraídas foram classificadas utilizando o algoritmo de vizinho mais próximo com distância de erro absoluto. É importante evidenciar que os resultados obtidos foram bem expressivos, resultando em uma taxa de acurácia de 98% para a base de dados do IITD.

Uma abordagem inovadora e eficiente é proposta no trabalho de [Hai-Long e Zhi-Chun \(2009\)](#), onde há uma combinação de dois algoritmos conhecidos como transformada de Wavelet Haar e *Orthogonal Centroid algorithm* (OC) para a extração de características das orelhas. Inicialmente, é realizado um pré-processamento das imagens, onde as imagens contidas nos bancos são redimensionadas para um tamanho padrão e têm seus histogramas equalizados. Em seguida, é aplicada a transformada de Wavelet Haar bidimensional (DWT), também conhecida como decomposição Wavelet 2D, com o intuito de obter sub-imagens de baixas frequências. Para finalizar, as características são extraídas através da aplicação do OC para as sub-imagens de baixa frequência e classificadas com a utilização do algoritmo K-NN. A métrica avaliada neste trabalho foi o tempo gasto para o reconhecimento da orelha, sendo comparado o tempo gasto pelo algoritmo proposto com outros existentes na literatura. Os testes foram realizados nas bases de dados I e II da USTB.

Um software em MATLAB que segue algumas técnicas da visão computacional foi desenvolvido por [Ali, Javed e Basit \(2007\)](#). Assim como outros trabalhos aqui já citados, primeiramente é realizado um pré-processamento das imagens, onde as imagens das orelhas são cortadas manualmente, redimensionadas para um tamanho padrão e convertidas para a escala de cinza a fim de evitar ruídos que podem dificultar no reconhecimento. Posteriormente, é aplicada a transformada de Wavelet Haar bidimensional (DWT) para a extração de características das orelhas. As características extraídas de cada imagem pela transformada são armazenadas em um vetor linha de tamanho 256 bytes, onde estas são usadas para treinar a base de dados. Por fim, as características da imagem de teste são extraídas e então para a classificação, as distâncias Euclidianas entre a imagem teste e os vetores de características das imagens treinadas são adquiridas,

sendo que a imagem testada é correspondente aquela que obteve a menor distância. A base de dados utilizada para os experimentos neste trabalho é da *University of Notre Dame (UND)*, onde os resultados alcançados foram significativos, com taxa de acerto de até 94.30%.

Os autores [Ghoulmi, Draa e Chikhi \(2016\)](#) propuseram uma nova técnica para normalização das imagens que mostrou ser eficaz em seus experimentos. Eles utilizaram um algoritmo da área de otimização conhecido como colônia de abelha artificial (ABC), juntamente com uma metaheurística, com o intuito de melhorar o contraste das imagens de orelha para que estas possam ser mais naturais e possuam um maior detalhamento de conteúdo. Para a extração de características das orelhas, o algoritmo SIFT foi utilizado para adquirir as características mais discriminantes e pertinentes das imagens de orelha. Finalmente, a distância Euclidiana classifica as imagens de teste. Em seus experimentos, realizaram uma comparação da abordagem proposta com abordagens tradicionais de realce de contraste de orelha, sendo elas: equalização de histograma (HE) e equalização adaptativa de histograma limitado ao contraste (CLAHE). Na base de dados II da USTB e na base do IITD, a abordagem proposta supera as outras no quesito de precisão de reconhecimento. Porém na base I da USTB, ela não supera a técnica de equalização de histograma (HE). Contudo, é válido ressaltar que os resultados obtidos nesse trabalho foram bem expressivos, alcançando uma taxa de acerto nas três bases de 97,15%, 94,79% e 99,60%.

O Método proposto por [Nosrati, Faez e Faradji \(2007\)](#) faz uma combinação de dois algoritmos conhecidos como Transformada de Wavelet Haar e *Análise do Componente Principal (PCA)* para a extração das características das orelhas. Preliminarmente, é realizado um pré-processamento das imagens, onde as imagens contidas nos bancos têm seus histogramas equalizados com o intuito de aliviar a variação de iluminação de uma imagem para outra. Em seguida, é utilizada a técnica de correspondência de modelos (do inglês, *template matching*) para a segmentação da orelha nas imagens, a fim de obter-se regiões de interesse para a extração das características. Posteriormente, é aplicada a transformada de Wavelet Haar bidimensional para obter três matrizes de características em três direções diferentes (horizontal, vertical e diagonal). Logo após, essas três matrizes de características são combinadas a fim de obter uma matriz de características que contenha maiores detalhes da orelha. Por fim, o algoritmo PCA é utilizado para reduzir a dimensionalidade das características e para classificar as mesmas. Os resultados obtidos pelos experimentos realizados nas bases de dados II da USTB e na base de dados disponibilizada por [Carreira-Perpinan \(1995\)](#) foram relevantes, com taxa de acerto de 90,5% e 95,05%, respectivamente.

No trabalho de [Damer e Fuhrer \(2012\)](#), diferente de outras propostas relatadas anteriormente, não foi necessária a etapa de pré-processamento, porque a base de dados do IIT utilizada possui as imagens de orelhas segmentadas e normalizadas. Inicialmente, as características das orelhas foram extraídas simultaneamente pelo histograma multi-escala de gradientes orientados (*Multi-Scale Histogram Of Oriented Gradients, MSHOG*) que simplesmente é um HOG de diferentes tamanhos de blocos e células que garantem a captura de estruturas diferentes e complicadas

das orelhas. Em seguida, um vetor contendo as características do MSHOG foi construído para descrever cada imagem da orelha. Para finalizar, foi utilizado o algoritmo PCA para reduzir a dimensionalidade das características e o algoritmo K-NN com distância Euclidiana para a classificação das mesmas. A taxa de acerto alcançada foi de 92%.

Uma nova abordagem para o reconhecimento de orelha, com a presença de oclusão parcial por cabelo e/ou brincos, foi desenvolvida por Othman, Alizadeh e Sutherland (2018). A ideia geral do trabalho é fazer primeiro um pré-processamento das imagens, onde as imagens das orelhas são redimensionadas para  $240 \times 320$  pixels, convertidas para a escala de cinza e equalizadas pela técnica de equalização de histograma, a fim de obter imagens com maiores contrastes e evitar ruído que podem dificultar no reconhecimento. Em seguida, foi aplicado o algoritmo desenvolvido por Viola e Jones (2001) para que a detecção da orelha seja em tempo real e automática, onde este é um algoritmo robusto e eficiente em detecção de objetos. Logo após, foi aplicado o detector de borda Canny para encontrar os pontos de borda da imagem da orelha para que as características da mesma possam ser extraídas. Posteriormente, o Contexto da Forma foi utilizado como o descritor para que as características sejam extraídas, sendo que este cria um histograma log-polar. Para finalizar, para classificar a imagem desconhecida, a matriz de dissimilaridade entre o histograma da imagem desconhecida da orelha e os histogramas das imagens conhecidas do banco de dados é calculada, fazendo uma correspondência ponto a ponto nos histogramas para encontrar a maior similaridade. A avaliação experimental contida neste trabalho abordaram 4 cenários diferentes sendo eles: orelha sem oclusão, orelha com oclusão de brinco, orelha com oclusão de cabelo e orelha com oclusão de brinco e cabelo. A taxa de acerto nestes cenários são respectivamente, 100%, 80%, 73,3% e 57,1%.

Padrão Binário Local (Local Binary Pattern, LBP) foi utilizado para a extração de características na proposta elaborada por Boodoo-Jahangeer e Baichoo (2013). Basicamente, este algoritmo divide a imagem da orelha em células que possuem blocos de tamanho  $3 \times 3$ . O pixel central desses blocos são comparados com os demais vizinhos, atribuindo-se 0 aos valores do vizinhos inferiores e 1 aos superiores, gerando assim uma matriz de 0 e 1s. Posteriormente, essa matriz é multiplicada por uma outra matriz que contém os pesos dados aos pixels correspondentes e os valores dos oito pixels são somados para obter o número de LBP para o pixel central. Para finalizar, é calculado o histograma sobre cada célula da imagem e ao obter todos os histogramas da imagem, esses são concatenados para fornecer o vetor de características a ser usado para a classificação. Por fim, o histograma do vetor de características gerado é comparado com os histogramas treinados da base de dados, fazendo uma correspondência para fazer o reconhecimento. Os experimentos foram realizados na base de dados do IITD e obteve uma taxa de reconhecimento de 93%, superando o algoritmo PCA também testado na base, que alcançou uma taxa de reconhecimento de apenas 85%.

Um resumo sucinto dos trabalhos citados anteriormente, incluindo as técnicas utilizadas, bancos de dados, parâmetros de avaliação de desempenho é apresentado na Tabela 2.1.

Tabela 2.1 – Comparação das técnicas recentes em reconhecimento biométrico de orelha.

Autores	Técnicas	Base de Dados	Parâmetros
Tariq, Anjum e Akram (2011)	<b>Pré-Processamento:</b> redimensionamento, escala de cinza e CLAHE. <b>Extração de Características:</b> DWT. <b>Classificação:</b> FNCC.	Conjunto I e II da USTB e IITD Ear.	<b>Acurácia</b> - 98,33%, 96,10% e 95,20%.
Anwara, Ghanyb e Elmahdy (2015)	<b>Pré-Processamento:</b> redimensionamento e filtragem gaussiana. <b>Segmentação:</b> Snake. <b>Extração de Características:</b> Canny + algoritmo dos autores. <b>Classificação:</b> K-NN.	IITD Ear.	<b>Acurácia</b> - 98%.
Hai-Long e Zhi-Chun (2009)	<b>Pré-Processamento:</b> redimensionamento e HE. <b>Extração de Características:</b> DWT + OC. <b>Classificação:</b> K-NN.	Conjunto I e II da USTB.	<b>Tempo Gasto.</b>
Ali, Javed e Basit (2007)	<b>Segmentação:</b> Manual. <b>Pré-Processamento:</b> redimensionamento e escala de cinza. <b>Extração de Características:</b> DWT. <b>Classificação:</b> Distância Euclidiana.	Conjunto da UND.	<b>Acurácia</b> - 94,30%.
Ghoulmi, Draa e Chikhi (2016)	<b>Pré-Processamento:</b> Melhoria de contraste com ABC. <b>Extração de Características:</b> SIFT. <b>Classificação:</b> Distância Euclidiana.	Conjunto I e II da USTB e IITD Ear.	<b>Acurácia</b> - 97,15%, 99,10% e 94,79%.
Nosrati, Faez e Faradji (2007)	<b>Pré-Processamento:</b> HE. <b>Segmentação:</b> Template Matching. <b>Extração de Características:</b> DWT. <b>Classificação:</b> PCA.	Conjunto II da USTB e Carreira-Perpinan (1995).	<b>Acurácia</b> - 90,5% e 95,05%.
Damer e Fuhrer (2012)	<b>Extração de Características:</b> MSHOG + PCA. <b>Classificação:</b> K-NN.	IITD Ear.	<b>Acurácia</b> - 92%.
Othman, Alizadeh e Sutherland (2018)	<b>Pré-Processamento:</b> redimensionamento, escala de cinza e HE. <b>Deteção:</b> algoritmo de Viola e Jones (2001). <b>Extração de Características:</b> Canny + Contexto de Forma. <b>Classificação:</b> Correspondência na matriz de dissimilaridade.	Conjunto I e II da USTB e IITD Ear.	<b>Acurácia</b> - 100,00%, 80,00% 73,30% e 57,1%.
Boodoo-Jahangeer e Baichoo (2013)	<b>Extração de Características:</b> LBP. <b>Classificação:</b> Correspondência entre histogramas.	IITD Ear.	<b>Acurácia</b> - 93%.

## 2.2 Aprendizagem em Profundidade

O aprendizado profundo (do inglês, deep learning) se tornou um dos pilares da pesquisa moderna na área de visão computacional, propondo uma arquitetura de várias camadas de aprendizado (JUNIOR, 2018). Segundo Lecun, Bengio e Hinton (2015), tal abordagem permite que modelos computacionais compostos de múltiplas camadas de processamento aprendam representações de dados com vários níveis de abstração, sendo que, ao final do processo de aprendizado têm-se uma função capaz de receber por entrada os dados brutos e fornecer como saída uma representação adequada para o problema em questão (PONTI; COSTA, 2017). Portanto, pode-se dizer que o aprendizado profundo é baseado no conceito de redes neurais artificiais, ou seja, sistemas computacionais que imitam a maneira como o cérebro humano funciona (LECUN; BENGIO; HINTON, 2015).

Devido ao sucesso das redes neurais na aprendizagem e classificação de características, muitos pesquisadores adotaram essa abordagem para reconhecimento de orelha (ALVA; SRINIVASARAGHAVAN; SONAWANE, 2019). Tian e Mu (2016) propuseram uma arquitetura de rede neural convolucional com três camadas convolucionais, três camadas de pooling e duas camadas totalmente conectadas para o reconhecimento automático de imagens de orelhas. Para melhorar o desempenho do reconhecimento e evitar ajustes excessivos da rede, algumas estratégias foram utilizadas para aumentar o conjunto de dados. Os resultados experimentais na base de dados III da USTB indicam que método proposto alcança uma alta acurácia e supera alguns métodos tradicionais, perfazendo uma taxa média de 98,27%.

O sistema desenvolvido por Chowdhury, Islam e Gao (2017) mostra-se robusto e eficiente à variações de pose, oclusão e mudanças de iluminação para o reconhecimento da orelha. Os autores utilizaram um algoritmo detector baseado em Adaboost desenvolvido por Liu e Liu (2010) para detectar e segmentar as orelhas, em que estes afirmam que a motivação para sua escolha é pela alta precisão e velocidade. Após a segmentação, as imagens são redimensionadas e filtradas por um filtro conhecido como difuso, onde este emprega regras difusas para decidir o nível de cinza dos pixels dentro de uma janela na imagem. Posteriormente, as características das orelhas são extraídas utilizando-se o detector de borda de Canny, resultando em características locais das bordas da orelhas. Para finalizar, as características locais são alimentadas em uma rede neural desenvolvida pelos autores para fins de treinamento e reconhecimento. Através dos experimentos realizados, pode-se observar que o método proposto alcança um desempenho superior, comparável a outros métodos similares existentes alcançando uma taxa de reconhecimento de 98,20% para a base de dados da UND.

Uma combinação de filtragem de ruído, extração de características e reconhecimento baseado em rede neural são utilizadas por Alaraj, Hou e Fukami (2010) para identificação humana com imagens de orelhas. Inicialmente, é realizado um pré-processamento das imagens, onde as imagens são redimensionadas e filtradas pela técnica de filtragem mediana para eliminar os ruídos. Em seguida, os conjuntos de dados são divididos treinamento e teste, e o algoritmo PCA

é então aplicado para obter as principais características de cada imagem presentes nestes. Por fim, a classificação é realizada por meio da rede neural de avanço multicamada (*Multilayer Feed Forward Neural Network, MFFNN*). A base de dados utilizada para os experimentos neste trabalho é da UND, obtendo uma taxa média de acurácia de 96,1%.

