



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS SOBRAL
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO
MESTRADO ACADÊMICO EM ENGENHARIA DE COMPUTAÇÃO

FRANCISCO GERARDO MEDEIROS NETO

EXTRAÇÃO AUTOMÁTICA DE CARACTERÍSTICAS DE ASAS DE MOSCAS DA
ESPÉCIE *DROSOPHILA MELANOGASTER*

SOBRAL

2017

FRANCISCO GERARDO MEDEIROS NETO

EXTRAÇÃO AUTOMÁTICA DE CARACTERÍSTICAS DE ASAS DE MOSCAS DA
ESPÉCIE *DROSOPHILA MELANOGASTER*

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação do Campus Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Computação. Área de Concentração: Sistemas de Informação

Orientador: Prof. Dr. Iális Cavalcante de Paula Júnior

SOBRAL

2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

M439e Medeiros Neto, Francisco Gerardo.

Extração automática de características de asas de moscas da espécie *Drosophila melanogaster* /
Francisco Gerardo Medeiros Neto. – 2017.
58 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Campus de Sobral, Programa de Pós-Graduação
em Engenharia Elétrica e de Computação, Sobral, 2017.

Orientação: Prof. Dr. Iális Cavalcante de Paula Júnior.

1. Processamento Digital de Imagens. 2. Reconhecimento de Padrões. I. Título.

CDD 621.3

FRANCISCO GERARDO MEDEIROS NETO

EXTRAÇÃO AUTOMÁTICA DE CARACTERÍSTICAS DE ASAS DE MOSCAS DA
ESPÉCIE *DROSOPHILA MELANOGASTER*

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação do Campus Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Computação. Área de Concentração: Sistemas de Informação

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Iális Cavalcante de Paula
Júnior (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Ajalmar Rêgo da Rocha Neto
Instituto Federal de Educação, Ciência e Tecnologia
do Ceará (IFCE)

Prof. Dr. Márcio André Baima Amora
Universidade Federal do Ceará (UFC)

Prof. Dr. Marcelo Marques Simões de Souza
Universidade Federal do Ceará (UFC)

Dedico este trabalho aos meus pais

AGRADECIMENTOS

Ao Prof. Dr. Iális Cavalcante de Paula Júnior por me orientar em minha dissertação de mestrado.

Aos meus pais e à Virgínia por todo apoio e dedicação.

Aos colegas do mestrado, em especial à Rosângela pelo apoio e palavras de incentivo.

Aos professores e funcionários da Universidade Federal do Ceará que permitiram minha formação profissional.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento da pesquisa de mestrado via bolsa de estudos.

“A educação é a arma mais poderosa que você
pode usar para mudar o mundo”

(Nelson Mandela)

RESUMO

Técnicas de biometria são utilizadas nos estudos de diversas espécies de animais, como exemplo a *Drosophila Melanogaster*, popularmente conhecida como mosca da fruta. Essa espécie tornou-se um organismo modelo para o estudo dos impactos que outros insetos produzem ao meio ambiente. Ademais, essas moscas possuem proteínas e genes similares aos dos seres humanos. Uma dificuldade para especialistas no estudo desses indivíduos é a semelhança das asas entre machos e fêmeas, que podem ser afetadas por mutações ou variações no genótipo. Este trabalho propõe um método de discriminação de gênero e genótipo de moscas da espécie *Drosophila melanogaster* a partir de características extraídas de imagens das asas. Essa abordagem se baseia na dimensão fractal extraída da segmentação por filtro de Canny das componentes da Transformada *Wavelet* Estacionária. A metodologia é validada com a divisão dos dados em grupos com taxas de treinamento e teste variáveis e na utilização de seis classificadores de abordagens diferentes: Floresta Aleatória, Máquinas de Vetor de Suporte, Perceptron Multicamadas, Análise por Discriminante Linear, Análise por Discriminante Quadrático e *K* Vizinhos Mais Próximos. Em seguida, a classificação é repetida com redução dos dados, apenas fêmeas para genótipo ou apenas indivíduos sem mutação para gênero. Os resultados obtidos foram satisfatórios, superando trabalhos da literatura com metodologias que não utilizam interação humana. Para genótipo, as taxas de acerto foram mais baixas devido à semelhança física entre as asas.

Palavras-chave: Dimensão fractal. Transformada *Wavelet* Estacionária. Filtro de Canny. *Drosophila melanogaster*.

