

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS  
ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO  
GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO



**PROPOSTA DE UMA BASE DE IMAGENS MULTIESPECTRAIS DAS MÃOS E  
SUA APLICAÇÃO EM CENÁRIOS DE CLASSIFICAÇÃO**

Paulo Vitor Santana da Silva

GOIÂNIA

2020

Paulo Vitor Santana da Silva

**PROPOSTA DE UMA BASE DE IMAGENS MULTIESPECTRAIS DAS MÃOS E  
SUA APLICAÇÃO EM CENÁRIOS DE CLASSIFICAÇÃO**

Trabalho de Conclusão de Curso apresentado à Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do grau de Bacharel em Engenharia de Computação.

Orientador(a): Prof. Dr. Arlindo Rodrigues Galvão Filho

GOIÂNIA

2020

Paulo Vitor Santana da Silva

**PROPOSTA DE UMA BASE DE IMAGENS MULTIESPECTRAIS DAS MÃOS E  
SUA APLICAÇÃO EM CENÁRIOS DE CLASSIFICAÇÃO**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do grau de Bacharel em Engenharia de Computação e aprovado em sua forma final pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, em 02 de dezembro de 2020.

---

Prof. Ma. Ludmilla Reis Pinheiro dos Santos  
Coordenador(a) de Trabalho de Conclusão de Curso

Banca examinadora:

---

Orientador(a): Prof. Dr. Arlindo Rodrigues Galvão  
Filho

---

Prof. Dr. Rafael Viana de Carvalho

---

Prof. Dr. Ricardo Augusto Pereira Franco

GOIÂNIA

2020

## RESUMO

Este trabalho tem por finalidade apresentar um estudo sobre classificação de imagens das mãos no espectro visível e infravermelho próximo (NIR), propondo, para isso, uma base de dados para tal. A base de dados em questão possui imagens capturadas de 80 indivíduos, sendo 20 tipos de imagens para cada qual, separadas em dez no espectro visível e dez no espectro NIR. Composta por imagens do dorso, da palma e do pulso, capturadas em duas alturas diferentes. Para a criação da base de dados foi construído um sistema de aquisição de imagens multiespectrais. A câmera utilizada foi a estação JAI AD-080E, que realiza a captura simultânea no espectro visível (entre 400 nm e 700 nm) e NIR (entre 700 nm e 1000 nm). Com a intenção de apresentar um *benchmark* inicial sobre a base de dados, foram realizadas algumas simulações com alguns classificadores, são eles: *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) e Árvore de decisão. Foram realizadas classificações utilizando imagens originais e pré-processadas, utilizando diferentes parametrizações nos modelos de classificação. Por meio das experimentações utilizando os classificadores citados, foi possível verificar bons resultados utilizando os modelos SVM e KNN. Para o modelo SVM, entre todos os tipos de imagens do espectro NIR, foi obtida uma acurácia média de 99,5%, para o espectro visível foi obtida uma acurácia média de 99,19%. Para o modelo KNN, entre todos os tipos de imagens do espectro NIR, foi obtida uma acurácia média de 99,5%, para o espectro visível foi obtida uma acurácia média de 98,63%. Em relação à base de dados, a grande quantidade de tipos de imagens em ambos os espectros (visível e NIR) e alta resolução utilizada, em comparação com outras bases de dados de mesmo propósito, se mostrou viável e bastante relevante ao estado da arte.

**Palavras-chave:** Base de dados. Mãos. NIR. Biometria. Aprendizado de máquina.

## ABSTRACT

This work aims to propose a database composed by images of hands in the visible and near infrared (NIR) spectrum, which can be used in classification applications, pattern recognition and any other relevant applications. The database in question has images obtained from 80 individuals, with 20 types of images for each, separated into 10 in the visible spectrum and 10 in the NIR spectrum. Composed by images of the back of hand, palm and wrist, captured in two different heights. For the creation of the database, a Multispectral Image Acquisition System was built. The camera used was the JAI AD-080E station, that makes simultaneous capture in the visible spectrum (between 400 nm and 700 nm) and NIR (between 700 nm and 1000 nm). To present an initial benchmark about the database, some simulations were performed with some classifiers, they are: Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Decision Tree. The classifications were realized from original and preprocessed images, using different parameterizations in the classification models. Through the experiments using the mentioned classifiers, it was possible to verify good results using the SVM and KNN models. For the SVM model, among all types of images in the NIR spectrum, an average precision of 99.5% was obtained and for the visible spectrum an average precision of 99.19% was obtained. For the KNN model, among all types of images in the NIR spectrum, an average precision of 99.5% was obtained and for the visible spectrum an average precision of 98.63% was obtained. About the database, the large number of types of images in both the specter visible and NIR and a high resolution used, in comparison with other databases with the same purpose, proved to be feasible and very relevant to the state of art.

**Keywords:** Database. Hands. NIR. Biometry. Machine Learning

## LISTA DE FIGURAS

Figura 1 – Características extraídas por meio do SURF em imagem de veias das mãos . . . . .	11
Figura 2 – Exemplo de agrupamentos criados pelo <i>k-means</i> . . . . .	12
Figura 3 – Criação do vocabulário por meio do <i>Bag of Featuers</i> . . . . .	13
Figura 4 – Criação do histograma de características por meio do <i>Bag of Featuers</i> . . . . .	13
Figura 5 – Exemplos de possíveis hiperplanos encontrados por meio do SVM . . . . .	14
Figura 6 – Exemplos de teste utilizando o KNN . . . . .	15
Figura 7 – Exemplo de estrutura da Árvore de decisão . . . . .	15
Figura 8 – Espectros Visível e NIR . . . . .	17
Figura 9 – Dimensões da estrutura . . . . .	18
Figura 10 – Ilustração da visão geral da estrutura . . . . .	19
Figura 11 – Lâmpada de filamento palito . . . . .	19
Figura 12 – Fita de LEDs . . . . .	20
Figura 13 – Imagens capturadas com a câmera aproximada da mão no espectro visível e NIR . . . . .	20
Figura 14 – Imagens capturadas com a câmera afastada da mão no espectro visível e NIR . . . . .	21
Figura 15 – Distribuição das Idades dos Indivíduos . . . . .	22
Figura 16 – Distribuição dos Sexos dos Indivíduos . . . . .	22
Figura 17 – Diagrama dos passos do pré-processamento das imagens . . . . .	24
Figura 18 – Resultados da Classificação utilizando o SVM . . . . .	27
Figura 19 – Resultados da Classificação utilizando o KNN . . . . .	28
Figura 20 – Resultados da Classificação utilizando a Árvore de Decisão . . . . .	28

## LISTA DE TABELAS

Tabela 1 – Variância explicada pelas componentes principais com imagens originais	23
Tabela 2 – Variância explicada pelas componentes principais com imagens pré-processadas . . . . .	25
Tabela 3 – Comparativo entre a variância explicada pelas componentes principais .	25
Tabela 4 – Parametrizações SVM . . . . .	25
Tabela 5 – Parametrizações KNN . . . . .	25
Tabela 6 – Parametrizações Árvore de Decisão . . . . .	26

## LISTA DE ABREVIATURAS E SIGLAS

CLAHE	<i>Contrast-limited Adaptive Histogram</i>
GDI	<i>Gini's Diversity Index</i>
KNN	<i>K-Nearest Neighbors</i>
NIR	<i>Near Infrared</i>
PCA	<i>Principal Component Analysis</i>
RGB	<i>Red, Blue and Green</i>
ROI	<i>Region of Interest</i>
SURF	<i>Speeded up Robust Features</i>
SVM	<i>Support Vector Machine</i>



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>9</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>11</b>
2.1	<i>Speeded up Robust Features</i> (SURF)	11
2.2	<i>K-means</i>	11
2.3	<i>Bag of Features</i> (BoF)	12
2.4	<i>Support Vector Machine</i> (SVM)	13
2.5	<i>K-Nearest Neighbors</i> (KNN)	14
2.6	Árvore de decisão (DTREE)	15
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>17</b>
3.1	Sistema de Aquisição de Imagens	17
3.2	Base de Dados	19
3.3	Análise Exploratória	21
3.4	Pré-processamento das Imagens	22
3.4.1	Extração de ROI	23
3.4.2	Suavização	23
3.4.3	Equalização de Histograma	23
3.4.4	Extração de Características	24
3.5	Análise Exploratória Após Pré-processamento	24
3.5.1	Comparativo	24
3.6	Modelos Propostos	25
3.7	Avaliação dos Modelos	26
<b>4</b>	<b>RESULTADOS OBTIDOS</b>	<b>27</b>
<b>5</b>	<b>CONCLUSÃO</b>	<b>30</b>
	<b>REFERÊNCIAS</b>	<b>31</b>

## 1 INTRODUÇÃO

O estudo de sistemas de classificação e reconhecimento de padrões tem tido grande destaque, como em aplicações de autenticação com biometria, que proporcionam uma camada de segurança mais confiável em sistemas nos quais acessos a informações venham a ser restringidos. Um dos padrões biométricos estudados exaustivamente nos últimos anos é o das mãos, no qual podem ser extraídas características como impressão palmar, impressões digitais e geometria das mãos.