Um sistema de reconhecimento automático da orelha utilizando a transformada de Wavelet Haar e uma rede neural de retropropagação (*Back-Propagation Neural Network, BPNN*) foi desenvolvido por [Daramola e Oluwaninyo \(2011\)](#). A imagem da orelha de entrada é decomposta em quatro sub-bandas usando a transformação de wavelet de Haar, onde as características são extraídas de cada uma das sub-bandas detalhadas. Por fim, as características de cada sub-bandas são unidas em um vetor e estas são usadas como entrada para rede neural para classificação eficaz da imagem da orelha. O sistema proposto foi testado usando a base de dados própria dos autores, a qual esta contém 1050 imagens de orelha coletadas de 350 pessoas. Resultados experimentais demonstraram a eficácia do sistema proposto, alcançando uma taxa média de acurácia de 97,42%.

Com o intuito de tornar mais fácil e rápido o processo de reconhecimento da orelha como biométrico, uma arquitetura de rede neural modular foi desenvolvida por [Gutierrez, Melin e Lopez \(2010\)](#). Para melhorar o desempenho de reconhecimento da orelha e fazer uma comparação com outros métodos existentes, foi utilizado a transformada de Wavelets Haar bidimensional para extração de características para alimentar a rede neural, sendo esta responsável pela classificação. Os experimentos foram realizados na base de dados II da USTB e mostram que método proposto alcançam uma alta acurácia, atingindo uma taxa média de 97%.

Uma nova abordagem de reconhecimento de orelha, incluindo um nova técnica extração de características é proposta por [Guo e Xu \(2008\)](#). A princípio uma rede neural celular é adotada para pré-processar imagens, cuja função é eliminar as informações irrelevantes presentes nas mesmas. Em seguida, o método proposto chamado de padrão binário de similaridade local (*Local Similarity Binary Pattern, LSBP*) juntamente com o LBP tradicional são utilizados para a etapa de extração de características. Por fim, a classificação das imagens é realizada empregando o uso do algoritmo K-NN. A partir dos experimentos conduzidos no banco de dados II da USTB é possível que a abordagem proposta supera alguns outros métodos conhecidos em termos da taxa de reconhecimento, alcançando uma média de 93,80%.

[Mayya e Saii \(2016\)](#) desenvolveram um eficiente sistema para o reconhecimento biométrico de orelha, sob diferentes condições de iluminação e pose. Primeiramente, as imagens da orelha foram obtidas por meio de um detector de pele de probabilidade desenvolvido por [Conaire, O'Connor e Smeaton \(2007\)](#) e algumas operações morfológicas. Posteriormente, é aplicada a transformada de Wavelet Haar bidimensional para a extração de características, onde é utilizado o PCA para reduzir a dimensionalidade das mesmas. Finalmente, a rede neural BPNN é treinada usando a características para realizar a classificação. Foram realizados experimentos para avaliar a eficiência e a robustez do sistema usando o banco de dados de orelhas feito pelos autores. O banco de dados contém 460 imagens de faces laterais correspondentes a 55 pessoas de 12 a 60

anos de idade. Os resultados mostram que o sistema atingiu uma taxa média de 96,73% para a segmentação da orelha e 98,90% para o reconhecimento.

Zhang et al. (2005) propuseram um sistema híbrido e robusto para a classificação de imagens de orelha empregando o uso de uma rede neural RBF. Inicialmente, as bases de dados utilizadas são decompostas em combinações lineares de várias imagens para aumentar a quantidade de imagens contidas nestas. Posteriormente, a extração de características é realizado por meio do algoritmo ICA, na qual estas são alimentadas em uma rede neural RBF que é responsável pela classificação das mesmas. Os testes foram realizados nas bases de dados I da USTB e na base do Carreira-Perpinan (1995), alcançando uma taxa média de acurácia de 90% e 94,11%, respectivamente.

Assim como Tian e Mu (2016), Galdamez, Raveane e Gonzalez (2016) propuseram um sistema envolvendo redes neurais convolucionais (CNN) para o reconhecimento automático de imagens de orelhas. A rede neural desenvolvida é responsável pela tarefa de extração e classificação das características. Para os autores, em comparação com outros métodos de visão computacional, as CNNs são muito mais robustas e tolerantes a formas e variações visuais das imagens ou objetos destinados a serem reconhecidos. Os experimentos foram realizados na base de dados própria dos autores, alcançando uma média de precisão de 97,71%.

A Tabela 2.2 mostra uma breve comparação entre os trabalhos, incluindo as técnicas utilizadas, bancos de dados, parâmetros de avaliação de desempenho.

Tabela 2.2 – Comparação das técnicas recentes em reconhecimento biométrico de orelha com o uso de aprendizado profundo.

Autores	Técnicas	Base de Dados	Parâmetros
Tian e Mu (2016)	<b>Pré-Processamento:</b> Aumento de dados. <b>Extração de Características e Classificação</b> realizadas pela CNN.	Conjunto III da USTB.	<b>Acurácia</b> - 98,27%.
Chowdhury, Islam e Gao (2017)	<b>Pré-Processamento:</b> redimensionamento e filtragem difusa. <b>Segmentação:</b> Adaboost. <b>Extração de Características:</b> Canny + algoritmo dos autores. <b>Classificação:</b> Rede Neural proposta pelos autores.	Conjunto da UND.	<b>Acurácia</b> - 98,20%.
Alaraj, Hou e Fukami (2010)	<b>Pré-Processamento:</b> redimensionamento e filtragem mediana. <b>Extração de Características:</b> PCA. <b>Classificação:</b> MFFNN.	Conjunto da UND.	<b>Acurácia:</b> 96,1%.
Daramola e Oluwaninyo (2011)	<b>Extração de Características:</b> DWT. <b>Classificação:</b> BPNN.	Própria dos autores.	<b>Acurácia</b> - 97,42%.
Gutierrez, Melin e Lopez (2010)	<b>Extração de Características:</b> DWT. <b>Classificação:</b> Rede Neural Modular.	Conjunto II da USTB.	<b>Acurácia</b> - 97%.
Guo e Xu (2008)	<b>Pré-Processamento:</b> Rede Neural Celular. <b>Extração de Características:</b> LBSP + LBP. <b>Classificação:</b> K-NN.	Conjunto II da USTB.	<b>Acurácia</b> - 93,80%.
Mayya e Saii (2016)	<b>Segmentação:</b> Detector de pele com probabilidade + operações morfológicas. <b>Extração de Características:</b> DWT + PCA. <b>Classificação:</b> BPNN	Própria dos autores.	<b>Acurácia da Segmentação</b> - 96,73% e <b>Acurácia do Reconhecimento</b> 98,90%.
Zhang et al. (2005)	<b>Extração de Características:</b> ICA. <b>Classificação:</b> Rede Neural RBF.	Conjunto I da USTB e Conjunto do (CARREIRA-PERPINAN, 1995).	<b>Acurácia</b> - 90% e 94,11%.
Galdamez, Ravane e Gonzalez (2016)	<b>Extração de Características e Classificação</b> realizadas pela CNN.	Própria dos autores.	<b>Precisão</b> - 97,71%.

## 3 Fundamentação Teórica

Neste capítulo é apresentada toda uma fundamentação teórica que visa explicar as estratégias aplicadas para a produção deste trabalho, bem como alguns conceitos imprescindíveis para entendimento da pesquisa desenvolvida.

### 3.1 Segmentação

A segmentação é uma das principais e mais estudadas operações em processamento de imagens, onde esta consiste em dividir a imagem em regiões que dizem respeito ao mesmo conteúdo e aplicação (FONSECA, 1998). Ela subdivide a imagem em regiões/objetos que são de real interesse para a aplicação, sendo extremamente importante para que o reconhecimento/interpretação da imagem seja eficiente.

Podemos dizer que os algoritmos de segmentação geralmente seguem duas estratégias genéricas baseadas nos níveis de cinza, afirma Sobral (2002), sendo elas:

- **Descontinuidade:** a partição da imagem é efetuada com base nas alterações bruscas de intensidade. Temos como exemplo a detecção de bordas;
- **Similaridade:** a partição é efetuada com base na similaridade entre *pixels*, seguindo um determinado critério. Temos como exemplos: binarização, crescimento de regiões, divisão e junção de regiões.

A construção de algoritmos com o intuito de automatizar processos de segmentação é considerada como sendo uma das tarefas mais difíceis em processamento de imagens (GONZALEZ; WOODS, 2006). Problemas como variação de iluminação, ambientes com o fundo muito complexo e objetos que ocluem a região interessada dificultam que os algoritmos viabilizem de maneira eficiente a segmentação.

Lei, You e Abdel-Mottaleb (2016) relatam que os maiores desafios para se elaborar um algoritmo que seja eficiente para segmentação da orelha estão relacionados a curvatura da orelha, por ser uma superfície 3D altamente curva gerando assim reflexos desiguais, bem como pelo e pele com pigmentação variada. A Figura 3.1 apresentada abaixo mostra um exemplo de segmentação da imagem de orelha proposto no trabalho de Nosrati, Faez e Faradji (2007), na qual foi utilizado a técnica de correspondência de modelos (do inglês, *template matching*) para a tarefa.

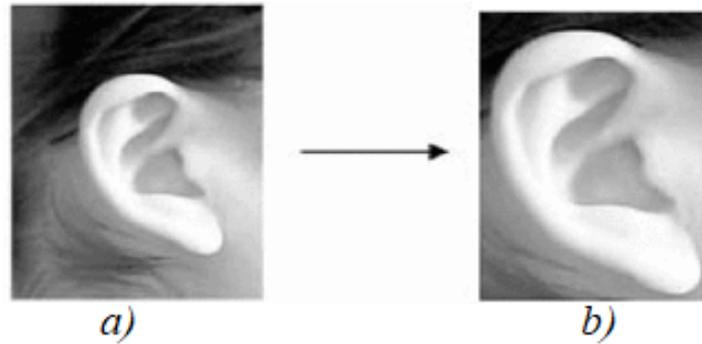


Figura 3.1 – Exemplo da segmentação de uma imagem da orelha (NOSRATI; FAEZ; FARADJI, 2007) - a) Imagem original e b) Segmentação com correspondência de modelos.

## 3.2 Extração de Características

Como sabemos os objetos ou segmentos são representados como uma coleção de *pixels* em uma imagem, possuindo forma, cor e textura. Estas propriedades podem ser mensuradas e tais medidas são denominadas características do objeto. Para extrairmos características específicas de um objeto com o intuito de reduzir a complexidade do mesmo, utilizamos o que denominamos descritores que normalmente são representados por um vetor de escalares. É importante que os descritores sejam invariantes a transformações afins como rotação, escala e posição para que os mesmos sejam robustos e eficientes nas descrições de características. Nesta perspectiva, será utilizado neste trabalho o descritor Histograma de Gradientes Orientados (*Histogram of Oriented Gradients, HOG*) para a extração de características devido ao seu amplo uso em detecção de objetos e reconhecimento de seres humanos (STALLKAMP et al., 2012).

O algoritmo HOG desenvolvido por Dalal e Triggs (2005) é um descritor que calcula os histogramas das orientações dos gradientes na imagem. O descritor final é um vetor de histogramas extraídos da imagem. A Figura 3.2 demonstra a imagem de entrada e o resultado da aplicação do HOG sobre a mesma.

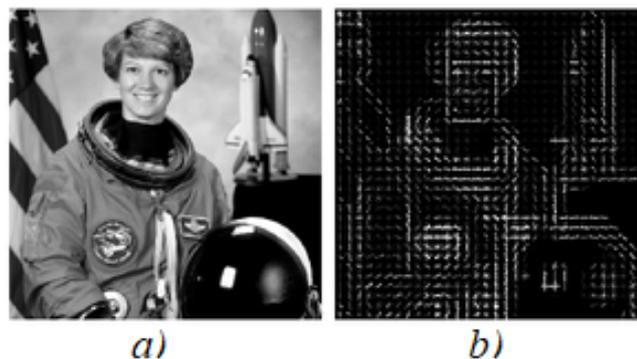


Figura 3.2 – Exemplo de aplicação do HOG (HYPRAPTIVE, 2017). a) Imagem de entrada e b) Imagem resultante.

O algoritmo se baseia na ideia de que a forma e a aparência de um objeto podem ser descritas muitas vezes pela intensidade dos gradientes ou a direção das bordas, sem um conhecimento prévio da posição de tais bordas. O processo para gerar o descritor pode ser dividido em quatro etapas: cálculo do gradiente em cada *pixel*, agrupamento dos *pixels* em células, agrupamento das células em blocos e obtenção do descritor.

O primeiro passo é a computação dos gradientes, um estágio crítico na formação do descritor. Dada uma imagem  $I$  são calculados os gradientes  $G_x$  e  $G_y$  em cada *pixel* de acordo com máscaras derivada discreta pontual tanto no eixo vertical como horizontal.

O passo seguinte é responsável por agrupar os *pixels* em regiões pequenas conectadas, criando-se o que se chama de células (por exemplo, de tamanho 8 *pixels*). Após a segunda etapa, os blocos são criados através do agrupamento de células de uma certa região.

Na etapa final, cria-se o descritor que nada mais é do que uma lista dos histogramas de todas as células de todos os blocos. A atenuação do problema das variações locais de iluminação ou de contraste entre o primeiro plano e o plano de fundo, se dá através da normalização de cada histograma de acordo com seus próprios valores. A Figura 3.3 ilustra o passo a passo da obtenção do descritor HOG.

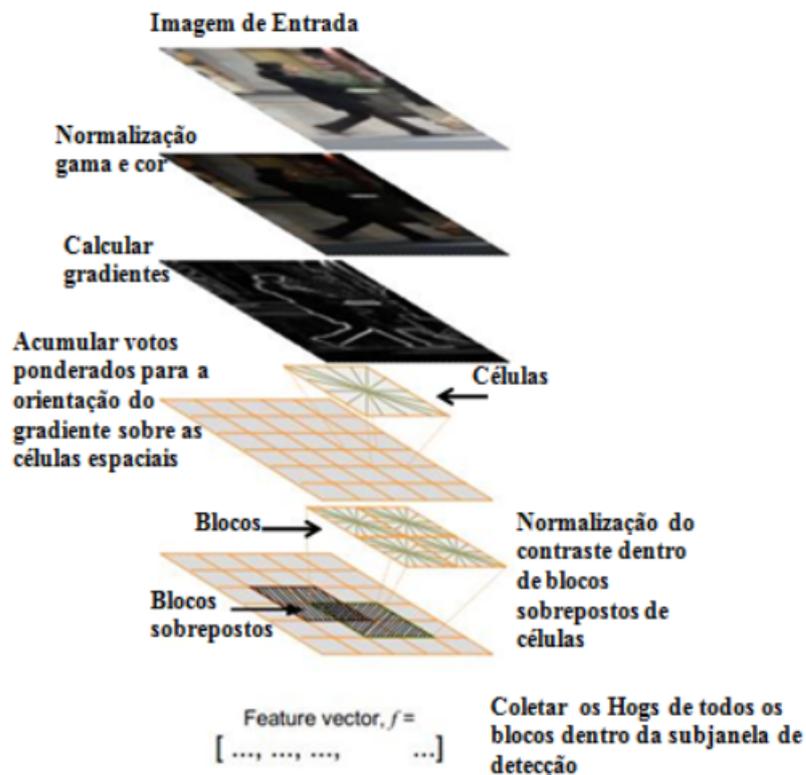


Figura 3.3 – Etapas do HOG (CERNA, 2014).

## 3.3 Classificação

Nesta seção, têm-se três subseções que serão retratadas a seguir, sendo elas, aprendizado de máquina (AM), classificadores convencionais e redes neurais convolucionais (CNN). Na primeira divisão temos uma sucinta explicação sobre conceitos referentes ao aprendizado de máquina bem como a classificação em si, que é considerada como uma técnica de AM. A subseção seguinte é composta por uma breve explicação sobre os classificadores convencionais escolhidos para serem combinados com o HOG. Para finalizar, na última subseção têm-se uma breve introdução sobre o funcionamento das redes neurais convolucionais bem como uma concisa explanação sobre as arquiteturas escolhidas para a classificação direta das imagens de orelha, isto é, sem a necessidade de descritores.

### 3.3.1 Aprendizado de Máquina

Segundo Ratsch (2014), visando a criação de sistemas computacionais inteligentes que sejam capazes de obter conhecimentos de forma automática, foi criado o aprendizado de máquina que pode ser considerado como uma área de estudo em expansão na Inteligência Artificial. De acordo com Mitchell (2006), os estudos teóricos nessa área tem permitido que novas técnicas e eficientes algoritmos sejam desenvolvidos para as variadas tarefas de aprendizado, porém ainda não é possível fazer com que os computadores aprendam tão bem quanto as pessoas.

As técnicas de aprendizado de máquina existentes nos dias de hoje empregam um princípio de inferência denominado indução, no qual obtém-se conclusões genéricas a partir de um conjunto particular de exemplos conforme descreve Lorena e Carvalho (2007). Pistori (2003) afirma que geralmente estas técnicas são empreendidas em problemas que não podem ser resolvidos por meio de métodos tradicionais de programação, tais como os algoritmos imperativos, funcionais ou orientados a objetos. Lorena e Carvalho (2007) retratam ainda que, conforme ilustrado na Figura 3.4, o aprendizado indutivo pode ser dividido em duas principais categorias: aprendizagem supervisionada e não-supervisionada.

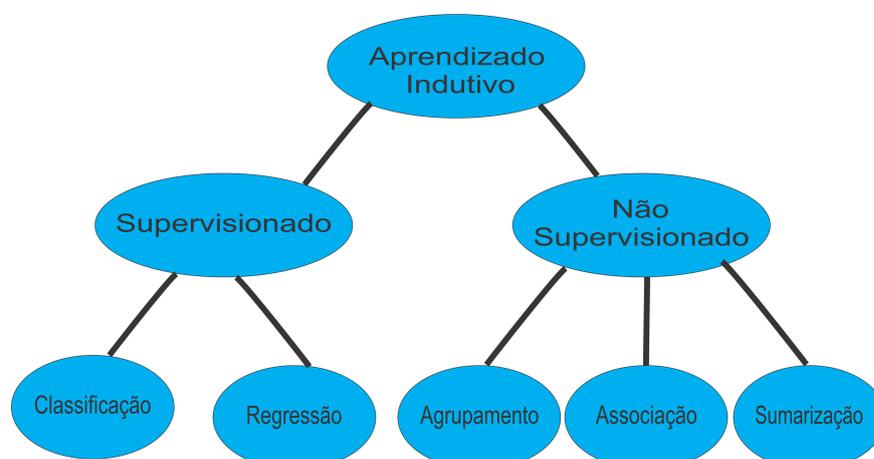


Figura 3.4 – A Hierarquia do aprendizado (MONTEIRO, 2016).

No aprendizado supervisionado temos a figura de um professor supervisor [Haykin \(1998\)](#), na qual este demonstra a ideia do ambiente por meio de um conjunto de dados na configuração entrada e saída desejada. Portanto, o objetivo do algoritmo é determinar um modelo que seja capaz de classificar corretamente exemplos não rotulados recebidos como entrada ([FERREIRA, 2016](#)).

Por outro lado, o aprendizado não-supervisionado que por vezes é conhecido como modelagem descritiva, não utiliza informações das variáveis de saída. Os dados de entradas são analisados e agrupados conforme a proximidade dos seus valores. Para cada grupo, ou cluster, um rótulo é utilizado, permitindo indicar a qual grupo cada registro pertence ([ROZA, 2016](#)).

Atualmente, a classificação é um dos assuntos mais abordados na área de aprendizado de máquina tornando-se presente em variados problemas reais, como por exemplo, reconhecer padrões em imagens ([AGGARWAL, 2014](#)). Segundo [Steinbruch \(2006\)](#) a finalidade de um algoritmo de aprendizado na etapa de classificação é determinar de maneira correta a qual classe pertence um determinado objeto ainda não rotulado, mediante conjuntos de classes e exemplos de treinamento. Sendo assim, classificadores convencionais encontrados na literatura como *K-Vizinhos mais próximos (K-NN)*, *Análise Discriminante Linear (LDA)*, *Máquinas de Vetores de Suporte (SVM)*, e diferentes arquiteturas de Redes Neurais Convolucionais também presentes na literatura (*AlexNet*, *SqueezeNet* e *GoogleNet*) foram escolhidos neste trabalho pelo fato de serem bem difundidos no reconhecimento de padrões.

### 3.3.2 Classificadores Convencionais

Existem na literatura diversos trabalhos que usam classificadores dos mais variados tipos para problemas presentes na área de Visão Computacional. Dentre as diversas técnicas existentes, os algoritmos (K-NN, LDA e SVM) têm sido amplamente usados na literatura, sendo estas representações populares de diferentes paradigmas de aprendizado ([WITTEN; FRANK; HALL, 2011](#)), o que justifica a utilização destes neste trabalho.

#### 3.3.2.1 Análise Discriminante Linear

A Análise Discriminante Linear (LDA) proposta por [Fisher \(1936\)](#) é uma técnica frequentemente utilizada para a redução de dimensionalidade de características e classificação de dados ([BALAKRISHNAMA; GANAPATHIRAJU, 1998](#)). A ideia básica da LDA envolve a obtenção de uma transformação linear que diferencie as classes o máximo possível, onde a classificação é submetida em um espaço transformado baseado em alguma métrica, como por exemplo a distância euclidiana ([SOUSA, 2016](#)). Ela fornece a separabilidade das classes desenhando uma região de decisão entre as diferentes classes, onde essa região é conhecida como hiperplanos, sendo que a separação por hiperplano é obtida buscando projeções que maximizam a distância entre as classes ([FUKUNAGA, 1990](#)).

O algoritmo LDA tem uma baixa exigência computacional, o que o torna adequado para sistemas em que o tempo para a classificação é importante (LOTTE et al., 2018), além disso, esta técnica de classificação é simples de usar e apresenta bom desempenho em um grande conjunto de aplicações, sendo considerado um dos classificadores mais populares (THEODORIDIS; KOUTROUMBAS, 2008) por permitir uma previsão direta do grupo que a variável pertence (SAMPAIO, 2006).

### 3.3.2.2 K-Vizinhos Mais Próximos

O K-Vizinhos Mais Próximos (K-NN) proposto por Fukunaga e Narendra (1975) é um dos algoritmos mais simples voltados para classificação, que possui como base a ideia de agrupar objetos com características semelhantes (JUNIOR, 2018). Ele tem sido usado numa extensa variedade de aplicações como finanças, saúde, ciência política, detecção de manuscritos, reconhecimento de imagem e reconhecimento de vídeo (SHAKHNAROVICH; DARRELL; INDYK, 2006). A ideia principal deste é determinar o rótulo de classificação de uma amostra baseado nas amostras vizinhas advindas de um conjunto de treinamento (SHAKHNAROVICH; DARRELL; INDYK, 2006).

De maneira geral, o funcionamento deste método de classificação consiste em calcular a distância entre a amostra desconhecida e as amostras contidas na base de treinamento com o intuito de identificar e selecionar os  $K$  vizinhos mais próximos, onde  $K$  é um valor definido pelo usuário. Os rótulos das classes pertencentes aos vizinhos mais próximos são utilizadas para determinar o rótulo da amostra desconhecida, onde a classe mais comum entre eles é atribuída a amostra (LIMA, 2012). Esse processo de eleger a classe mais comum é conhecido como procedimento de votação majoritária (JUNIOR, 2018). Para exemplificar o funcionamento do K-NN, a Figura 3.5 apresenta um problema de classificação com dois rótulos de classe, sendo eles  $A$  e  $B$ . No exemplo, as amostras contidas na base de treinamento são representadas pelas bolinhas azuis e amarelas, e uma nova amostra é representada por uma estrela vermelha. Como o valor de  $K$  é igual a 7, os sete vizinhos mais próximos são encontrados com base na distâncias aferidas entre as amostras de treino e a amostra desconhecida. Com isso, das sete amostras de treinamento mais próximas da nova amostra, 4 são do rótulo  $A$  e 3 do rótulo  $B$ . Portanto, como existem mais vizinhos do rótulo  $A$ , a nova amostra receberá o mesmo rótulo deles, ou seja,  $A$ .

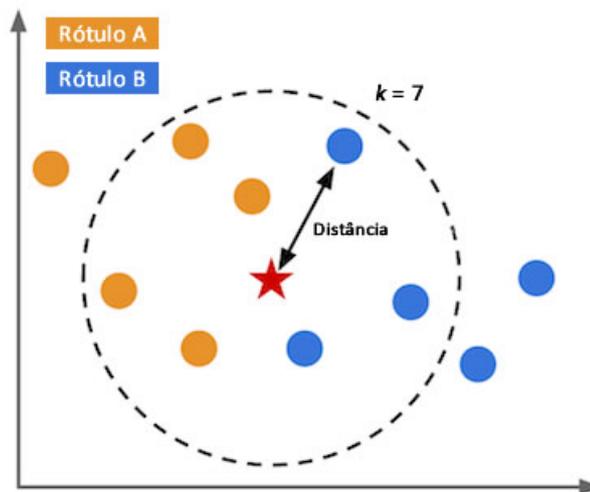


Figura 3.5 – Exemplo de classificação do KNN com dois rótulos de classe e  $k = 7$  (PACHECO, 2019).

### 3.3.2.3 Máquinas de Vetores de Suporte

As Máquinas de Vetores de Suporte (SVMs, do inglês *Support Vector Machines*) instituem uma técnica de aprendizado que vem recebendo crescente atenção da comunidade de AM (MITCHELL, 1997). Cerna (2014) discorre que estas são utilizadas na literatura na fase de classificação dos dados de treino, entretanto, a falta do conhecimento sobre a teoria e o funcionamento da mesma pelo usuário pode acarretar em uma insatisfatória classificação dos dados.

As SVMs mapeiam as características do conjunto de dados de entrada a um espaço dimensional, representados como pontos (CORTES; VAPNIK, 1995). Portanto, sempre que temos duas classes, podemos escolher uma fronteira de separação se os dados forem linearmente separáveis. Analisando a Figura 3.6, suponha que a fronteira escolhida é *A*. Como ela está bem próxima da classe azul, seu poder de generalização é baixo. Já escolhendo a fronteira *B*, podemos notar que o poder de generalização é bem melhor.

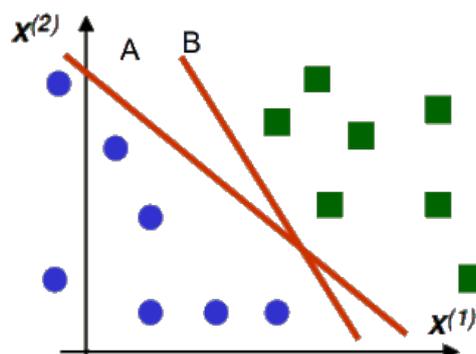


Figura 3.6 – Exemplo de escolha de fronteiras de separação (MENOTTI, 2011).

A ideia geral da SVM, basicamente falando, refere-se a maximização da margem, ou seja, maximizar a distância da margem dos dados de treinamento da reta. Neste contexto, um hiperplano ótimo seria aquele no qual a distância da margem para o exemplo da classe positiva é igual a distância da margem para o exemplo da classe negativa.

Os vetores de suporte são formados pelos exemplos da base de treinamento mais próximos do hiperplano. O hiperplano é definido unicamente pelos vetores de suporte, os quais são encontrados durante o treinamento. No entanto, a grande maioria dos problemas reais não são linearmente separáveis.

Tendo em vista que os problemas reais não costumam ser linearmente separáveis faz-se necessário projetar os dados em outra dimensão usando uma função de transformação (kernel). Essa função basicamente busca encontrar um hiperplano que separe os dados dentro de um espaço de características. Essa transformação mapeia o espaço original dos padrões para um novo espaço de atributos  $m$ -dimensional e nesse novo espaço os padrões passam a ser linearmente separáveis, como ilustra a Figura 3.7. Com a função de transformação o problema de otimização recai para um SVM Linear.

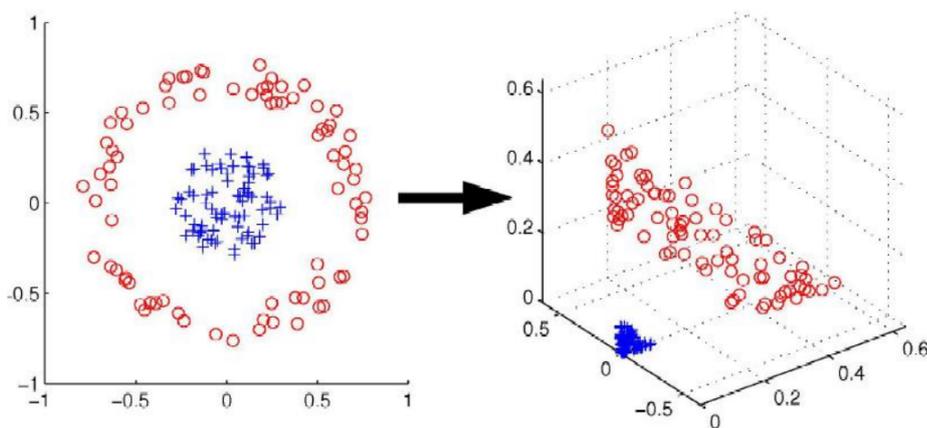


Figura 3.7 – Exemplo de mapeamento bidimensional para tridimensional (MENOTTI, 2011).