Atualmente a impressão digital é a característica mais difundida quando se fala de identificação biométrica de indivíduos, entretanto, existem algumas limitações. As impressões digitais podem sofrer alterações ao longo do tempo, sejam por queimaduras, cicatrizes ou até mesmo desgaste por idade. Contudo, novas tecnologias vêm emergindo e se provando mais seguras e suficientemente versáteis para se tornarem também um padrão (Jain; Ross; Prabhakar, 2004). Como a classificação biométrica baseada no padrão de veias das mãos (Jain; Flynn; Ross, 2007).

Sistemas de classificação de indivíduos comumente utilizam inteligência artificial para a determinação dos indivíduos. O processo da criação do sistema consiste em ter dados para que seja realizado o treinamento do modelo, outros dados para que sejam realizados o teste e a validação do modelo. Essa necessidade de dados muitas das vezes é um empecilho na criação do sistema, pode ser oneroso criar a própria base de dados, as bases de dados existentes podem não atender a necessidade da aplicação, a quantidade de dados é insuficiente, entre outros motivos. Portanto, em alguns casos torna-se viável utilizar uma base de dados já construída que atenda as especificações do projeto.

Existem algumas bases de dados neste segmento de classificação de indivíduos utilizando as mãos como *Tongji Contactless Palmprint Dataset* (Zhang et al., 2017), *VERA Spoofing Palmvein Database* (Tome; Marcel, 2015) e *PUT Vein Dataset* (Kabaciński; Kowalski, 2011).

O *Tongji Contactless Palmprint Dataset* (Zhang et al., 2017) possui imagens de 300 voluntários capturadas no espectro visível, coletadas em duas sessões, dez imagens de cada palma dos indivíduos foram obtidas em cada sessão, totalizando 12.000 imagens, sendo 40 imagens por indivíduo. Nesta base foi utilizado um sistema *contactless* (sem contato), ou seja, o indivíduo posiciona a mão em algum lugar acima da câmera, que captura a imagem. As imagens desta base de dados estão, em geral em distâncias diferentes em relação à câmera, algumas estão próximas, cortando metade dos dedos, outras estão distante, mostrando grande parte dos dedos. Apesar de possuir uma grande quantidade de amostras, a base de dados proposta (Zhang et al., 2017) possui baixa resolução nas imagens, apenas 600x800 pixels.

O *VERA Spoofing Palmvein Database* (Tome; Marcel, 2015) possui imagens de

110 voluntários capturadas no espectro NIR, coletadas em duas sessões, cinco imagens de cada palma dos indivíduos foram obtidas em cada sessão, totalizando 2.200 imagens, sendo 20 imagens por indivíduo. Esta base, assim como a proposta por (Zhang et al., 2017) utiliza do sistema *contactless* para aquisição. Logo, possui a mesma característica de as amostras estarem em distâncias variáveis. Esta base de dados, em comparação com a proposta por (Zhang et al., 2017), possui uma quantidade de amostras inferior, além de também possuir baixa resolução nas imagens, apenas 480x680 pixels.

O *PUT Vein Dataset* (Kabaciński; Kowalski, 2011) possui imagens de 50 voluntários capturadas no espectro NIR, coletadas em três sessões. Esta base possui 1200 imagens da palma e 1200 imagens do pulso dos indivíduos, totalizando 2400 imagens, sendo 48 amostras por indivíduo. Possui resolução de 1280x960 pixels. Além de possuir dois tipos de imagens, da palma e pulso, possui uma resolução maior que as utilizada nas bases propostas por (Zhang et al., 2017) e (Tome; Marcel, 2015). Entretanto, possui uma quantidade pequena de imagens.

O objetivo deste trabalho é propor uma base de dados com imagens das mãos para fins de treinamento de modelos de classificação e reconhecimento de padrões. A base de dados em questão é composta por imagens de ambas as mãos utilizando duas alturas fixas para a câmera, uma mais distante, contemplando a mão, os dedos e o pulso, e uma mais próxima, contemplando a mão e parte dos dedos. Foram capturadas imagens da palma, pulso e dorso, tanto no espectro visível (RGB) quanto no infravermelho próximo (NIR). A base possui imagens capturadas de 80 voluntários, sendo composta por 20 tipos de imagens, sendo dez no visível e dez no NIR, possuindo 5 exemplares de cada tipo, totalizando 8000 imagens. A resolução das imagens é de 1024x768 pixels.

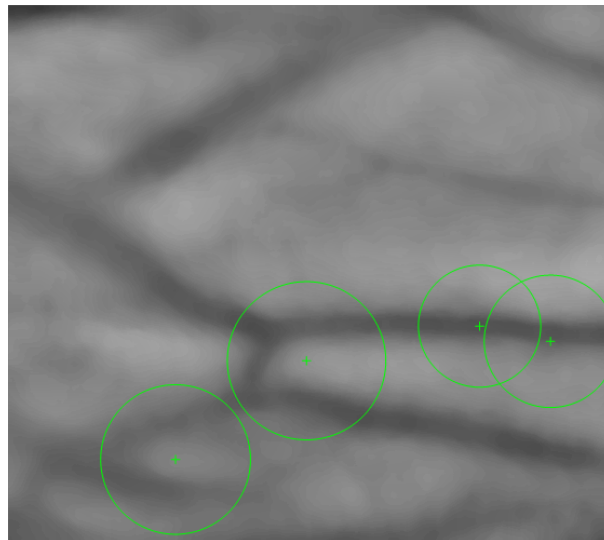
Para proporcionar uma análise exploratória sobre a base de dados foram realizadas simulações de classificação com todos os tipos de imagens utilizando três algoritmos de *machine learning* diferentes, são eles: *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), e *Árvore de decisão* (DTREE). Estas foram realizadas utilizando imagens originais e imagens pré-processadas.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 *Speeded up Robust Features (SURF)*

O algoritmo SURF foi proposto por Bay, Tuytelaars e Gool (2008). Trata-se de um “detector e descritor invariante de escala e rotação”(Bay et al., 2008). Este é utilizado para a extração e descrição de pontos de interesse em uma imagem. Estes resultados são obtidos utilizando imagens integrais para a convolução das imagens. Através da matriz hessiana, são selecionados os pontos fortes dos principais detectores e descritores existentes. Um exemplo de características extraídas pelo SURF em um ROI extraído de uma imagem do dorso de uma mão no espectro NIR pode ser observado na Figura 1.

Figura 1 – Características extraídas por meio do SURF em imagem de veias das mãos



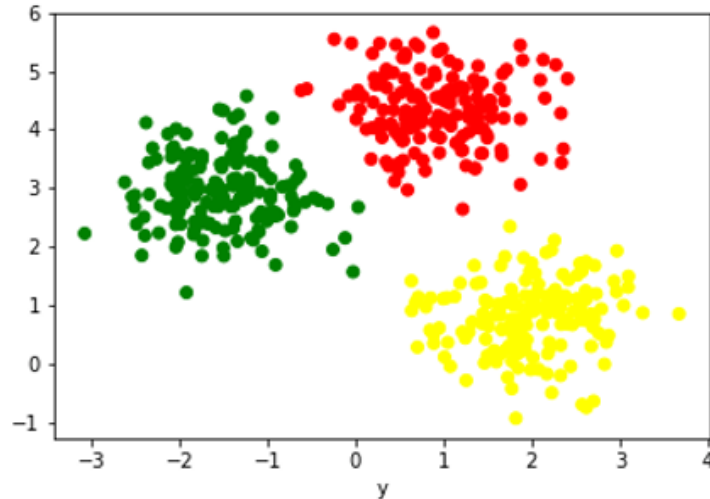
Fonte: Autoria Própria

### 2.2 *K-means*

O *K-means* é um algoritmo de *machine learning* com aprendizado não supervisionado vastamente utilizado para aplicações de agrupamento proposto por (MacQueen, 1967). O algoritmo em questão itera sobre os pontos de dados na tentativa de minimizar a soma das distâncias quadradas dentro de todos os agrupamentos baseando-se em alguma medida de similaridade. São definidos k-pontos aleatórios sobre o espaço, são os chamados centroides. Para cada dado calcula-se o centroide mais próximo e rotula esse dado como pertencente a esse centroide. Em seguida recalculam-se as coordenadas dos centroides, com base na distância média entre seus dados relacionados. Estas etapas são executadas até que alcance um limite de iterações ou satisfaça algum critério de erro. Na Figura 2 pode

ser visualizado um exemplo de agrupamentos propiciados pela execução do *k-means*, os pontos de mesmas cores representam amostras com mesmas classes.