No entanto, SVMs são classificadores binários. Sendo assim, em casos como o deste trabalho faz-se necessário o uso de múltiplos SVMs binários para construir um classificador multiclases, ou seja, o problema de classificação maior é subdividido em vários problemas binários. Neste contexto existem diversas abordagens para realizar esta tarefa, sendo mais utilizadas as estratégias *um-contra-todos* e *um-contra-um* (BONESSO, 2013). De maneira geral, a primeira abordagem consiste em particionar um problema de  $n$  classes em dois grupos, sendo um grupo formado por uma classe (positiva) e o outro formado pelas classes restantes (negativas). Com isso, um classificador binário é treinado para rotular esses dois grupos e este procedimento é repetido para cada uma das  $n$  classes, criando-se então  $n$  SVMs (LORENA; CARVALHO, 2007). Na segunda abordagem, ao contrário da primeira, um classificador é treinado para cada par de classes, o que corresponde um total de  $K(K-1)/2$  classificadores (SOUSA, 2016). Como o número de

classificadores gerados nesta metodologia em geral é grande, a avaliação de seu resultado pode ser lenta (LORENA; CARVALHO, 2007).

### 3.3.3 Redes Neurais Convolucionais

Uma Rede Neural Convolucional pode ser considerada como sendo um tipo de rede neural artificial variante do perceptron de múltiplas camadas, que foi inspirada no processo biológico contido no córtex visual dos animais (SOLAZZO et al., 2016). Ela é um dos algoritmos mais populares de aprendizado profundo (*do inglês, deep learning*), sendo atualmente a abordagem mais avançada para reconhecimento e classificação de imagens pertencentes à classe de Aprendizado Supervisionado (SOLAZZO et al., 2016).

Segundo Lin et al. (2017) pode-se dizer que a CNN é um poderoso algoritmo que aprende a executar tarefas de classificação diretamente de imagens, onde três fatores tornam o seu uso tão popular (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), sendo eles:

1. Não necessidade de extração manual de características pelo fato destas serem aprendidas diretamente pela CNN;
2. Resultados de reconhecimento são de última geração;
3. CNNs podem ser treinadas novamente para novas tarefas de reconhecimento, permitindo que sejam construídas redes pré-existentes.

As CNNs possuem uma estrutura específica dividida em dois blocos principais: convolucional e linear. O primeiro determina a principal diferença em relação às redes neurais artificiais clássicas e pode ser considerado como o ponto chave dessas redes (SOLAZZO et al., 2016). Ele é composto basicamente pelas camadas convolucionais que são responsáveis pela extração de características, geralmente alternado com as camadas de subamostragem (*pooling*). Já o segundo bloco é formado por uma rede neural totalmente conectada (*também conhecida como Multi-Layer Perceptron (MLP)*) que é usada para coletar as informações extraídas e realizar a classificação (LIN et al., 2017). Seguindo esta estrutura básica, Cun et al. (1990) construíram uma arquitetura de Rede Neural Convolucional para reconhecimento de dígitos que ficou conhecida como LeNet-5 (Figura 3.8), na qual obtiveram resultados que alcançaram o estado-da-arte para o problema.

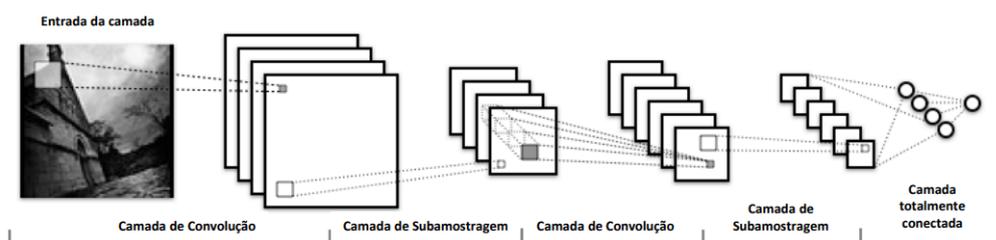


Figura 3.8 – Arquitetura LeNet-5, destacando as operações aplicadas ao longo da rede (JUNIOR, 2018).

No entanto, muitos avanços já foram feitos na área, gerando diferentes organizações de camadas, operações, e por consequência, diferentes arquiteturas (GU et al., 2015). Desta forma, as seções a seguir visa explicar o funcionamento de uma arquitetura básica de uma CNN bem como algumas arquiteturas encontradas na literatura que foram escolhidas neste trabalho.

### 3.3.3.1 Camada Convolutiva

Segundo Vargas, Carvalho e Vasconcelos (2016), a camada convolutiva é composta por um conjunto de neurônios que são responsáveis por fazer o mesmo que alguns processos tradicionais de visão computacional, aplicar filtros sobre a imagem de entrada. Estes filtros (também conhecidos como kernels) geralmente são matrizes pequenas como, por exemplo, matrizes  $3 \times 3$  nos 3 canais (R, G e B) de cor, sendo estas compostas por valores reais que serão responsáveis pela operação de convolução sobre a imagem de entrada (CHAVES, 2019). Basicamente, esta operação realiza uma espécie de varredura na imagem, multiplicando a matriz da imagem pela matriz kernel com o intuito de obter mapas de características (do inglês *feature maps*) (VARGAS; CARVALHO; VASCONCELOS, 2016). É importante evidenciar que os filtros possuem dois parâmetros principais que podem influenciar no tamanho dos mapas de características (EBERMAM; KROHLING, 2018), sendo eles:

- **preenchimento (do inglês *padding*):** inserção de zeros em torno da imagem de entrada;
- **passo (do inglês *stride*):** tamanho do deslocamento do filtro em pixel sobre a imagem.

Na Figura 3.9 é possível visualizar como é realizada a operação de convolução, em que um filtro com estes parâmetros é aplicado sobre uma determinada região da matriz da imagem.

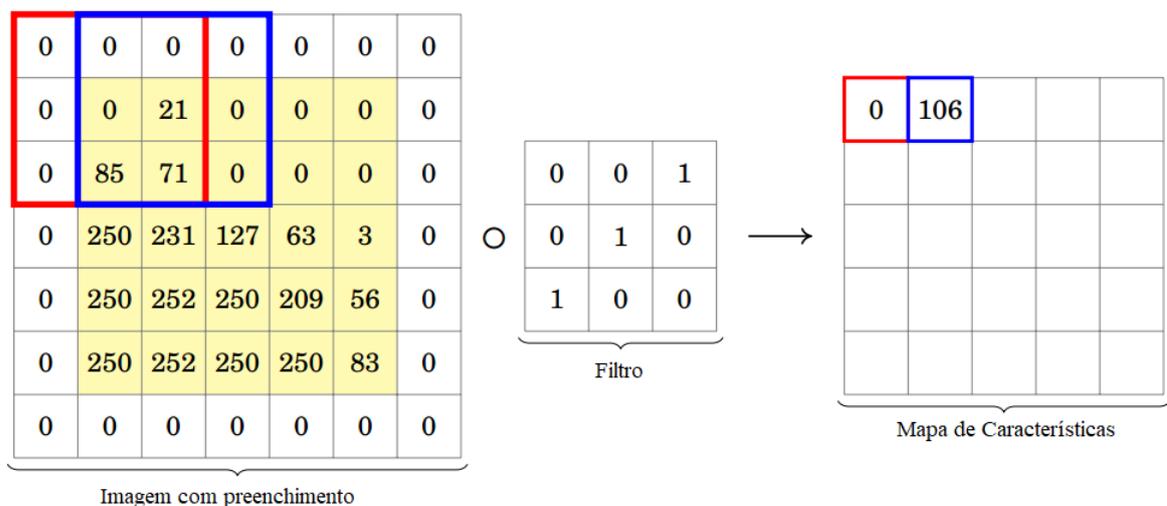


Figura 3.9 – Camada convolutiva parcial, utilizando um filtro  $3 \times 3$  (PAVLOVSKY, 2017).

Ainda na camada de convolução, para que seja possível a resolução de problemas não-lineares pela rede, comumente é utilizada uma função de ativação sobre os mapas de características (EBERMAM; KROHLING, 2018). Segundo Nwankpa et al. (2018), a função de ativação mais utilizada em redes neurais convolucionais é a unidade linear retificada (*Rectified Linear Unit*, ReLU) na qual os valores negativos são substituídos por zero e os valores positivos são mantidos, conforme é visto na Figura 3.10.

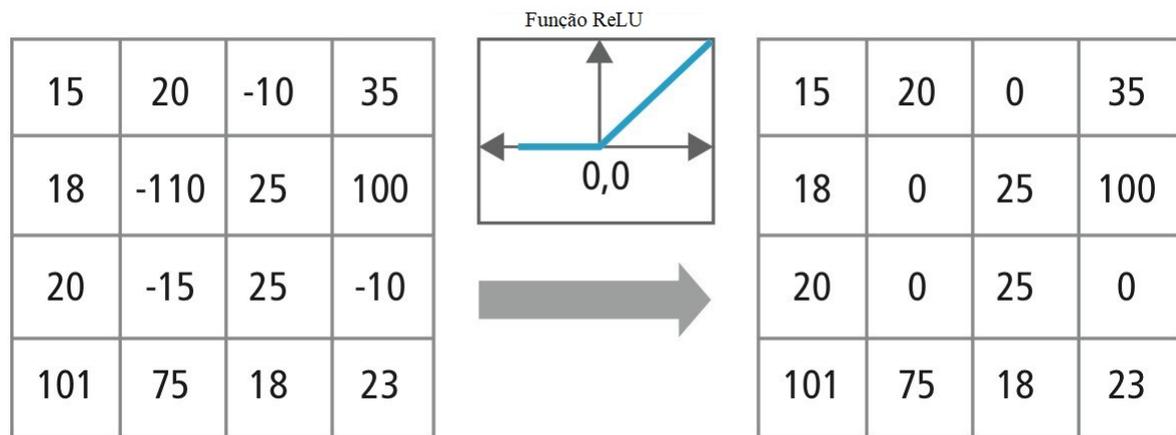


Figura 3.10 – Exemplo de aplicação da função ReLU sobre uma mapa de características (SINGHAL, 2017).

### 3.3.3.2 Camada de Subamostragem

Conforme Junior (2018), a camada de subamostragem (também conhecida como *pooling*) é aplicada após a camada de convolução com o objetivo de diminuir o tamanho espacial das matrizes. Desta forma, a quantidade de parâmetros a serem aprendidos pela rede é reduzido, contribuindo para evitar o sobreajuste da rede (do inglês *overfitting*). Pode-se dizer que a redução produz um treinamento mais rápido e eficaz, porém sua maior importância é pelo fato de criar uma invariância espacial (VARGAS; CARVALHO; VASCONCELOS, 2016).

De um modo geral, esta camada seleciona uma região do mapa de características através de um filtro e aplica uma subamostragem, sendo os tipos mais comuns *average pooling* e *max pooling* (SINGHAL, 2017). No *average pooling* calcula-se a média dos valores de uma determinada região do mapa de características, enquanto no *max pooling* seleciona-se o maior valor. Assim como na camada convolutiva, é necessário definir o tamanho do filtro e o valor do passo que será aplicado (EBERMAM; KROHLING, 2018). A Figura 3.11 ilustra o resultado da aplicação das operações de *average pooling* e *max pooling* sobre um mapa de características, resultando em um mapa com menor dimensionalidade.

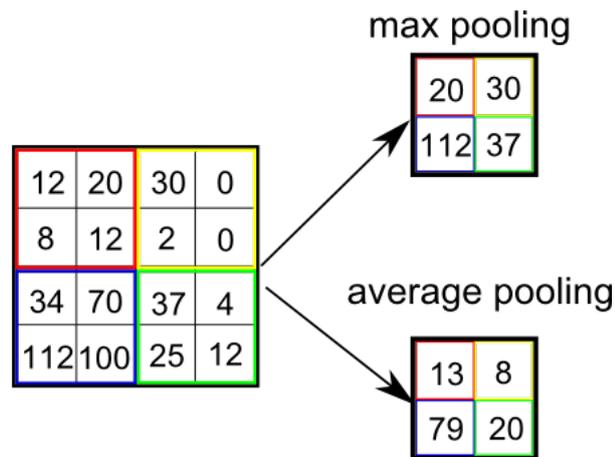


Figura 3.11 – Exemplo de aplicação do *max pooling* e *average pooling* de tamanho  $2 \times 2$  com passo 2 sobre uma mapa de características (SINGHAL, 2017).

### 3.3.3.3 Camada Totalmente Conectada

De modo geral, as camadas totalmente conectadas são formadas por uma rede neural totalmente conectada (também conhecida como *Multi-Layer Perceptron (MLP)*) (LIN et al., 2017). Como os neurônios nesta camada possuem conexões com todos os neurônios da camada anterior (geralmente subamostragem), é necessário realizar a conversão dos mapas de características em só um vetor, o qual servirá de entrada para a rede neural MLP (EBERMAM; KROHLING, 2018). Esta é responsável pela classificação das características recebidas, utilizando geralmente nas unidades de saída uma função softmax, que obtém a probabilidade de uma imagem de entrada pertencer a uma determinada classe (GU et al., 2015).

### 3.3.3.4 AlexNet

A AlexNet desenvolvida por Krizhevsky, Sutskever e Hinton (2012), é uma arquitetura de CNN com mais de 60 milhões de parâmetros que foi vencedora do desafio de reconhecimento visual em grande escala do ImageNet em 2012. Ela superou a performance de métodos estado-da-arte da época na competição, ficando mais de dez pontos percentuais à frente do segundo colocado (MARTINS, 2018). Basicamente, esta rede é composta por oito camadas, sendo as cinco primeiras compostas por camadas convolucionais (algumas delas seguidas por camadas de *max pooling*) e as três últimas formadas por camadas totalmente conectadas, como pode ser visto na Figura 3.12.

Segundo Krizhevsky, Sutskever e Hinton (2012), a AlexNet utiliza a unidade linear retificada (ReLU) como função de ativação por apresentar melhor desempenho de treinamento em relação as funções de tangente hiperbólica e sigmoide. Além disso, ela emprega o uso da técnica de dropout para evitar o sobreajuste da rede, na qual esta consiste em zerar os valores de saída de um neurônio de uma camada com 50% de chance. Desta forma, a rede aprende

descritores mais robustos, já que elimina a contribuição de alguns neurônios e tira a garantia de que um neurônio possa depender em outro (MARTINS, 2018).

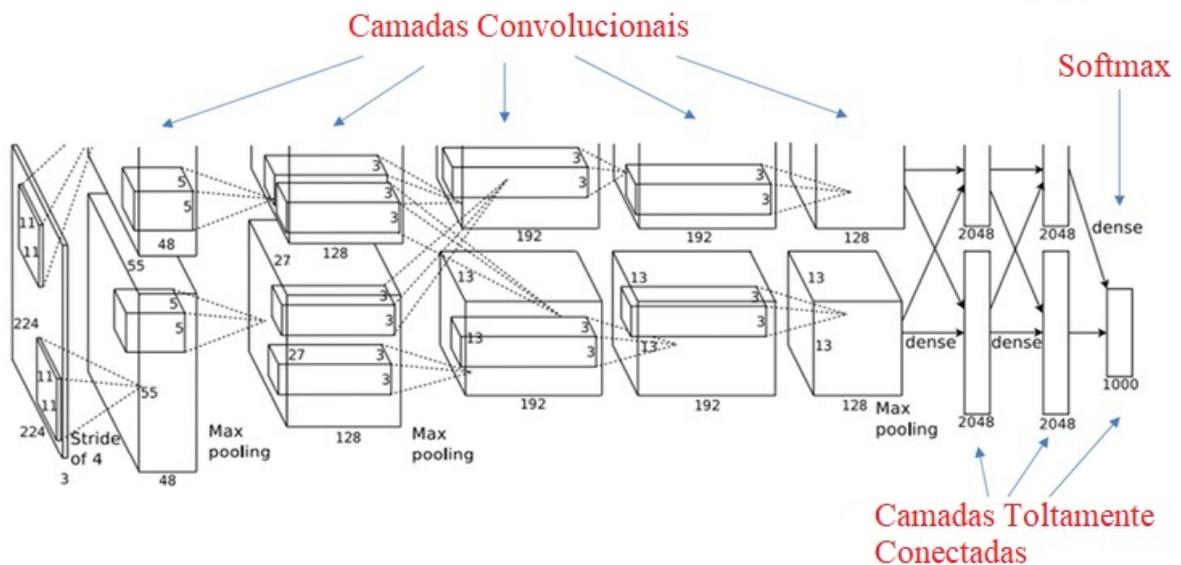


Figura 3.12 – Arquitetura da Rede Neural Alexnet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

### 3.3.3.5 GoogleNet

A GoogLeNet também conhecida como Inception-v1 foi desenvolvida por (SZEGEDY et al., 2015), sendo a rede vencedora do desafio de reconhecimento visual em grande escala do ImageNet no ano de 2014. Trata-se de uma rede composta por 22 camadas de profundidade que desviou da abordagem tradicional de empilhar camadas convolutivas e camadas de agrupamento, em uma estrutura sequencial (MASSUCATTO, 2018). A principal característica dessa arquitetura é a melhor utilização dos recursos de computação dentro da rede, na qual foi desenvolvido um módulo conhecido como *Inception* que reduziu drasticamente o número de parâmetros na rede, sendo 12 vezes menor comparado a rede AlexNet (SZEGEDY et al., 2015). De acordo com Silva (2018), o objetivo principal deste módulo *Inception* é atuar como um extrator de características em vários níveis, computando convoluções muito pequenas, como por exemplo  $1 \times 1$ ,  $3 \times 3$  e  $5 \times 5$ , dentro do mesmo módulo da rede. Na Figura 3.13 pode-se visualizar um exemplo do módulo *Inception* presente na rede.

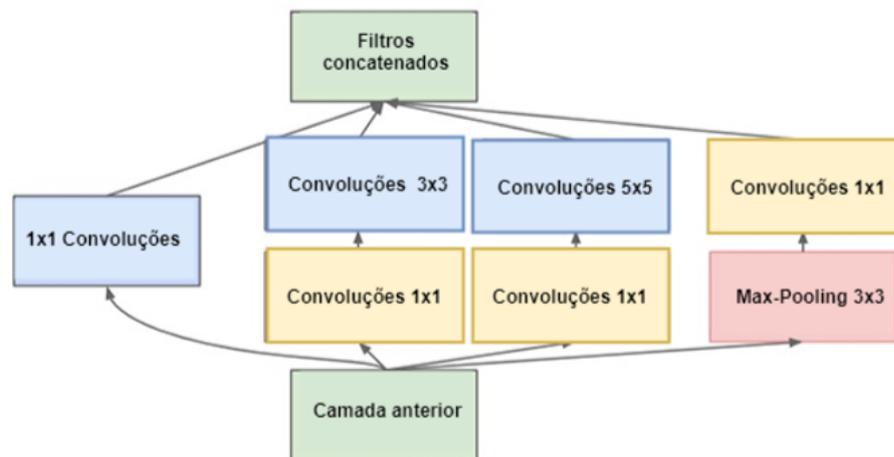


Figura 3.13 – Módulo *Inception* (SZEGEDY et al., 2015).

### 3.3.3.6 SqueezeNet

Em 2016, Iandola et al. (2016) desenvolveram a SqueezeNet com o intuito de obter-se uma rede que mantém a precisão competitiva mesmo com poucos parâmetros. Segundo Iandola et al. (2016), a SqueezeNet atinge a precisão do nível AlexNet no ImageNet com 50 vezes menos parâmetros. Ao projetá-la, o objetivo era criar uma rede neural menor que pudessem se encaixar sem dificuldades na memória do computador e ser transmitida mais facilmente através de uma rede de computadores (GANESH, 2016). Para conseguir alcançar este objetivo, os autores se inspiraram na idéia do módulo Inception para projetar um módulo chamado Incêndio (do inglês *Fire*), sendo este composto por uma camada de compressão e uma camada de expansão (KWON et al., 2017), conforme é mostrado na Figura 3.14. Basicamente, a camada de compressão utiliza um filtro de convolução 1x1 para reduzir e compactar os elementos de entrada. Já a camada de expansão emprega o uso de filtros de convolução  $1 \times 1$  e  $3 \times 3$  para o aprendizado, realizando a concatenação dessas convoluções.

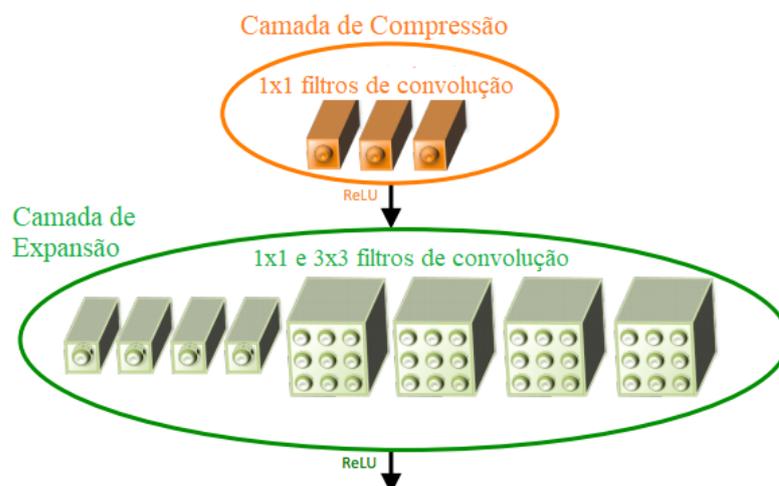


Figura 3.14 – Módulo *Incêndio* (IANDOLA et al., 2016).

De acordo com [Iandola et al. \(2016\)](#), a estrutura básica da SqueezeNet é composta por vários módulos de incêndio, alternados com camadas de subamostragem (pooling), sendo aplicado a técnica de dropout sobre cada módulo de incêndio. Além disso, é aplicado a unidade linear retificada (ReLU) como função de ativação sobre cada camada de compressão e expansão.

## 4 Metodologia

Como visto no Capítulo 1, diversas pesquisas na área de biometria têm resultado em trabalhos que exploram diversas características distintas dos seres humano como, por exemplo, a face e a íris. Porém, a captura dos dados de muitos dessas fontes são considerados intrusivos o que dificulta a aceitação destes métodos em sistemas que possuem alto nível de segurança. Entretanto, métodos que utilizam imagens de orelhas para identificação biométrica tem alcançado uma boa aceitação por serem intuitivos e menos intrusivos. Sendo assim, este capítulo visa explicar com mais detalhes a metodologia adotada para a implementação dos métodos em MATLAB. Na Figura 4.1, pode-se visualizar de forma detalhada a metodologia proposta.

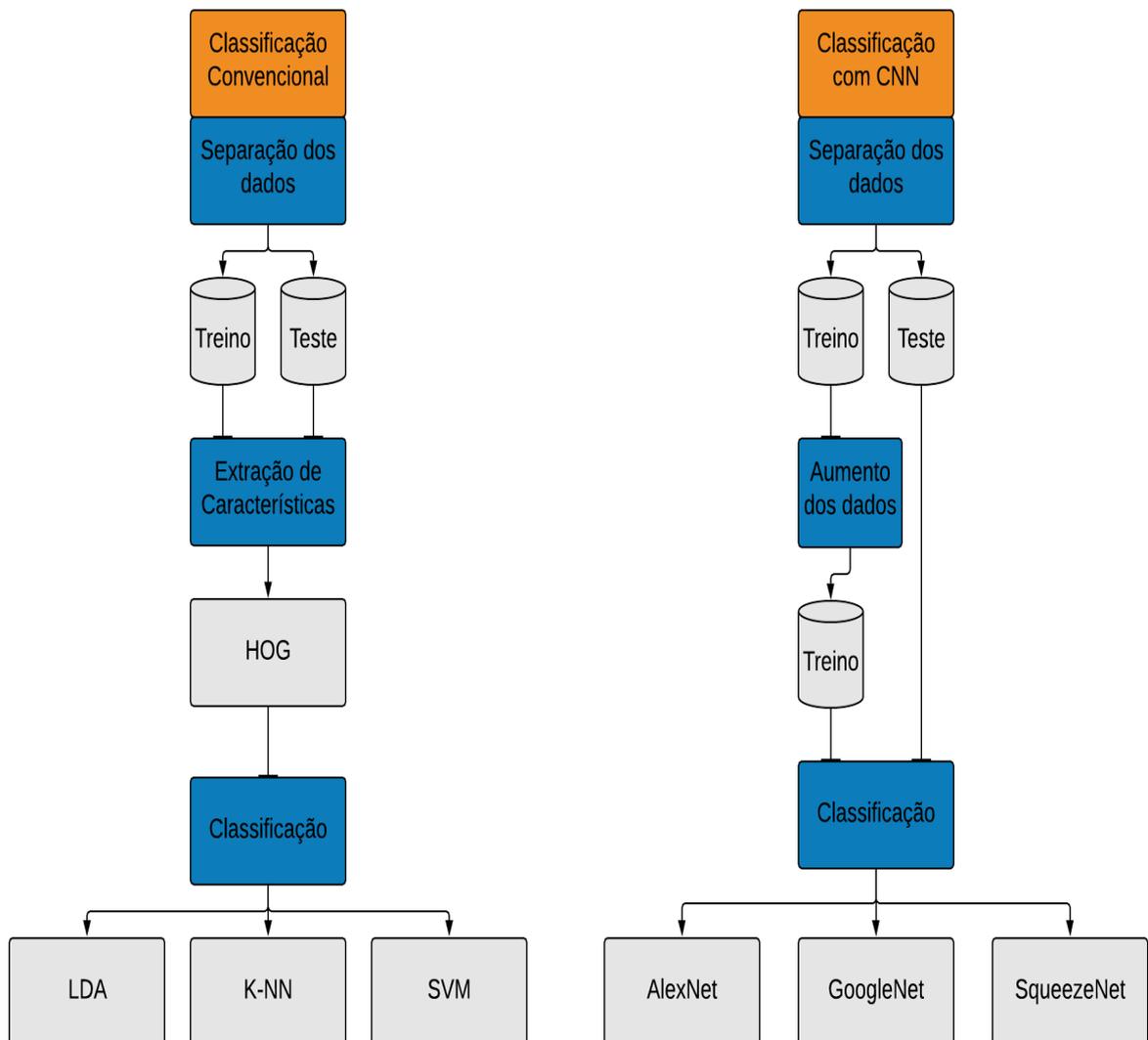


Figura 4.1 – Metodologia proposta para os métodos.

## 4.1 Separação dos Dados

Conforme pode-se ver na Figura 4.1, tanto para as abordagens de classificação convencional quanto para as abordagens de CNN, primeiramente faz-se necessário a separação da base de dados de imagens em dois grupos, sendo eles, treinamento e teste. Segundo Massucatto (2018), existem diversas maneiras de realizar esta divisão, sendo as principais formas o método *holdout*, *k-fold* e *leave-one-out*. Portanto, para a realização dessa tarefa, utilizou-se a técnica de validação cruzada *k-fold* por ser indicada quando se tem uma quantidade pequena de dados (BACKES, 2018) e por ser frequentemente utilizada para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados (KOHAVI, 1995).

Mediante à essas informações, a técnica foi utilizada para dividir o conjunto de dados em  $k$  partes iguais sendo que em cada iteração uma das partes é usada para a avaliação do classificador e as demais para treinar o mesmo, conforme é mostrado na Figura 4.2. Desta forma, os  $k$  resultados obtidos são então combinados para produzir uma única estimativa de erro.

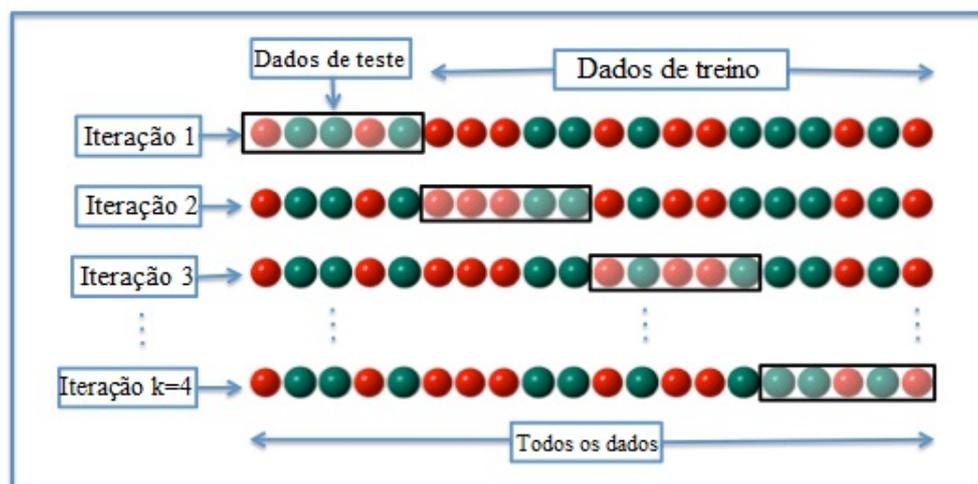


Figura 4.2 – Validação cruzada 4-fold (SOUZA, 2016).

## 4.2 Classificação Convencional

### 4.2.1 Extração de Características

Um dos quesitos mais fundamentais para a elaboração de um sistema de reconhecimento de padrões é a escolha do método de extração de características, visto que a eficiência de tal sistema está atrelado ao poder de discriminação do mesmo. Diante disso, posterior à etapa de separação da base de dados, foi executada a extração de características das orelhas através do descritor HOG. Segundo Dalal e Triggs (2005), a extração do mesmo em uma imagem começa com a repartição da imagem em pequenas regiões espaciais denominadas como células que podem ser ajustadas de acordo com a necessidade. Assim como as células, os blocos também

podem ser ajustados. Portanto, utilizou-se os tamanhos padrões de células e blocos do HOG, sendo estes  $8 \times 8$  e  $2 \times 2$ , respectivamente. A Figura 4.3 exemplifica a extração de características do descritor HOG em relação à uma imagem da orelha contida na base de dados IITD Ear (KUMAR; WU, 2012).