Figura 2 – Exemplo de agrupamentos criados pelo *k-means*



Fonte: <https://cmdlinetips.com/2019/05/k-means-clustering-in-python>

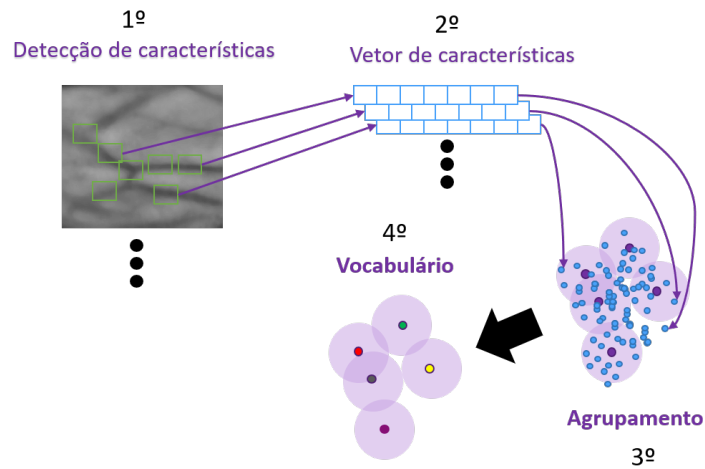
### 2.3 *Bag of Features (BoF)*

O *Bag of Features* é uma técnica utilizada para extração de características. Consiste na aplicação do algoritmo SURF e utilização do *K-means* para agrupamento das características encontradas. As representações de BoF tornaram-se populares para classificação de imagens baseada em conteúdo devido à sua simplicidade e bom desempenho (Nowak; Jurie; Triggs, 2006).

Primeiramente são detectadas e extraídas as características de todas as imagens do conjunto de treinamento utilizando o SURF. Estas características extraídas são submetidas ao algoritmo *K-means*, para que sejam agrupadas de acordo com medidas de similaridade abstraídas pelo algoritmo. Após a finalização da execução do agrupamento, de modo a diminuir a complexidade e a repetição de características semelhantes, são consideradas apenas os centroides de cada agrupamento de características, este conjunto de centroides é denominado ‘vocabulário’. Na Figura 3 pode ser visto um diagrama dos passos para a criação do vocabulário utilizando o BoF.

Após possuir o vocabulário, cada imagem é percorrida, tendo suas características detectadas e extraídas por meio do SURF, tendo suas características submetidas ao *K-means* para verificar a quais centroides do vocabulário suas características pertencem. Possuindo a quantidade de características em cada um dos centroides do vocabulário, é possível criar um histograma de características. Este histograma é comumente utilizado como vetor de características por algoritmos de classificação ou regressão. Na Figura 3

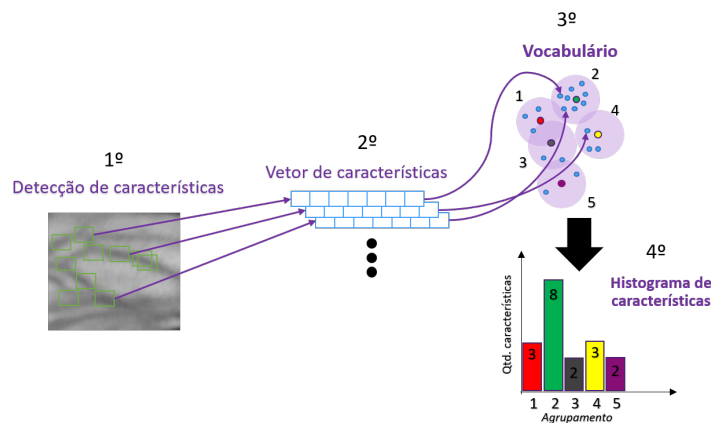
Figura 3 – Criação do vocabulário por meio do *Bag of Features*



Fonte: Autoria Própria

pode ser visto um diagrama dos passos para a criação do histograma de características utilizando o BoF.

Figura 4 – Criação do histograma de características por meio do *Bag of Features*



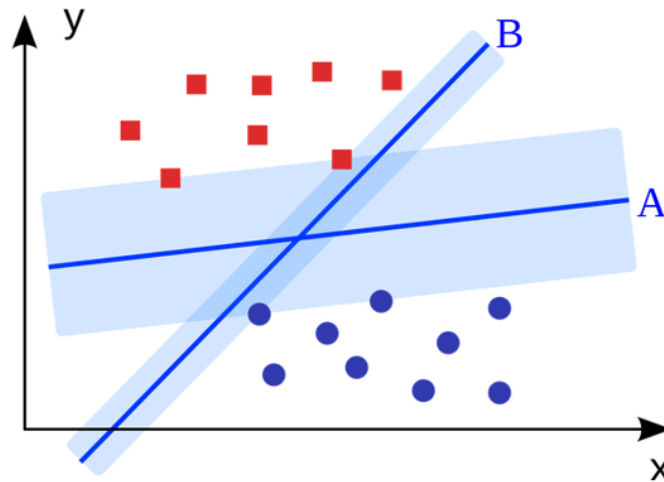
Fonte: Autoria Própria

## 2.4 *Support Vector Machine* (SVM)

O SVM é um algoritmo de *machine learning* com aprendizado supervisionado utilizado para classificação e regressão. A técnica consiste basicamente em tentar encontrar hiperplanos no espaço que separem corretamente as classes. Após encontrar os hiperplanos, são determinados os vetores de suporte, que são os pontos de dados mais próximos aos hiperplanos encontrados. Por fim, o algoritmo procura maximizar a distância entre o hiperplano e os vetores de suporte, de modo a selecionar o hiperplano ideal. Os resultados obtidos por meio dessa técnica são comparáveis e muitas vezes superiores aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs) (Braga,

2000; Haykin; Network, 2004). Na 5 podem ser observadas duas classes, a de quadrados vermelhos e a de círculos azuis. O hiperplano B é o hiperplano encontrado que separa as classes corretamente, o hiperplano A é o hiperplano ideal, após a maximização da distância entre os vetores de suporte e o hiperplano.