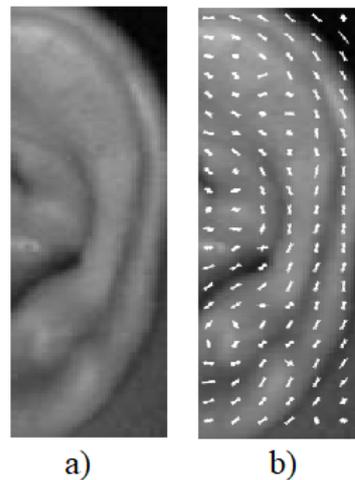


Figura 4.3 – Exemplo de extração de características - a) Imagem da orelha e b) Descrição resultante.

### 4.2.2 Classificação

Em seguida, para efetuação da classificação das imagens foi utilizado os classificadores LDA, K-NN e SVM. É válido ressaltar que para a utilização dos classificadores K-NN e SVM, conforme é explicado no Capítulo 3, algumas estratégias devem ser definidas, como métrica de distância e números de vizinhos para o K-NN e a estratégia de treinamento juntamente com a função de transformação (kernel) para a SVM. Portanto, foi definida a distância Euclidiana como a métrica de distância por ser frequentemente utilizada na literatura e o valor de  $k$  igual um para o K-NN. Já para a SVM, foi adota a estratégia um-contra-todos e utilizado três kernels mais usuais no meio científico (SOUSA, 2016), sendo eles, linear, polinomial e gaussiano. Por fim, é gerada a matriz de confusão para averiguar o desempenho dos classificadores.

## 4.3 Classificação com Redes Neurais Convolucionais

### 4.3.1 Aumento dos Dados

Posterior à etapa de separação dos dados, as imagens do conjunto de treinamento e teste são redimensionadas para os padrões das redes e são aplicadas técnicas de aumento de dados (do inglês *data augmentation*) sobre o conjunto de treinamento. Estas técnicas fazem modificações nas imagens com o objetivo de fazer com que as redes não memorizem detalhes específicos das imagens de treinamento, para que assim, as redes não percam sua capacidade de generalização se

tornando especialista apenas nos dados de treinamento (SILVA, 2018). Desta forma, algumas transformações nas imagens foram realizadas, como: espelhar as imagens aleatoriamente em relação ao eixo horizontal, rotacionar aleatoriamente as imagens horizontalmente ou verticalmente em até 30 pixels e aumentar a escala da imagem nos eixos horizontais e verticais aleatoriamente no intervalo de 1.1 e 1.5. Na Figura 4.4 é possível visualizar o resultado dessas transformações sobre uma imagem de orelha contida na base de dados IITD Ear (KUMAR; WU, 2012).

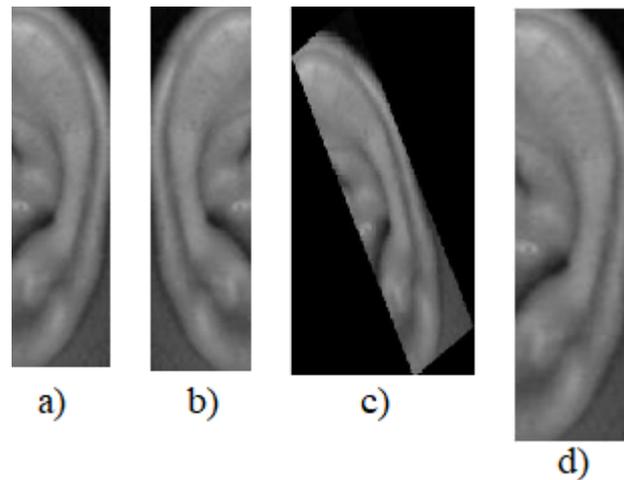


Figura 4.4 – Exemplo do aumento dos dados realizados - a) Imagem original, b) Espelhamento, c) Rotação em 30 graus e d) Aumento de escala (1.2x).

### 4.3.2 Classificação

Segundo Karpathy (2016), é muito raro treinar uma CNN completamente do zero para realizar a tarefa de classificação, pelo fato destas demandarem muitos recursos computacionais e uma base de dados suficientemente grande. Desta forma, a técnica de transferência de aprendizado (do inglês *transferring learning*) tem sido amplamente utilizado no dias atuais, na qual uma rede previamente treinada em uma tarefa com sua aprendizagem passa a ser utilizada em outra tarefa relacionada (OLIVAS et al., 2009). Mediante à essas informações, após a realização do aumento dos dados, as imagens resultantes desta etapa são alimentadas nas redes pré-treinadas Alexnet, Googlenet e Squeezenet para retreinar as mesmas para a classificação do conjunto de teste. Como as redes são pré-treinadas no ImageNet (DENG et al., 2009), a saída das camadas totalmente conectadas possui o número de classes da base do ImageNet. Sendo assim, as camadas totalmente conectadas foram substituídas por novas camadas totalmente conectadas que possui o número de saídas (classes) adequado para o problema deste trabalho.

É importante evidenciar que para acelerar o treinamento das redes e reduzir o custo computacional desta tarefa, foi empregado o uso do Gradiente Descendente Estocástico (*Stochastic Gradient Descent, SGD*) utilizando amostras aleatórias dos dados com um tamanho fixo (*mini-batches*). Esta abordagem tem sido uma boa opção para treino de modelos de CNN (JUNIOR, 2018), já que esta ao invés de analisar todos dados disponíveis, analisa-se apenas uma amostra

da base por iteração de treino, adicionando aleatoriedade ao processo (PONTI; COSTA, 2017). Sendo assim, alguns parâmetros tiveram que ser ajustados, sendo eles: taxa de aprendizado (*learning rate*), momento (*momentum*), tamanho da amostra (*mini-batch*) e quantidade de ciclos de aprendizagem também conhecido como épocas (*epochs*). Os três primeiros parâmetros e seus respectivos valores foram modificados seguindo as recomendações de Ruder (2016) e Bengio (2012), ao contrário do último parâmetro, onde o valor foi definido empiricamente. A Tabela 4.1 apresenta os parâmetros utilizados para o treinamento dos redes.

Tabela 4.1 – Parâmetros usados para treinamento dos modelos de CNN.

<b>Parâmetro</b>	<b>Valor</b>
Taxa de aprendizado	0.0001
Momento	0.9
Tamanho da amostra	32
Épocas	50

Assim como nos métodos convencionais de classificação, a matriz de confusão é gerada para analisar o desempenho das redes.

# 5 Experimentos Computacionais

Neste capítulo são apresentados os experimentos realizados para a validação do método proposto. Todos os experimentos foram realizados em uma máquina com processador Intel Xeon E5-2603 de 1.8 GHz com 12 GB de RAM e disco rígido de 600 GB, utilizando o sistema operacional Windows 7. Na Seção 5.1 é apresentada uma sucinta explicação sobre as bases de dados utilizada para os experimentos. Na Seção seguinte, é descrita as métricas de avaliação utilizadas para examinar a eficiência dos métodos abordados no Capítulo 4. Por fim, na Seção 5.3 são mostrados os resultados obtidos através dos experimentos bem como a comparação destes com os resultados da literatura.

## 5.1 Base de Dados

Nesta seção será apresentada uma sucinta descrição das bases de dados utilizadas neste trabalho, sendo elas IIT Delhi Ear disponibilizada pelo *Indian Institute of Technology Delhi* e AMI Ear Database disponibilizada pela *Universidad de Las Palmas de Gran Canaria (ULPGC)*. Ambas se encontram disponíveis publicamente para trabalhos científicos.

### 5.1.1 IITD Ear

A base de dados IITD Ear desenvolvido por [Kumar e Wu \(2012\)](#) é constituído de imagens de orelhas coletadas dos funcionários e alunos do *Indian Institute of Technology Delhi*, onde estas foram adquiridas à distância usando uma configuração simples de imagem. Esta base de dados contém dois conjuntos de imagens de orelhas, sendo que o primeiro é formado pelas imagens originais de tamanho  $272 \times 204$  pixels e o segundo por imagens cortadas e normalizadas automaticamente de tamanho  $180 \times 50$  pixels. Os dois conjuntos são formados por 125 pessoas que estão na faixa etária de 14 a 58 anos na qual cada pessoa tem pelo menos três imagens da orelha, totalizando em 493 imagens. Exemplos de amostras contidas na base de dados são mostradas na Figura 5.1, sendo escolhido para a utilização o conjunto composto por imagens de orelhas cortadas e normalizadas.

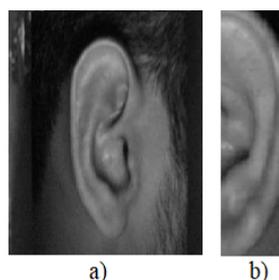


Figura 5.1 – Amostras da base IITD Ear - a) Imagem do conjunto I e b) Imagem do conjunto II.

### 5.1.2 AMI Ear

A base de dados AMI Ear foi desenvolvida por Gonzalez, Alvarez e Mazorra (2012) para a tese de doutorado em Ciência da Computação da Esther Gonzalez. Este conjunto de dados é constituído de 700 imagens que envolve 100 pessoas diferentes com idade entre 19 e 65 anos, sendo 6 imagens da orelha direita e 1 da orelha esquerda. Todas as imagens foram capturadas por uma câmara Nikon D100, sob as mesmas condições de iluminação e ângulos diferentes, onde estas possuem resolução de  $702 \times 492$  pixels. Exemplos de amostras contidas na base de dados são mostradas na Figura 5.2, na qual cinco das imagens capturadas foram do perfil do lado direito do indivíduo com este voltado para a frente (Fig 5.2 (a)), olhando para cima (Fig 5.2 (b)), para baixo (Fig 5.2 (c)), olhando para a esquerda (Fig 5.2 (d)) e para a direita (Fig 5.2 (e)). A sexta imagem (Fig 5.2 (f)) foi tirada com um comprimento focal diferente da câmara (zoom) mas também com a pessoa virada para a frente. A última imagem (Fig 5.2 (g)) foi tirada da orelha esquerda sendo que a pessoa está virada para frente e com a mesma distância focal da câmara das cinco primeiras imagens.

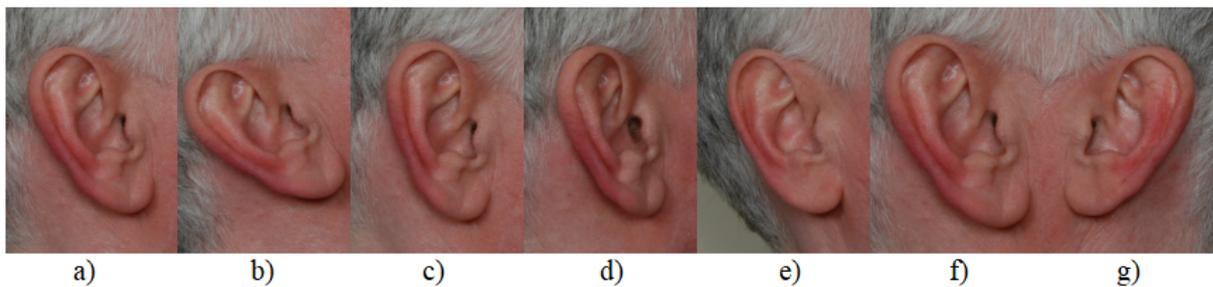


Figura 5.2 – Exemplos de imagens da AMI Ear.

## 5.2 Métricas de Avaliação

Com o intuito de analisar a eficiência do método proposto para a comparação com outros propostos na literatura, algumas medidas foram calculadas a partir da matriz de confusão. Esta mede o desempenho do classificador através dos acertos e erros, indicando quatro possibilidades:

- **Verdadeiros Positivos (VP):** exemplos classificados como positivos e realmente são da classe positiva;
- **Verdadeiros Negativos (VN):** exemplos classificados como negativos e realmente são da classe negativa;
- **Falsos Positivos (FP):** exemplos classificados como positivos mas pertencem a classe negativa;

- **Falsos Negativos (FN):** exemplos classificados como negativos mas pertencem a classe positiva.

Nesta perspectiva, as medidas calculadas foram precisão (do inglês, *precision*), revocação (do inglês, *recall*), acurácia (do inglês, *accuracy*) e medida-f1 (do inglês, *f1-measure*). Manning, Raghavan e Schütze (2008) faz a seguinte definição sobre as mesmas:

- **Precisão:** número de amostras corretamente classificadas dividida pelo número total de amostras retornadas. Pode ser calculada pela seguinte fórmula:  $\frac{VP}{VP+FP}$ ;
- **Revocação:** número de amostras corretamente classificadas divididas pelo número de amostras relevantes que deveriam ter sido retornadas. Pode ser calculada pela seguinte fórmula:  $\frac{VP}{VP+FN}$ ;
- **Acurácia:** percentual de acerto obtido na classificação das amostras. Pode ser calculada pela seguinte fórmula:  $\frac{VP+VN}{VP+VN+FP+FN}$ ;
- **Medida-f1:** média harmônica da precisão e revocação, podendo assumir valores entre 0 e 1 na qual quanto mais próximo de 1 melhor o desempenho do classificador. Pode ser calculada pela seguinte fórmula:  $\frac{2*Precision*Recall}{Precision+Recall}$ .

### 5.3 Resultados Alcançados

Foram realizados 10 experimentos para cada uma das bases de dados com o intuito de avaliar a eficiência dos métodos propostos, sendo que os melhores resultados foram obtidos com o valor de k igual a 3 para a técnica de validação cruzada *k-fold*. As Tabelas 5.1 e 5.2 apresentam a média dos resultados alcançados pelos métodos através dos experimentos realizados nas bases de dados IITD Ear e AMI Ear, onde valores mais próximos de 1 correspondem a um bom desempenho dos métodos.

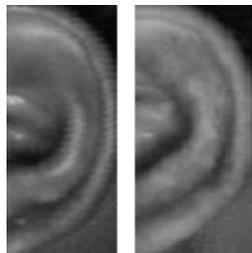
Tabela 5.1 – **IITD Ear** - Média dos resultados em relação à Precisão, Revocação, Acurácia e Medida-F1.

Classificador	Precisão	Revocação	Acurácia	Medida-F1
HOG + k-NN (K=1)	0,9250	0,9020	0,9289	0,9134
HOG + LDA	0,9265	0,9028	0,9297	0,9145
HOG + SVM (Kernel linear)	0,9230	0,9006	0,9259	0,9117
HOG + SVM (Kernel gaussiano)	0,9270	0,9053	0,9301	0,9160
<b>HOG + SVM (Kernel polinomial)</b>	<b>0,9349</b>	<b>0,9125</b>	<b>0,9378</b>	<b>0,9236</b>
AlexNet	0,8712	0,9047	0,8712	0,8876
GoogleNet	0,8872	0,9104	0,8872	0,8987
SqueezeNet	0,8456	0,8895	0,8456	0,8670

Tabela 5.2 – **AMI Ear** - Média dos resultados em relação à Precisão, Revocação, Acurácia e Medida-F1.

Classificador	Precisão	Revocação	Acurácia	Medida-F1
HOG + k-NN (K=1)	0,6420	0,6945	0,6420	0,6672
HOG + LDA	0,6500	0,7068	0,6500	0,6772
HOG + SVM (Kernel linear)	0,8080	0,8596	0,8080	0,8330
HOG + SVM (Kernel gaussiano)	0,7830	0,8352	0,7830	0,8083
HOG + SVM (Kernel polinomial)	0,8340	0,8877	0,8340	0,8640
AlexNet	0,9520	0,9749	0,9520	0,9633
<b>GoogleNet</b>	<b>0,9540</b>	<b>0,9755</b>	<b>0,9540</b>	<b>0,9646</b>
SqueezeNet	0,9340	0,9662	0,9340	0,9498

Os resultados alcançados para a base de dados IITD Ear, mostrados na Tabela 5.1, indicam que os métodos de classificação convencional obtiveram um desempenho superior aos métodos de CNNs. Pode-se notar que a melhor abordagem de CNN, que é por certo a GoogleNet, comparando com os métodos convencionais em relação à taxa de reconhecimento (acurácia), fica aproximadamente 4% atrás da pior abordagem (HOG + SVM (Kernel linear)) e 5% atrás da melhor abordagem (HOG + SVM (Kernel polinomial)). É válido ressaltar que tanto os métodos convencionais quanto as CNNs encontraram maiores dificuldades na classificação para orelhas que possuem formas parecidas, como é mostrado na Figura 5.3.

Figura 5.3 – **IITD Ear** - Exemplo de confusão na classificação.

Ao fazer a análise em relação a base de dados AMI Ear, pode-se verificar que os métodos propostos de CNNs superaram os métodos convencionais. Isso se dá pelo fato desta base conter algumas imagens com oclusão parcial por cabelo e brincos, como pode ser visto na Figura 5.4, o que causa maiores dificuldades de aprendizado nos métodos convencionais, resultando em muitas classificações errôneas.

Figura 5.4 – **AMI Ear** - Exemplo de oclusão parcial por brinco.

Como as redes neurais profundas precisam de grande quantidades de dados rotulados para aprender conceitos simples, em alguns cenários os métodos convencionais ainda se mostram mais eficientes, como foi observado no cenário da base IITD Ear. Mesmo realizando técnicas de aumento de dados, as CNNs não superaram nenhum dos métodos convencionais propostos. Porém, em cenários mais realistas como encontrados na segunda base, as CNNs mostraram-se mais eficientes e robustas, o que justifica a ampla utilização nos dias atuais, já que grandes dificuldades são encontradas nesses cenários.

### 5.3.1 Resultados em Comparação ao Estado da Arte

Para uma análise abrangente, os resultados obtidos pelo melhor método convencional proposto (HOG + SVM (Kernel polinomial)) e de CNN (GoogleNet) foram comparados a outros métodos encontrados na literatura como é mostrado na Tabela 5.3. Ao examiná-la em relação a base de dados IIT Delhi Ear, pode-se ver que a melhor abordagem proposta com CNN supera somente o método proposto por Hurley, Nixon e Carter (2002), possuindo um desempenho inferior a diversos métodos. Já o método convencional alcança um bom desempenho comparado a outros métodos similares existentes na literatura, sendo que a taxa de reconhecimento da abordagem difere somente aproximadamente 2% do segundo melhor resultado e 4% do melhor resultado. Em relação à base de dados AMI Ear, a abordagem proposta com CNN supera todos os métodos, ficando aproximadamente 5% do segundo melhor resultado e à 12% do melhor método convencional proposto, demonstrando a sua eficiência em alguns cenários de oclusão parcial da orelha.

Tabela 5.3 – Comparação ao estado da arte com base na porcentagem de acurácia.

Método	Base	Acurácia
Hurley, Nixon e Carter (2002)	IIT Delhi Ear	74.93%
<b>GoogleNet</b>	IIT Delhi Ear	88.72%
Damer e Fuhrer (2012)	IIT Delhi Ear	92%
Sarangi, Mishra e Dehuri (2017)	IIT Delhi Ear	92.76%
Boodoo-Jahangeer e Baichoo (2013)	IIT Delhi Ear	93%
<b>HOG + SVM (Kernel polinomial)</b>	IIT Delhi Ear	93.78%
Kumar e Wu (2012)	IIT Delhi Ear	96.27%
Anwara, Ghanyb e Elmahdy (2015)	IIT Delhi Ear	<b>98%</b>
Chowdhury et al. (2017)	AMI Ear	70.58%
Rodrigues (2018)	AMI Ear	79%
<b>HOG + SVM (Kernel polinomial)</b>	AMI Ear	83.40%
Rodrigues (2018)	AMI Ear	90%
Soni et al. (2014)	AMI Ear	90%
<b>GoogleNet</b>	AMI Ear	<b>95.40%</b>

## 6 Conclusão

Nesta trabalho foi realizado um estudo sobre a biometria da orelha bem como a revisão de alguns métodos existentes na literatura para identificação humana com o uso de imagens da orelha. Métodos que utilizam orelhas para identificação biométrica tem alcançado uma boa aceitação por serem intuitivos e menos intrusivos. Podemos dizer que este estudo foi fundamental para um melhor entendimento sobre o tema em questão, ampliando o conhecimento e possibilitando a realização de uma comparação entre métodos convencionais e métodos utilizando CNNs no processo de identificação de pessoas através de imagens de orelha. Sendo assim, dentre as diversas arquiteturas de CNN disponíveis na literatura, foram escolhidas AlexNet, GoogleNet e Squeezenet. Da mesma forma, para os métodos convencionais foram selecionados três classificadores para serem combinados com o descritor HOG, sendo eles K-NN, LDA e SVM.

É válido ressaltar que o processo de identificação biométrica utilizando imagens da orelha apresenta certas vantagens quando comparado a outros que utilizam medidas biométricas mais tradicionais como: face e impressão digital. Primeiramente, pelo fato da orelha ser menor que a região da face e não sofrer com rugas, envelhecimento e expressões faciais e em segundo, por esta possuir uma região maior que a impressão digital e não sofrer com degradação durante o tempo. Porém, como no caso das mulheres, as orelhas podem sofrer oclusões de cabelo e/ou brinco, o que causa grandes dificuldades na identificação resultando em um maior interesse dos pesquisadores neste cenário, já que estes são considerados mais realistas.

A eficácia dos métodos propostos foi averiguada através de 10 experimentos realizados na base de dados IIT Delhi Ear e 10 experimentos realizados na base de dados AMI Ear. Para a análise deste desempenho, foram calculadas algumas métricas de avaliação como precisão, revocação, acurácia e medida-f1. A avaliação experimental confirma que em cenários mais realistas como encontrados na base AMI Ear, as CNNs alcançaram um desempenho superior comparado aos métodos convencionais e a outros similares existentes na literatura, alcançando uma média de acurácia de 95.40%. Porém na base IITD Ear, os métodos convencionais mostraram-se mais eficazes, o que mostra que alguns cenários são melhores que as CNNs.

Em resumo, visto que este trabalho realizou um estudo comparativo entre métodos no processo de identificação biométrica, os resultados obtidos podem contribuir com pesquisadores iniciantes na área, servindo como um ponto de partida para resultados mais promissores.

### 6.1 Trabalhos Futuros

Para trabalhos futuros, propõe-se:

- Aplicar outras técnicas de extração de características além do HOG, como por exemplo o

*SIFT* e o *SURF*;

- Selecionar mais classificadores convencionais (*Naive Bayes* e *Floresta Aleatória*) e arquiteturas de CNN (*VGGNet* e *Resnet*) para avaliação;
- Utilizar as bases não controladas *AWE*, *ITWE* e *UERC*.

# Referências

- AGGARWAL, C. C. *Data Classification: Algorithms and Applications*. 1. ed. The address: Chapman & Hall/CRC, 2014. ISBN 1466586745, 9781466586741.
- ALARAJ, M.; HOU, J.; FUKAMI, T. A neural network based human identification framework using ear images. In: *TENCON 2010 - 2010 IEEE Region 10 Conference*. [S.l.: s.n.], 2010. p. 1595–1600.
- ALI, M.; JAVED, M. Y.; BASIT, A. Ear recognition using wavelets. *Proceedings of Image and Vision Computing New Zealand*, 01 2007.
- ALMEIDA, O. C. P. de. *Técnicas de processamento de imagens para localização e reconhecimento de faces*. Dissertação (Mestrado em Ciências - Ciências da Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2006.
- ALVA, M.; SRINIVASARAGHAVAN, A.; SONAWANE, K. A review on techniques for ear biometrics. In: *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. [S.l.: s.n.], 2019. p. 1–6.
- ANWARA, A. S.; GHANYB, K. K. A.; ELMAHDY, H. Human ear recognition using geometrical features extraction. In: *Elsevier*. [S.l.: s.n.], 2015. v. 65.
- BACKES, A. *Aula 3 - Classificadores Elementares*. 2018. Disponível em: <<http://www.facom.ufu.br/~backes/pgc204/Aula03-ClassificadoresElementares.pdf>>.
- BALAKRISHNAMA, S.; GANAPATHIRAJU, A. Linear discriminant analysis—a brief tutorial. v. 11, 01 1998.
- BENGIO, Y. Practical recommendations for gradient-based training of deep architectures. In: \_\_\_\_\_. *Neural Networks: Tricks of the Trade: Second Edition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 437–478. ISBN 978-3-642-35289-8. Disponível em: <[https://doi.org/10.1007/978-3-642-35289-8\\_26](https://doi.org/10.1007/978-3-642-35289-8_26)>.
- BERTILLON, A. *La photographie judiciaire: avec un appendice sur la classification et l'identification anthropométriques*. [S.l.]: Paris: Gauthier-Villars, 1890.
- BONESSO, D. *Estimação dos Parâmetros do Kernel em um Classificador Svm na Classificação de Imagens Hiperespectrais em uma Abordagem Multiclasse*. Dissertação (Mestrado em Sensoriamento Remoto) — Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.
- BOODOO-JAHANGEER, N. B.; BAICHO, S. Lbp-based ear recognition. In: *13th IEEE International Conference on Bioinformatics and BioEngineering*. [S.l.: s.n.], 2013. p. 1–4.
- BURGE, M.; BURGER, W. Ear biometrics in computer vision. In: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. [S.l.: s.n.], 2000. v. 2, p. 822–826 vol.2. ISSN 1051-4651.

CARREIRA-PERPINAN, M. A. Compression neural networks for feature extraction: Application to human recognition from ear images. *MSc thesis, Faculty of Informatics, Technical University of Madrid*, 1995.

CERNA, L. R. *Fusão de descritores de histogramas de gradientes para a detecção de faces baseado em uma cascata de classificadores*. 113 p. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Ouro Preto, Ouro Preto, 2014.

CHAVES, E. de L. *Detecção de câncer de mama por meio de imagens infravermelhas utilizando Redes Neurais Convolucionais*. Monografia (Bacharel em Ciência da Computação) — Universidade Federal de Uberlândia, Uberlândia, 2019.

CHOWDHURY, D. P.; BAKSHI, S.; GUO, G.; SA, P. K. On applicability of tunable filter bank based feature for ear biometrics: A study from constrained to unconstrained. *Journal of Medical Systems*, v. 42, n. 1, p. 11, Nov 2017. ISSN 1573-689X. Disponível em: <<https://doi.org/10.1007/s10916-017-0855-8>>.

CHOWDHURY, M.; ISLAM, R.; GAO, J. Robust ear biometric recognition using neural network. In: *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. [S.l.: s.n.], 2017. p. 1855–1859. ISSN 2158-2297.

CONAIRE, C. O.; O’CONNOR, N. E.; SMEATON, A. F. Detector adaptation by maximising agreement between independent data sources. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2007. p. 1–6.

CORTES, C.; VAPNIK, V. Support-vector networks. *Mach. Learn.*, Kluwer Academic Publishers, Norwell, MA, USA, v. 20, n. 3, p. 273–297, set. 1995. ISSN 0885-6125. Disponível em: <<https://doi.org/10.1023/A:1022627411411>>.

CUN, Y. L.; BOSER, B.; DENKER, J. S.; HOWARD, R. E.; HABBARD, W.; JACKEL, L. D.; HENDERSON, D. Advances in neural information processing systems 2. In: TOURETZKY, D. S. (Ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990. cap. Handwritten Digit Recognition with a Back-propagation Network, p. 396–404. ISBN 1-55860-100-7. Disponível em: <<http://dl.acm.org/citation.cfm?id=109230.109279>>.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. [S.l.: s.n.], 2005. v. 1, p. 886–893 vol. 1. ISSN 1063-6919.

DAMER, N.; FUHRER, B. Ear recognition using multi-scale histogram of oriented gradients. In: *2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. [S.l.: s.n.], 2012. p. 21–24.

DARAMOLA, S.; OLUWANINYO, O. Automatic ear recognition system using back propagation neural network. *International Journal of Video Image Processing and Network Security IJVIPNS-IJENS*, v. 11, 01 2011.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09*. [S.l.: s.n.], 2009.

EBERMAM, E.; KROHLING, R. A. Uma introdução compreensiva às redes neurais convolucionais: Um estudo de caso para reconhecimento de caracteres alfabéticos. *Revista de Sistemas de Informação da FSMA*, p. 49–59, 2018.

FERREIRA, L. D. *Técnicas de aprendizado de máquina aplicadas à identificação de perfis de aprendizado em um ambiente real de ensino*. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2016.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, n. 2, p. 179–188, 1936. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x>>.

FONSECA, M. S. da. *Segmentação de Imagem*. 1998. Disponível em: <<http://www2.ic.uff.br/~aconci/limiarizacao.htm>>.

FUKUNAGA, K. *Introduction to Statistical Pattern Recognition (2Nd Ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990. ISBN 0-12-269851-7.

FUKUNAGA, K.; NARENDRA, P. M. A branch and bound algorithm for computing k-nearest neighbors. *IEEE Transactions on Computers*, C-24, n. 7, p. 750–753, July 1975. ISSN 2326-3814.

GALDAMEZ, P.; RAVEANE, W.; GONZALEZ, A. A brief review of the ear recognition process using deep neural networks. *Journal of Applied Logic*, v. 24, 11 2016.

GANESH, A. *Deep Learning Reading Group: SqueezeNet*. 2016. Disponível em: <<https://www.kdnuggets.com/2016/09/deep-learning-reading-group-squeezenet.html>>.

GHOUALMI, L.; DRAA, A.; CHIKHI, S. An ear biometric system based on artificial bees and the scale invariant feature transform. *Expert Systems with Applications*, v. 57, 03 2016.