Figura 5 – Exemplos de possíveis hiperplanos encontrados por meio do SVM

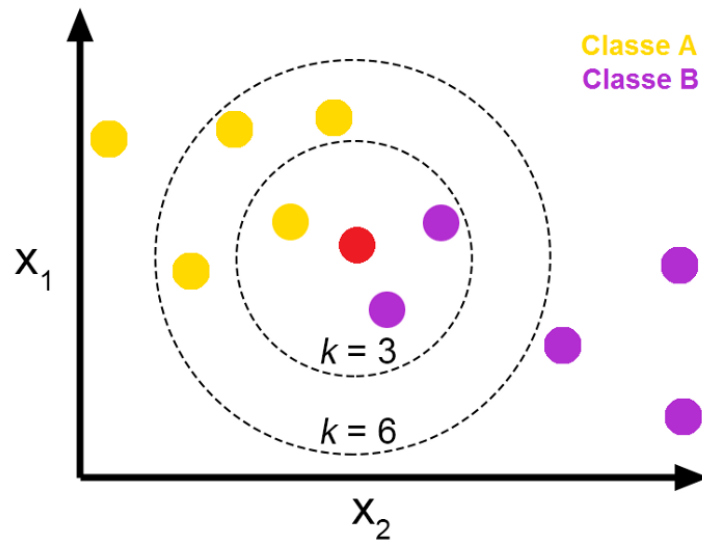


Fonte: <https://minerandodados.com.br/spotify-svm-python>

## 2.5 *K-Nearest Neighbors* (KNN)

O KNN é um algoritmo de *machine learning* com aprendizado supervisionado utilizado para classificação. A técnica consiste basicamente em, à partir de dados de treinamento, cujos rótulos são conhecidos, realizar o treinamento. O KNN trabalha com vetores no espaço, logo, após o treinamento, possuiremos no espaço vetores com seus respectivos rótulos. Para realizar a classificação de um dado com rótulo desconhecido, o KNN calcula a distância entre o vetor do dado desconhecido e os vetores dos dados conhecidos pelo modelo (dados de treinamento) seja utilizando distância Euclidiana, Manhattan, Minkowski, entre outras. Após possuir as distâncias entre o vetor do dado desconhecido e os vetores de todos os dados conhecidos pelo modelo, são então elegidos os  $K$  vizinhos mais próximos ao dado desconhecido no espaço. A classe a qual o dado desconhecido pertencerá será a predominante entre os  $K$  vizinhos mais próximos (dados conhecidos mais próximos) no espaço. O classificador KNN é bastante popular na comunidade de reconhecimento de padrões devido ao seu bom desempenho e simplicidade de uso (Chou; Shen, 2006). Na Figura 6 pode ser visto um exemplo, no qual possuímos 2 classes, classe A e classe B, e um dado desconhecido (ponto vermelho) submetido ao teste. No exemplo em questão, ao verificar a qual classe o ponto desconhecido pertence podemos perceber que, utilizando  $k = 3$ , pertenceria à classe B e utilizando  $k = 6$ , pertenceria à classe A.

Figura 6 – Exemplos de teste utilizando o KNN

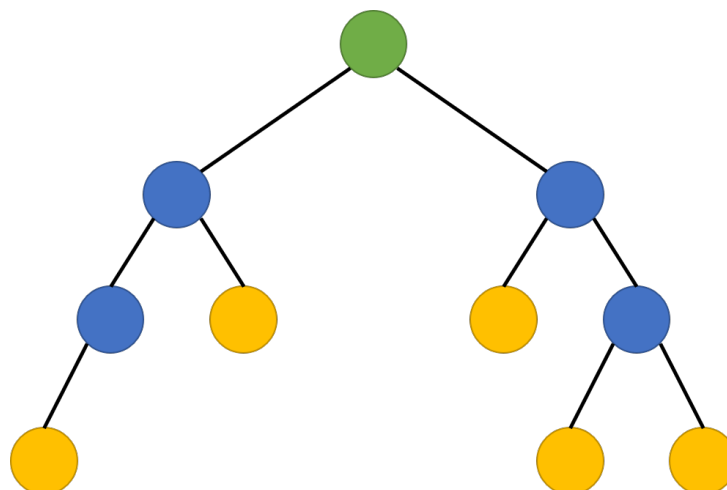


Fonte: <https://medium.com/brasil-ai/knn-k-nearest-neighbors-1-e140c82e9c4e>

## 2.6 Árvore de decisão (DTREE)

A Árvore de decisão é um algoritmo de *machine learning* com aprendizado supervisionado utilizado para classificação e regressão. A estrutura de uma Árvore de decisão é composta por nós e ramos. Os nós que não possuem descendentes são chamados de folhas, o primeiro nó da DTREE é chamado de raiz, o único que não possui ascendentes. Os ramos são as ligações parentais entre os nós. Na Figura 7 pode ser observado o nó verde como o nó raiz, os nós amarelos como folhas, e os azuis como nós.

Figura 7 – Exemplo de estrutura da Árvore de decisão



Fonte: Autoria Própria

A técnica consiste em dividir os dados com a realização de perguntas em relação às suas características. Cada nó equivale à uma pergunta, logo, existirá uma divisão para cada possível resposta para o questionamento do nó. A recomendação é a de fazer as



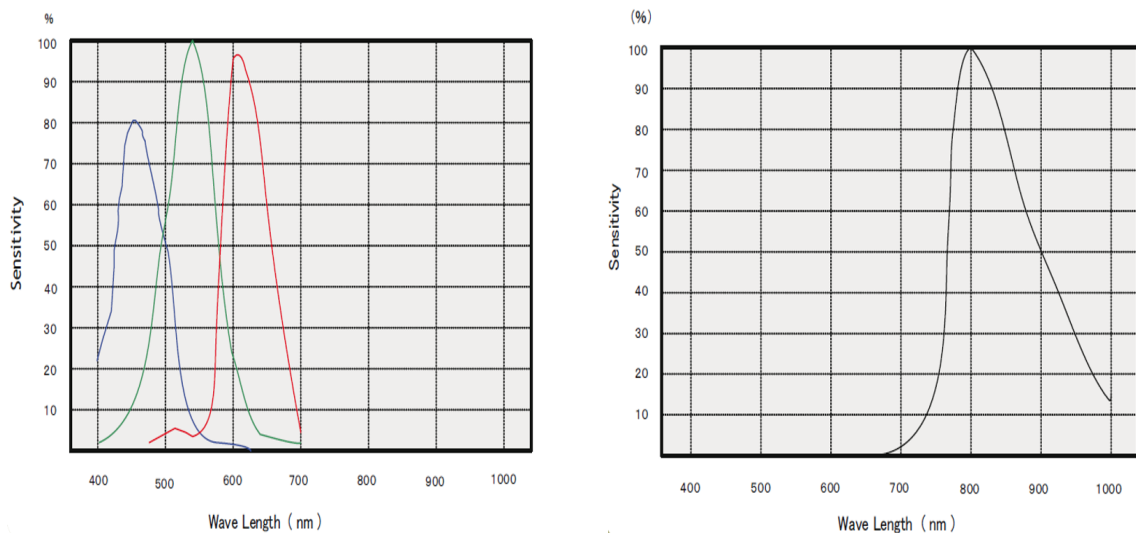
perguntas que melhor separem os dados. No processo de treinamento podem ser escolhidos entre diversos critérios para saber o quão boa foi a pergunta, em relação à quão bom foi a separação propiciada por ela. No trabalho em questão, o critério para escolha das melhores perguntas foi o Gini's Diversity Index (GDI).

### 3 MATERIAIS E MÉTODOS

#### 3.1 Sistema de Aquisição de Imagens

Para a aquisição das imagens das mãos no espectro visível e NIR foi construído um sistema de aquisição de imagens multiespectrais. A câmera utilizada foi a estação JAI AD-080E, que realiza a captura simultânea no espectro visível (entre 400 nm e 700 nm) e NIR (entre 700 nm e 1000 nm), como pode ser observado na Figura 8. Os comprimentos de onda nas da Figura 8(a) nas cores vermelho, verde e azul representam o comprimento de onda em relação à sensibilidade da câmera para o sistema de cores RGB respectivamente. Na Figura 8(b) o comprimento de onda na cor preta representa o comprimento de onda em relação à sensibilidade da câmera para o espectro NIR.

Figura 8 – Espectros Visível e NIR



(a) Comprimento de onda do espectro visível

(b) Comprimento de onda do espectro NIR

Fonte: [http://www.altavision.com.br/Arquivos/JAI/Datasheet/GigE/AD-080GE\\_datasheet.pdf](http://www.altavision.com.br/Arquivos/JAI/Datasheet/GigE/AD-080GE_datasheet.pdf)

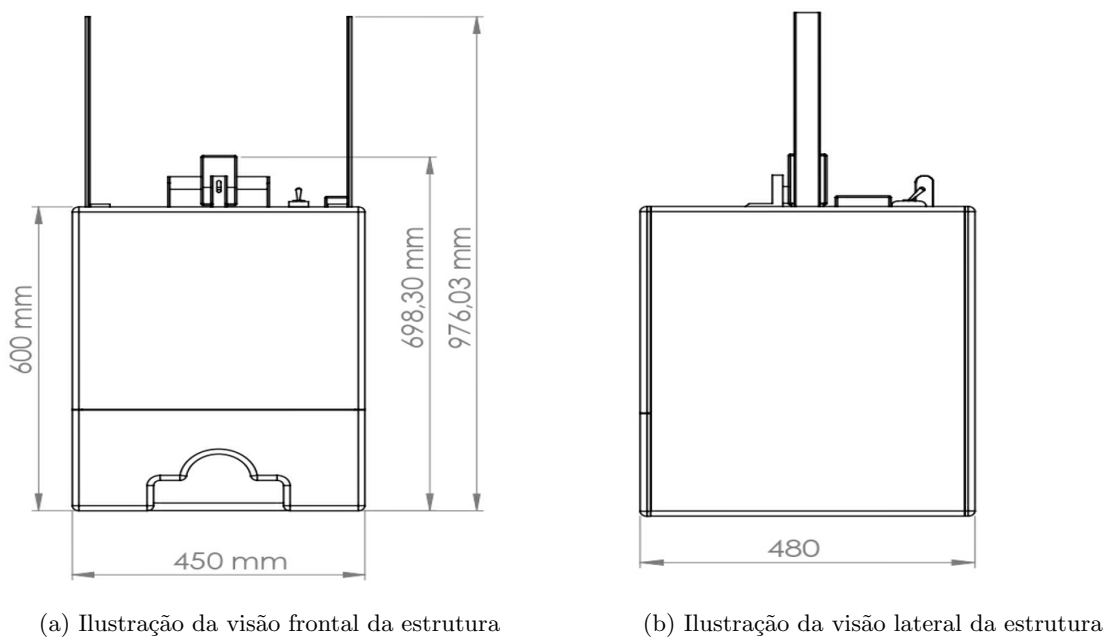
Com base em experimentos, foi percebido que para obtenção de imagens com a câmera aproximada da palma, pulso e dorso a distância entre a câmera e o objeto alvo da captura deveria ser de 200 mm. Bem como para a obtenção de imagens com a câmera afastada da palma e do dorso a distância deveria ser de 600 mm.