GONZALEZ, E.; ALVAREZ, L.; MAZORRA, L. *AMI Ear Database*. 2012. Disponível em: <[http://ctim.ulpgc.es/research\\_works/ami\\_ear\\_database/#basic](http://ctim.ulpgc.es/research_works/ami_ear_database/#basic)>.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X.

GU, J.; WANG, Z.; KUEN, J.; MA, L.; SHAHROUDY, A.; SHUAI, B.; LIU, T.; WANG, X.; WANG, G.; CAI, J.; CHEN, T. Recent advances in convolutional neural networks. *ArXiv*, abs/1512.07108, 2015.

GUO, Y.; XU, Z. Ear recognition using a new local matching approach. In: *15th IEEE International Conference on Image Processing*. [S.l.: s.n.], 2008. p. 289–292.

GUTIERREZ, L.; MELIN, P.; LOPEZ, M. Modular neural network integrator for human recognition from ear images. In: *The 2010 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2010. p. 1–5.

HAI-LONG, Z.; ZHI-CHUN, M. Combining wavelet transform and orthogonal centroid algorithm for ear recognition. In: *2009 2nd IEEE International Conference on Computer Science and Information Technology*. [S.l.: s.n.], 2009. p. 228–231.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2nd. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. ISBN 0132733501.

HURLEY, D.; CARTER, J.; NIXON, M. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, v. 98, p. 491–512, 06 2005.

- HURLEY, D. J.; NIXON, M. S.; CARTER, J. N. Force field energy functionals for image feature extraction. *Image and Vision Computing*, v. 20, n. 5, p. 311 – 317, 2002. ISSN 0262-8856. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0262885602000033>>.
- HYPRAPTIVE. *Find the Bears: dlib*. 2017. Disponível em: <<https://hypraptive.github.io/2017/02/02/find-the-bears-dlib.html>>.
- IANDOLA, F. N.; MOSKEWICZ, M. W.; ASHRAF, K.; HAN, S.; DALLY, W. J.; KEUTZER, K. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. *CoRR*, abs/1602.07360, 2016. Disponível em: <<http://arxiv.org/abs/1602.07360>>.
- JAIN, A. K.; KUMAR, A. Biometric recognition: an overview. In: *Second generation biometrics: The ethical, legal and social context*. [S.l.]: Springer, 2012. p. 49–79.
- JUNIOR, P. R. S. C. *Avaliação de arquiteturas de Redes Neurais CONvolucionais para classificação de imagens de gráficos*. Dissertação (Mestrado em Ciência da Computação) — Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, 2018.
- KARPATHY, A. *Cs231n convolutional neural networks for visual recognition*. 2016. Disponível em: <<http://cs231n.github.io/convolutional-networks/>>.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143. ISBN 1-55860-363-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=1643031.1643047>>.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGESS, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (Ed.). *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012. p. 1097–1105. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>>.
- KUMAR, A.; WU, C. Automated human identification using ear imaging. In: *Pattern Recognition*. [S.l.: s.n.], 2012. v. 41, n. 5.
- KURNIAWAN, F.; SHAFRY, M.; RAHIM, M. A review on 2d ear recognition. In: *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*. [S.l.: s.n.], 2012. p. 204–209.
- KWON, S.-K.; JUNG, H.-S.; BAEK, W.-K.; KIM, D. Classification of forest vertical structure in south korea from aerial orthophoto and lidar data using an artificial neural network. *Applied Sciences*, v. 7, n. 10, 2017. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/7/10/1046>>.
- LANNARELLI, A. *Ear identification. Forensic identification series*. [S.l.]: California: Paramount Publishing Company, Fremont, 1989.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. In: *Nature*. [S.l.: s.n.], 2015. v. 521, p. 436–444.
- LEI, J.; YOU, X.; ABDEL-MOTTALEB, M. Automatic ear landmark localization, segmentation, and pose classification in range images. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, v. 46, n. 2, p. 165–176, Feb 2016. ISSN 2168-2216.

- LIMA, E. S. de. *Aula 16 - K-Nearest Neighbor (KNN)*. 2012. Disponível em: <[http://edirlei.3dgb.com.br/aulas/ia\\_2012\\_1/IA\\_Aula\\_16\\_KNN.pdf](http://edirlei.3dgb.com.br/aulas/ia_2012_1/IA_Aula_16_KNN.pdf)>.
- LIN, Y.; SAKR, C.; KIM, Y.; SHANBHAG, N. Predictivenet: An energy-efficient convolutional neural network via zero prediction. In: *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. [S.l.: s.n.], 2017. p. 1–4. ISSN 2379-447X.
- LIU, H.; LIU, D. Notice of retraction: Improving adaboost ear detection with skin-color model and multi-template matching. In: *2010 3rd International Conference on Computer Science and Information Technology*. [S.l.: s.n.], 2010. v. 8, p. 106–109.
- LORENA, A. C.; CARVALHO, A. C. P. L. F. de. Uma introdução às support vector machines. In: *Revista de Informática Teórica e Aplicada*. [S.l.: s.n.], 2007.
- LOTTE, F.; BOUGRAIN, L.; CICHOCKI, A.; CLERC, M.; CONGEDO, M.; RAKOTOMAMONJY, A.; YGER, F. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, IOP Publishing, v. 15, n. 3, p. 031005, apr 2018. Disponível em: <<https://doi.org/10.1088/1741-2552/15/3/031005>>.
- MANNING, C.; RAGHAVAN, P.; SCHÜTZ, H. *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.
- MARTINS, V. E. *Deep Learning para Classificação de Imagens*. Monografia (Bacharel em Ciência da Computação) — Universidade Estadual de Londrina, Londrina, 2018.
- MASSUCATTO, J. D. E. *Aplicação de Conceitos de Redes Neurais Convolucionais na Classificação de Imagens de Folhas*. Monografia (Bacharel em Engenharia da Computação) — Universidade Tecnológica Federal do Paraná, Pato Branco, 2018.
- MAYYA, A.; SAI, M. Human recognition based on ear shape images using pca-wavelets and different classification methods. *Journal of Medical Devices*, v. 1, 09 2016.
- MENOTTI, D. *Classificadores Lineares*. 2011. Disponível em: <<https://slideplayer.com.br/slide/356517/>>.
- MITCHELL, T. The discipline of machine learning. *School of Computer Science*. [S.l.]: Carnegie Mellon University, Pittsburgh, PA, 2006.
- MITCHELL, T. M. *Machine Learning*. [S.l.]: McGraw Hill, 1997.
- MONTEIRO, R. B. *Comparação de técnicas de aprendizado de máquina para predição da disponibilidade de bicicletas no projeto biciletar*. Monografia (Bacharel em Sistemas de Informação) — Universidade Federal do Ceará, Quixadá, 2016.
- NOSRATI, M. S.; FAEZ, K.; FARADJI, F. Using 2d wavelet and principal component analysis for personal identification based on 2d ear structure. In: *2007 International Conference on Intelligent and Advanced Systems*. [S.l.: s.n.], 2007. p. 616–620.
- NWANKPA, C.; IJOMAH, W.; GACHAGAN, A.; MARSHALL, S. Activation functions: Comparison of trends in practice and research for deep learning. *CoRR*, abs/1811.03378, 2018. Disponível em: <<http://arxiv.org/abs/1811.03378>>.

- OLIVAS, E. S.; GUERRERO, J. D. M.; SOBER, M. M.; BENEDITO, J. R. M.; LOPEZ, A. J. S. *Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques - 2 Volumes*. Hershey, PA: Information Science Reference - Imprint of: IGI Publishing, 2009. ISBN 1605667668, 9781605667669.
- OTHMAN, R. N.; ALIZADEH, F.; SUTHERLAND, A. A novel approach for occluded ear recognition based on shape context. In: *2018 International Conference on Advanced Science and Engineering (ICOASE)*. [S.l.: s.n.], 2018. p. 93–98.
- PACHECO, A. G. C. *K vizinhos mais próximos - KNN*. 2019. Disponível em: <<http://computacaointeligente.com.br/algoritmos/k-vizinhos-mais-proximos/>>.
- PAVLOVSKY, V. *Introduction To Convolutional Neural Networks*. 2017. Disponível em: <<https://www.vaetas.cz/blog/intro-convolutional-neural-networks>>.
- PISTORI, H. *Tecnologia Adaptativa em Engenharia de Computação: estado da arte e aplicações*. Tese (Doutorado em Ciência da Computação) — Escola Politécnica da Universidade de São Paulo, São Paulo, 2003.
- PONTI, M. A.; COSTA, G. B. P. *Como funciona o Deep Learning*. 2017. Disponível em: <[http://conteudo.icmc.usp.br/pessoas/moacir/papers/Ponti\\_Costa\\_Como-funciona-o-Deep-Learning\\_2017.pdf](http://conteudo.icmc.usp.br/pessoas/moacir/papers/Ponti_Costa_Como-funciona-o-Deep-Learning_2017.pdf)>.
- RAPOSO, R. G. *Reconhecimento Biométrico da Orelha Análise de Factores de Qualidade de Imagem*. Dissertação (Mestrado em Engenharia Informática) — Universidade da Beira Interior, Covilhã, 2012.
- RATSCH, G. A brief introduction into machine learning. In: *In: 21st Chaos Communication Congress*. [S.l.: s.n.], 2014.
- RODRIGUES, D. K. de S. *Identificação biométrica através de imagens de orelhas*. Monografia (Bacharel em Ciência da Computação) — Universidade Federal de Ouro Preto, Ouro Preto, 2018.
- ROSS, A. An introduction to multibiometrics. In: *2007 15th European Signal Processing Conference*. [S.l.: s.n.], 2007. p. 20–24.
- ROZA, F. S. da. *Aprendizagem de máquina para apoio à tomada de decisão em vendas do varejo utilizando registros de vendas*. Monografia (Bacharel em Engenharia de Controle e Automação) — Universidade Federal de Santa Catarina, Florianópolis, 2016.
- RUDER, S. An overview of gradient descent optimization algorithms. *ArXiv*, 2016.
- SAMPAIO, O. M. *Diferenciação entre rums cubanos e não cubanos*. Dissertação (Mestrado em Ciências - Química Analítica) — Instituto de Química, Universidade de São Paulo, São Carlos, 2006.
- SARANGI, P. P.; MISHRA, B. S. P.; DEHURI, S. Ear recognition using pyramid histogram of orientation gradients. In: *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*. [S.l.: s.n.], 2017. p. 590–595.
- SHAKHNAROVICH, G.; DARRELL, T.; INDYK, P. *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)*. [S.l.]: The MIT Press, 2006. ISBN 026219547X.

- SILVA, R. E. V. d. *Um Estudo Comparativo Entre Redes Neurais Convolucionais para a Classificação de Imagens*. Monografia (Bacharel em Sistema de Informação) — Universidade Federal do Ceará, Quixadá, 2018.
- SINGHAL, H. *Convolutional Neural Network with TensorFlow implementation*. 2017. Disponível em: <<https://medium.com/data-science-group-iitr/building-a-convolutional-neural-network-in-python-with-tensorflow-d251c3ca8117>>.
- SOBRAL, J. L. *Segmentação de Imagem*. 2002. Disponível em: <<http://gec.di.uminho.pt/lesi/vpc0607/Aula07Segmentaç~ao.pdf>>.
- SOLAZZO, A.; SOZZO, E. D.; ROSE, I. D.; SILVESTRI, M. D.; DURELLI, G. C.; SANTAMBROGIO, M. D. Hardware design automation of convolutional neural networks. In: *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. [S.l.: s.n.], 2016. p. 224–229. ISSN 2159-3477.
- SOLOMON, C.; BRECKON, T. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. 1st. ed. [S.l.]: Wiley Publishing, 2011. ISBN 0470844736, 9780470844731.
- SONI, K.; GUPTA, S. K.; KUMAR, U.; AGRWAL, S. L. A new gabor wavelet transform feature extraction technique for ear biometric recognition. In: *2014 6th IEEE Power India International Conference (PIICON)*. [S.l.: s.n.], 2014. p. 1–3.
- SOUSA, P. H. T. de O. *Classificação utilizando técnicas de aprendizado estatístico: Estudo de casos*. Monografia (Bacharel em Estatística) — Universidade de Brasília, Brasília, 2016.
- SOUZA, A. *Validação Cruzada: Conceito e Exemplo em R*. 2016. Disponível em: <<https://pessoalex.wordpress.com/2019/04/16/validacao-cruzada-conceito-e-exemplo-em-r/>>.
- STALLKAMP, J.; SCHLIPSING, M.; SALMEN, J.; IGEL, C. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*, v. 32, p. 323 – 332, 2012. ISSN 0893-6080. Selected Papers from IJCNN 2011. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0893608012000457>>.
- STEINBRUCH, D. *Um estudo de algoritmos para a classificação automática para textos utilizando naives-bayes*. Dissertação (Mestrado em Ciência da Computação) — Pontifícia Universidade Católica, Rio de Janeiro, 2006.
- SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: *Computer Vision and Pattern Recognition (CVPR)*. [s.n.], 2015. Disponível em: <<http://arxiv.org/abs/1409.4842>>.
- TARIQ, A.; ANJUM, M. A.; AKRAM, M. U. Personal identification using computerized human ear recognition system. In: *Proceedings of 2011 International Conference on Computer Science and Network Technology*. [S.l.: s.n.], 2011. v. 1, p. 50–54.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition, Fourth Edition*. 4th. ed. Orlando, FL, USA: Academic Press, Inc., 2008. ISBN 1597492728, 9781597492720.
- TIAN, L.; MU, Z. Ear recognition based on deep convolutional network. In: *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. [S.l.: s.n.], 2016. p. 437–441.

- UOL. *Orelha Externa*. 2015. Disponível em: <<https://drauziovarella.uol.com.br/corpo-humano/orelha-externa/>>.
- VARGAS, A. C. G.; CARVALHO, A. M. P.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: *Proceedings of the XXIX Conference on Graphics, Patterns and Images*. [S.l.: s.n.], 2016. p. 1–4.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. [S.l.: s.n.], 2001. v. 1, p. I–I. ISSN 1063-6919.
- WANG, X.; XIA, H.; WANG, Z. The research of ear identification based on improved algorithm of moment invariant. In: *2010 Third International Conference on Information and Computing*. [S.l.: s.n.], 2010. v. 1, p. 58–60. ISSN 2160-7451.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN 0123748569, 9780123748560.
- YAN, P.; BOWYER, K. W. Biometric recognition using 3d ear shape. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 29, n. 8, p. 1297–1308, 2007.
- YAQUBI, M.; FAEZ, K.; MOTAMED, S. Ear recognition using features inspired by visual cortex and support vector machine technique. In: IEEE. *2008 International Conference on Computer and Communication Engineering*. [S.l.], 2008. p. 533–537.
- ZHANG, H.-J.; MU, Z.-C.; QU, W.; LIU, L.-M.; ZHANG, C.-Y. A novel approach for ear recognition based on ica and rbf network. In: *2005 International Conference on Machine Learning and Cybernetics*. [S.l.: s.n.], 2005. v. 7, p. 4511–4515 Vol. 7.