Com o intuito de criar uma estrutura de transporte facilitado e com maior leveza, foi escolhido placas do tipo *Medium-Density Fiberboard* (MDF) de 15 mm de espessura como material para sua construção. Tendo em vista a aplicação e a possibilidade de atender a diversidade da anatomia humana, foi definida uma estrutura cúbica com as seguintes dimensões: 450 mm de largura, 480 mm de profundidade e 600 mm de altura.

Na face frontal da estrutura foi criada uma porta de 200 mm de altura e 450 mm de largura, dedicada a oferecer acesso a parte interna da estrutura e abrigar o bocal por

onde as mãos são inseridas na estrutura. Na face superior desta foi criada uma abertura circular destinada ao posicionamento da lente da câmera apontando para seu interior. Nas faces laterais foram fixadas internamente duas corrediças telescópicas de 400 mm ligadas a face superior da estrutura, possibilitando o ajuste de altura da câmera entre 200 mm e 600 mm, para auxiliar no ajuste da altura da câmera foi instalada na parte externa da face superior da estrutura uma alça. Nas Figuras 9 e 10 podem ser observadas ilustrações do sistema de aquisição de imagens multiespectrais com altura da câmera ajustada em 600 mm e suas respectivas dimensões.

Figura 9 – Dimensões da estrutura

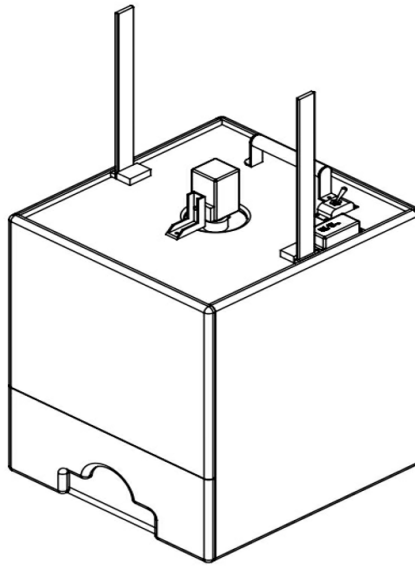


Fonte: Autoria Própria

De modo a mitigar a influência da iluminação externa na obtenção das imagens, que proporcionaria variação entre capturas em diferentes horários e condições climáticas, foi criada uma estrutura que proporcionasse isolamento da iluminação externa, utilizando assim uma iluminação padronizada em seu interior.

Para o sistema de iluminação foram utilizados dois tipos de lâmpada, a de filamento de 100W do tipo palito e fitas de LEDs de 5600 Kelvin. A lâmpada de filamento é utilizada para a iluminação durante a captura das imagens no espectro NIR, as fitas de LEDs são utilizadas para a iluminação durante a captura das imagens no espectro visível. Para propiciar a possibilidade da troca de iluminação foi criado um circuito elétrico com uma chave do tipo alavanca bipolar, de modo que cada lado da alavanca representasse a utilização de um tipo de iluminação. Foram utilizadas duas lâmpadas de filamento, estas foram fixadas em paralelo às faces laterais da parte interna da estrutura, de modo a atenuar o efeito de sombra. Mesmo motivo a determinar a distribuição das fitas de LEDs, no caso, foram utilizados quatro fitas com 12 LEDs cada, que por sua vez, foram fixados

Figura 10 – Ilustração da visão geral da estrutura



Fonte: Autoria Própria

em paralelo à cada uma das faces laterais da parte interna da estrutura.

Figura 11 – Lâmpada de filamento palito



Fonte: <https://images.app.goo.gl/PbuRpVF2Hz5Yz8gd7>

### 3.2 Base de Dados

A base de dados proposta neste trabalho possui o total de 8000 imagens, divididas igualmente entre 80 indivíduos, conferindo a cada indivíduo 100 imagens. Cada indivíduo possui imagens capturadas de ambas as mãos e um total de dez posições de captura diferentes para cada mão, sendo estas no espectro visível e também no NIR, totalizando 20 tipos de imagens por indivíduo. Para cada tipo de imagem de mão existe cinco exemplares,

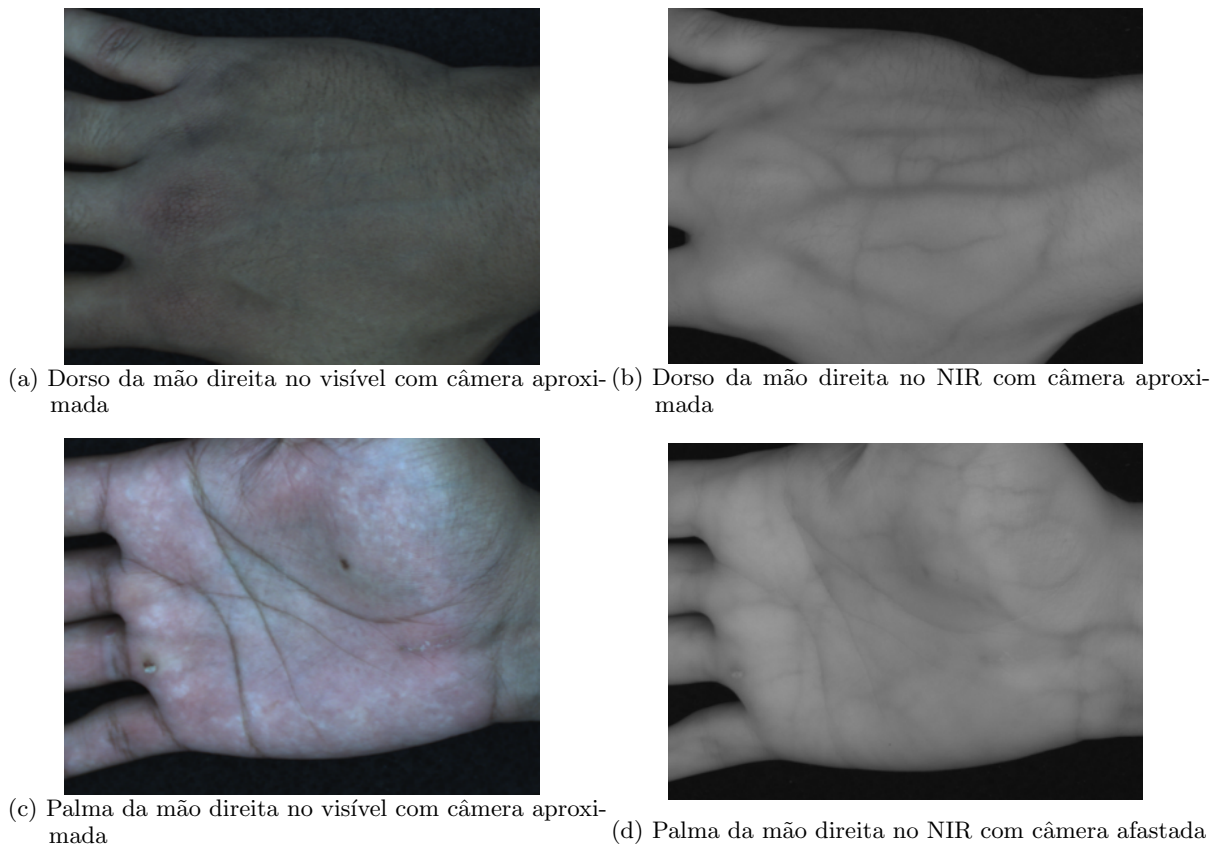
Figura 12 – Fita de LEDs



Fonte: <https://images.app.goo.gl/FPAv11H9qCVWNZzWA>

capturados separadamente com leve variação na maneira com que a mão está disposta na superfície a qual está apoiada. A captura foi realizada utilizando duas alturas para a câmera, uma aproximada e uma mais distante. Utilizando a câmera aproximada foram capturadas imagens do dorso, da palma e do pulso, de ambas as mãos do indivíduo nos espectros visível e NIR, conforme pode ser visto na Figura 13. Utilizando a câmera afastada foram capturadas imagens do dorso e da palma de ambas as mãos no espectro visível e NIR, conforme pode ser visto na Figura 14.

Figura 13 – Imagens capturadas com a câmera aproximada da mão no espectro visível e NIR



Fonte: Autoria Própria

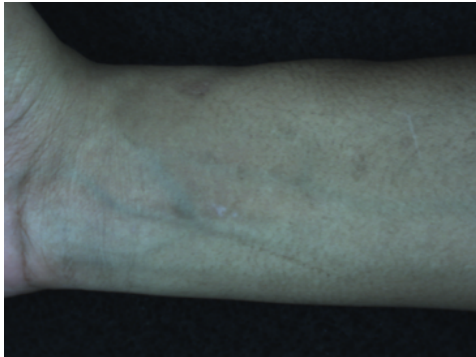
Figura 14 – Imagens capturadas com a câmera afastada da mão no espectro visível e NIR



(a) Dorso da mão direita no visível com câmera afastada (b) Dorso da mão direita no NIR com câmera afastada



(c) Palma da mão direita no visível com câmera afastada (d) Palma da mão direita no NIR com câmera afastada



(e) Pulso da mão direita no visível

(f) Pulso da mão direita no NIR

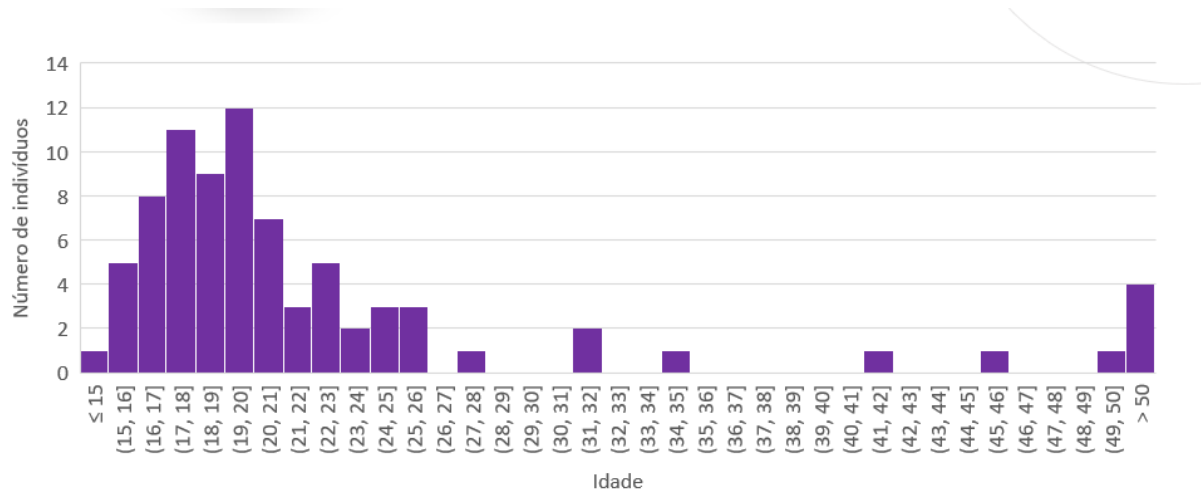
Fonte: Autoria Própria

São possuídas acerca das imagens da base de dados um rótulo de identificação de cada indivíduo, mais especificamente o nome; e características secundárias como idade e sexo. Estão contidas imagens de indivíduos entre 13 e 64 anos conforme Figura 15. A distribuição dos sexos pode ser observada conforme Figura 16. Em relação à tonalidade da pele dos indivíduos, existem tanto tonalidades claras quanto escuras.

### 3.3 Análise Exploratória

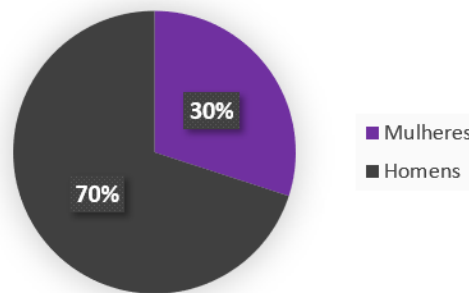
Para ter conhecimento sobre o comportamento dos dados e como estão dispostos, foi utilizado como ferramenta para análise exploratória a PCA utilizando o algoritmo

Figura 15 – Distribuição das Idades dos Indivíduos



Fonte: Autoria Própria

Figura 16 – Distribuição dos Sexos dos Indivíduos



Fonte: Autoria Própria

SVD para análise após os valores estarem centrados na média. Foram utilizadas as 400 componentes referentes às 400 imagens submetidas. Foram usadas como exemplo as imagens do dorso aproximado da mão esquerda dos indivíduos no espectro NIR. Após geradas as componentes principais, foram obtidas algumas informações úteis, como a porcentagem da variância explicada por cada componente principal conforme a Tabela 1.

Podemos observar por meio da Tabela 1 que com as imagens originais, utilizando as primeiras três PC's podemos explicar 75,9163% da variância total, bem como utilizando as primeiras dez PC's podemos explicar 89,2179% da variância total.

### 3.4 Pré-processamento das Imagens

Com o intuito de aprimorar o desempenho dos classificadores utilizados para o *benchmark*, foram realizados alguns pré-processamentos nas imagens, a fim de evidenciar as regiões de interesse e diminuir a complexidade a ser processada pelos modelos. Nas imagens pertencentes ao espectro visível foi realizado a evidenciação de características como veias e impressões palmares. Bem como nas imagens pertencentes ao espectro NIR

Tabela 1 – Variância explicada pelas componentes principais com imagens originais

Componente (PC)	Variância explicada (%)
1	65,8973
2	5,2936
3	4,7227
4	3,9957
5	2,6720
6	2,1232
7	1,6341
8	1,1540
9	0,8858
10	0,8395

o pré-processamento foi realizado visando a evidenciação especificamente das veias. As etapas que compõe o pré-processamento das imagens tanto no visível quanto no NIR são apresentadas nas seções seguintes e podem ser visualizadas na Figura 17.

#### 3.4.1 Extração de ROI

De modo a diminuir a dimensão das imagens e concentrar-se somente em regiões significativas para o trabalho, foi realizada a extração de uma região de interesse à partir das imagens. A região extraída foi pertinente a cada tipo de imagem, como no caso das imagens do dorso das mãos, foi extraída uma região retangular no centro do dorso omitindo os dedos e o pulso.

#### 3.4.2 Suavização

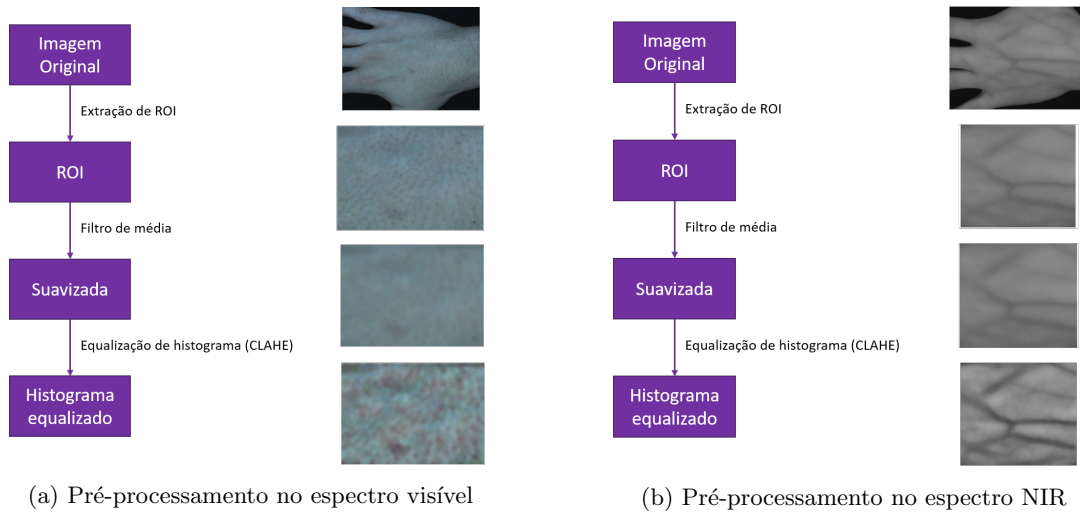
No âmbito de mitigar a influência de pelos, cicatrizes e marcas na pele, foi realizada uma suavização nas imagens por meio do filtro de média, especificamente de tamanho 16x16 nas imagens do espectro visível e 15x15x3 nas imagens do espectro NIR, em ambos os casos do tipo simétrico.

#### 3.4.3 Equalização de Histograma

A equalização de histograma foi utilizada com intuito de aprimorar o contraste das imagens distribuindo de uma forma mais igualitária as intensidades das componentes. Neste trabalho foi utilizada a Equalização de Histograma Limitada por Contraste (CLAHE) proposto por (Pizer et al., 1990). Nas imagens do espectro visível foram realizadas as equalizações individualmente em cada uma de suas componentes, nas imagens do espectro NIR foi realizada a equalização em sua componente única.



Figura 17 – Diagrama dos passos do pré-processamento das imagens



Fonte: Autoria Própria

### 3.4.4 Extração de Características

Foi utilizada uma técnica para extração de características das imagens chamada *Bag of Features*. Que consiste na aplicação do algoritmo SURF para extração de características e utilização do *K-means* para agrupamento dessas características. Foi utilizado para a localização dos pontos de características o método *grid*, com passo [8 8] e largura do bloco [32 64 96 128]. Foram mantidas 80% das características mais fortes de cada categoria. Na aplicação do *K-means* foram utilizados 500 agrupamentos.

### 3.5 Análise Exploratória Após Pré-processamento

Para ter conhecimento sobre a eficácia do pré-processamento nos dados, foi utilizado novamente como ferramenta para análise exploratória a PCA utilizando o algoritmo SVD para análise após os valores estarem centrados na média. Foram utilizadas as 400 componentes referentes às 400 imagens submetidas. Foram usadas como exemplo as imagens do dorso aproximado da mão esquerda dos indivíduos no espectro NIR, assim como na análise com imagens originais. Podemos observar a porcentagem da variância explicada por cada componente principal conforme a Tabela 2.

Podemos observar por meio da Tabela 2 que com as imagens pré-processadas, utilizando as primeiras três PC's podemos explicar 77,1541% da variância total, bem como utilizando as primeiras dez PC's podemos explicar 88,5048% da variância total.

#### 3.5.1 Comparativo

Observando a comparação entre a variância explicada pelas componentes principais antes e depois do pré-processamento como pode ser visto na Tabela 3, foi percebida uma representatividade maior nas três primeiras PC's após o pré-processamento, pode-se

Tabela 2 – Variância explicada pelas componentes principais com imagens pré-processadas

Componente (PC)	Variância explicada (%)
1	63,9222
2	8,3640
3	4,8680
4	3,0692
5	2,2263
6	1,5166
7	1,4471
8	1,1157
9	1,0254
10	0,9504

concluir que a classificação a partir das imagens pré-processadas pode ser realizada com mais precisão utilizando modelos menos complexos.

Tabela 3 – Comparativo entre a variância explicada pelas componentes principais

Componentes (PC's)	Imagens Originais (%)	Imagens Pré-processadas (%)
3 Primeiras	75,9163	77,1541
10 Primeiras	89,2179	88,5048
20 Primeiras	93,8069	93,3755
30 Primeiras	95,9663	95,5933

### 3.6 Modelos Propostos

Para a criação do *benchmark* inicial da base de dados foram proposto três modelos de classificação, sendo eles: SVM, KNN e Árvore de Decisão. Estes foram escolhidos de maneira arbitrária de modo a prover resultados utilizando diferentes abordagens algorítmicas. Os dados foram separados em dois subconjuntos, um para treinamento, outro para teste, com separação igual a 60% e 40% respectivamente. As parametrizações utilizadas nos modelos podem ser observadas nas Tabelas 4 à 6.

Tabela 4 – Parametrizações SVM

Parametrização	Tipo de Kernel	Restrição da caixa	Escala do Kernel	Codificação
SVM-1	Linear	1,0	1,0	onevsone
SVM-2	Linear	0,018576	0,0023518	onevsall

Tabela 5 – Parametrizações KNN

Parametrização	Número de vizinhos	Tipo de distância
KNN-1	1	Euclidiana
KNN-2	1	Spearman

Tabela 6 – Parametrizações Árvore de Decisão

Parametrização	Tamanho mínimo da folha
TREE-1	1
TREE-2	2

### 3.7 Avaliação dos Modelos

A avaliação da acurácia dos modelos foi realizada de maneira geral, realizando o teste com todo o subconjunto de teste. À partir dos resultados obtidos nos testes, foi calculada a relação entre a quantidade de acertos e a quantidade de amostras submetidas ao teste.

Sendo:

$a = \text{quantidade de acertos no subconjunto de testes}$

$n = \text{quantidade de amostras do subconjunto de testes}$

A avaliação da acurácia dos modelos é representada pela Equação 1

$$y = \frac{a}{n} \quad (1)$$

## 4 RESULTADOS OBTIDOS

Nas Figuras 18 (a) e 18 (b) podem ser observadas as acurácias do classificador SVM utilizando todos os tipos de imagens do espectro NIR e do espectro visível respectivamente.

Figura 18 – Resultados da Classificação utilizando o SVM

			SVM-1	SVM-1	SVM-2				SVM-1	SVM-1	SVM-2
			Original	Pré- processada	Pré- processada				Original	Pré- processada	Pré- processada
Mão Direita	Câmera afastada	Dorso	22,50%	99,38%	100,00%	Câmera afastada	Dorso	1,25%	96,89%	97,52%	
		Palma	10,00%	96,88%	98,75%		Palma	1,25%	98,76%	98,14%	
	Câmera aproximada	Dorso	17,50%	97,50%	98,13%	Câmera aproximada	Dorso	3,73%	100,00%	100,00%	
		Palma	6,83%	100,00%	100,00%		Palma	3,77%	99,37%	100,00%	
		Pulso	12,50%	98,75%	98,75%		Pulso	2,50%	100%	100,00%	
	Mão Esquerda	Câmera afastada	Dorso	11,25%	99,38%	100,00%	Câmera afastada	Dorso	2,50%	97,52%	96,89%
Palma			7,50%	99,38%	99,38%	Palma		2,53%	98,73%	99,37%	
Câmera aproximada		Dorso	7,50%	99,38%	100,00%	Câmera aproximada	Dorso	3,75%	99,38%	100,00%	
		Palma	1,25%	99,38%	100,00%		Palma	2,47%	100,00%	100,00%	
		Pulso	12,42%	100,00%	100,00%		Pulso	1,25%	99,38%	100,00%	
<b>Média</b>			<b>10,92%</b>	<b>99,00%</b>	<b>99,50%</b>	<b>Média</b>			<b>2,50%</b>	<b>99,00%</b>	<b>99,19%</b>

(a) Resultados da classificação das imagens no NIR

(b) Resultados da classificação das imagens no Visível

Pode ser constatado ao observar a Figura 18 que tanto para a análise no NIR quanto no visível tiveram baixa acurácia média com a utilização das imagens originais, e que com o pré-processamento, houve uma melhora considerável. Com a utilização da parametrização SVM-2 e imagens pré-processadas foi possível aprimorar ainda mais a acurácia média do modelo.

Nas Figuras 19 (a) e 19 (b) podem ser observadas as acurácias do classificador KNN utilizando todos os tipos de imagens do espectro NIR e do espectro visível respectivamente.

Pode ser constatado ao observar a Figura 19 que tanto para a análise no NIR quanto no visível tiveram acurácia média considerável com a utilização das imagens originais, e que com o pré-processamento, houve uma certa melhora. Entretanto, com a utilização da parametrização KNN-2 e imagens pré-processadas não foi possível aprimorar a acurácia média do modelo, no espectro NIR a acurácia média se manteve igual e no espectro visível a acurácia foi levemente comprometida.

Nas Figuras 20 (a) e 20 (b) podem ser observadas as acurácias do classificador Árvore de Decisão utilizando todos os tipos de imagens do espectro NIR e do espectro

Figura 19 – Resultados da Classificação utilizando o KNN

			KNN-1	KNN-1	KNN-2				KNN-1	KNN-1	KNN-2	
			Original	Pré- processada	Pré- processada				Original	Pré- processada	Pré- processada	
Mão Direita	Câmara afastada	Dorso	93,13%	98,75%	99,38%	Mão Direita	Câmara afastada	Dorso	92,55%	97,52%	96,89%	
		Palma	94,38%	98,13%	98,75%			Câmara afastada	Palma	88,82%	98,76%	97,52%
	Câmara aproximada	Dorso	96,88%	100,00%	99,38%		Câmara aproximada		Dorso	98,14%	100,00%	100,00%
		Palma	95,65%	100,00%	100,00%			Câmara aproximada	Palma	97,48%	100,00%	100,00%
		Pulso	95,63%	99,38%	99,38%				Câmara aproximada	Pulso	96,25%	100,00%
	Mão Esquerda	Câmara afastada	Dorso	96,25%	99,38%		99,38%	Mão Esquerda		Câmara afastada	Dorso	95,03%
Palma			93,75%	99,38%	100,00%	Câmara afastada	Palma		92,41%		97,47%	97,47%
Câmara aproximada		Dorso	96,25%	100,00%	100,00%		Câmara aproximada		Dorso	96,88%	99,38%	99,38%
		Palma	98,13%	100,00%	98,75%	Câmara aproximada			Palma	98,77%	100,00%	100,00%
		Pulso	95,03%	100,00%	100,00%				Câmara aproximada	Pulso	95,63%	100,00%
<b>Média</b>			<b>95,51%</b>	<b>99,50%</b>	<b>99,50%</b>	<b>Média</b>				<b>95,20%</b>	<b>98,82%</b>	<b>98,63%</b>

(a) Resultados da classificação das imagens no NIR

(b) Resultados da classificação das imagens no Visível

visível respectivamente.

Figura 20 – Resultados da Classificação utilizando a Árvore de Decisão

			TREE-1	TREE-1	TREE-2				TREE-1	TREE-1	TREE-2	
			Original	Pré- processada	Pré- processada				Original	Pré- processada	Pré- processada	
Mão Direita	Câmara afastada	Dorso	37,50%	33,13%	33,13%	Mão Direita	Câmara afastada	Dorso	39,13%	38,51%	38,51%	
		Palma	41,25%	28,13%	26,88%			Câmara afastada	Palma	42,24%	30,43%	29,19%
	Câmara aproximada	Dorso	41,88%	28,13%	26,88%		Câmara aproximada		Dorso	40,37%	26,71%	22,36%
		Palma	31,06%	31,68%	31,68%			Câmara aproximada	Palma	32,08%	20,13%	28,30%
		Pulso	25,00%	26,88%	31,88%				Câmara aproximada	Pulso	28,13%	20,00%
	Mão Esquerda	Câmara afastada	Dorso	43,13%	31,88%		33,13%	Mão Esquerda		Câmara afastada	Dorso	38,51%
Palma			41,88%	35,00%	33,13%	Câmara afastada	Palma		43,67%		21,52%	29,11%
Câmara aproximada		Dorso	39,38%	27,50%	25,63%		Câmara aproximada		Dorso	45,63%	31,25%	31,88%
		Palma	38,13%	29,38%	35,63%	Câmara aproximada			Palma	30,25%	29,01%	22,22%
		Pulso	32,95%	25,47%	33,54%				Câmara aproximada	Pulso	38,13%	36,25%
<b>Média</b>			<b>37,22%</b>	<b>29,72%</b>	<b>31,15%</b>	<b>Média</b>				<b>37,81%</b>	<b>28,42%</b>	<b>28,26%</b>

(a) Resultados da classificação das imagens no NIR

(b) Resultados da classificação das imagens no Visível

Pode ser constatado ao observar na Figura 20 que tanto para a análise no NIR quanto no visível tiveram baixa acurácia média com a utilização das imagens originais, e que com o pré-processamento, houve uma piora. Com a utilização da parametrização TREE-2 no espectro NIR a acurácia média foi levemente melhorada em relação à utilização com as

imagens pré-processadas e parametrização TREE-1. Utilizando a parametrização TREE-2 no espectro visível não foi possível aprimorar a acurácia média do modelo, apresentando leve piora em relação à utilização com imagens pré-processadas e parametrização TREE-1.

## 5 CONCLUSÃO

Para a obtenção das imagens, o Sistema de Aquisição de imagens multiespectrais se mostrou promissor, puderam ser obtidas imagens com resolução de 1024x768 pixels, considerada alta em comparação às bases de dados de mesmo intuito. A base de dados possui imagens capturadas de 80 indivíduos, possuindo 100 imagens por indivíduo, composta por 20 tipos diferentes de imagens, sendo dez imagens no espectro visível e dez no espectro NIR. Apresenta uma grande quantidade de tipos de imagens, quando comparada às bases de dados similares e ao estado da arte.

Por meio de experimentações utilizando os modelos de classificação SVM, KNN e Árvore de Decisão, tendo como objeto a base de dados proposta no trabalho, foi possível avaliar métricas de acurácia dos modelos propostos. Foram constatados valores relevantes de acurácia utilizando os modelos SVM e KNN.

Para o modelo SVM, entre todos os tipos de imagens do espectro NIR, foi obtida uma acurácia média de 99,5%. Para o mesmo modelo, utilizando todos os tipos de imagens no espectro visível, foi obtida uma acurácia média de 99,19%.

Para o modelo KNN, entre todos os tipos de imagens do espectro NIR, foi obtida uma acurácia média de 99,5%. Para o mesmo modelo, utilizando todos os tipos de imagens no espectro visível, foi obtida uma acurácia média de 98,63%.

Para o modelo Árvore de Decisão, entre todos os tipos de imagens do espectro NIR, foi obtida uma acurácia média de 37,22%. Para o mesmo modelo, utilizando todos os tipos de imagens no espectro visível, foi obtida uma acurácia média de 37,81%. Utilizando o modelo Árvore de decisão não foram obtidos bons resultados, devido a simplicidade do método.

Por meio das experimentações também foi possível observar o aprimoramento da acurácia dos modelos propiciado pela realização de um pré-processamento demasiado simples. A base de dados proposta neste trabalho se mostrou relevante ao estado da arte, conforme pôde ser visto ao longo do trabalho.

## REFERÊNCIAS

- Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (surf). **Computer Vision and Image Understanding**, v. 110, n. 3, p. 346 – 359, 2008. ISSN 1077-3142. Similarity Matching in Computer Vision and Multimedia. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1077314207001555>>.
- Braga, A. d. P. **Redes neurais artificiais: teoria e aplicações**. [S.l.]: Livros Técnicos e Científicos, 2000.
- Chou, K.-C.; Shen, H.-B. Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic k-nearest neighbor classifiers. **Journal of proteome research**, ACS Publications, v. 5, n. 8, p. 1888–1897, 2006.
- Haykin, S.; Network, N. A comprehensive foundation. **Neural networks**, v. 2, n. 2004, p. 41, 2004.
- Jain, A. K.; Flynn, P.; Ross, A. **Handbook of Biometrics**. New York: Springer, 2007.
- Jain, A. K.; Ross, A.; Prabhakar, S. An introduction to biometric recognition. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 14, n. 1, p. 4–20, 2004.
- Kabaciński, R.; Kowalski, M. Vein pattern database and benchmark results. **Electronics Letters**, v. 47, n. 20, p. 1127–1128, 2011.
- MacQueen, J. B. Some methods for classification and analysis of multivariate observations. In: . [S.l.: s.n.], 1967.
- Nowak, E.; Jurie, F.; Triggs, B. Sampling strategies for bag-of-features image classification. In: LEONARDIS, A.; BISCHOF, H.; PINZ, A. (Ed.). **Computer Vision – ECCV 2006**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 490–503. ISBN 978-3-540-33839-0.
- Pizer, S. M.; Johnston, R. E.; Ericksen, J. P.; Yankaskas, B. C.; Muller, K. E. Contrast-limited adaptive histogram equalization: speed and effectiveness. In: **[1990] Proceedings of the First Conference on Visualization in Biomedical Computing**. [S.l.: s.n.], 1990. p. 337–345.
- Tome, P.; Marcel, S. Palm vein database and experimental framework for reproducible research. In: **2015 International Conference of the Biometrics Special Interest Group (BIOSIG)**. [S.l.: s.n.], 2015. p. 1–7.
- Zhang, L.; Li, L.; Yang, A.; Shen, Y.; Yang, M. Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach. **Pattern Recognit.**, v. 69, p. 199–212, 2017.