ABSTRACT

Biometrics techniques are used in studies of several animal species, such as the *Drosophila melanogaster*, popularly known as fruit fly. This species has become a model organism for the study of the impacts that other insects produce to the environment. In addition, these flies have proteins and genes similar to those of humans. A challenge for specialists in the study of these individuals is the similarity of the wings between males and females, which may be affected by mutations or variations in the genotype. This work proposes a method of discrimination for gender and genotype of flies of the species *Drosophila melanogaster* from features extracted from images of the wings. This approach is based on the fractal dimension extracted from the Canny filter segmentation of the components of the Stationary Wavelet Transform. The methodology is validated by dividing the data into groups with variable training and test rates and using six different classifiers: Random Forest, Support Vector Machines, Multi-layer Perceptron, Linear Discriminant Analysis, Quadratic Discriminant Analysis and K Nearest Neighbors. Then, the classification is repeated with data reduction, only females for genotype or only individuals without mutation for gender. The results were satisfactory, surpassing works of the literature with automatic methodologies. For genotype, the hit rates were lower due to the physical similarity between the wings.

Keywords: Fractal dimension. Stationary wavelet transform. Canny filter. *Drosophila melanogaster*.

LISTA DE ILUSTRAÇÕES

| | | |
|-----------|---|----|
| Figura 1 | – Exemplares de moscas machos (M) e fêmeas (F) selvagens e com variação de genótipo. Imagens do microscópio Olympus e ampliação de 40X | 20 |
| Figura 2 | – Asas da mesma mosca fêmea tiradas com microscópios Olympus (direita) e Leica (esquerda) com ampliação de 40X (cima) e 20X (baixo) | 21 |
| Figura 3 | – Vetores médios ampliados das mutações (vermelho) em relação ao tipo selvagem <i>SAM</i> (preto) e magnitude da diferença entre eles em PD (SONNENS-CHEIN <i>et al.</i> , 2015a) | 22 |
| Figura 4 | – Exemplo de processamento com filtro mediana | 22 |
| Figura 5 | – Comparação da segmentação das componentes de segundo nível da TWE com operador de Canny de uma imagem tratada com filtro da mediana e da gaussiana. O valor $d = 7$ significa EE do tipo disco com raio 7 | 23 |
| Figura 6 | – Representação de uma TWD de 1 nível com banco de 2 filtros | 24 |
| Figura 7 | – Decomposição de TWE em 3 níveis | 25 |
| Figura 8 | – Funções de escalamento e <i>wavelet</i> da transformada de Haar para um sinal Dirac (STARCK <i>et al.</i> , 2007) | 25 |
| Figura 9 | – Decomposição com TWE de Haar de 2 níveis | 26 |
| Figura 10 | – Operador de Sobel na aplicação do filtro de Canny | 27 |
| Figura 11 | – Segmentação das componentes de segundo nível da TWE com variação de σ do filtro de Canny. Valores destacados foram utilizados na metodologia proposta | 27 |
| Figura 12 | – Comparação de métodos de segmentação com componentes de segundo nível da TWE | 28 |
| Figura 13 | – Exemplo de fractal: triângulo de Sierpinski | 29 |
| Figura 14 | – Vetores de suporte, em preto, em uma classificação com SVM linear | 32 |
| Figura 15 | – Rede MLP com 1 camada oculta. o_i é a ativação do nó i e f é a função de ativação | 33 |
| Figura 16 | – Separação de amostras utilizando LDA (superfície de separação linear) e QDA (superfície de separação quadrática) | 34 |
| Figura 17 | – Exemplo de abordagem kNN com $k = 3$ vizinhos mais próximos destacados | 35 |
| Figura 18 | – Definição da metodologia proposta. | 37 |
| Figura 19 | – Fluxograma da análise dos resultados obtidos. | 38 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Acerto obtido nos testes da metodologia proposta com amostras de ambos os gêneros. | 40 |
| Tabela 2 – Acerto obtido nos testes da metodologia proposta com amostras do gênero feminino apenas. | 41 |
| Tabela 3 – Acerto obtido nos testes da metodologia proposta com todos genótipos. | 42 |
| Tabela 4 – Acerto obtido nos testes da metodologia proposta com amostras de SAM. | 42 |
| Tabela 5 – Comparação dos resultados obtidos para classificação de genótipo com trabalho na literatura. | 43 |
| Tabela 6 – Comparação dos resultados obtidos para classificação de gênero com trabalhos na literatura. | 44 |
| Tabela 7 – Acerto obtido na classificação de genótipo com ambos os gêneros e imagens do microscópio Olympus e ampliação 20X. | 52 |
| Tabela 8 – Acerto obtido na classificação de genótipo com apenas fêmeas e imagens do microscópio Olympus e ampliação 20X. | 53 |
| Tabela 9 – Acerto obtido na classificação de gênero com todos os genótipos] e imagens do microscópio Olympus e ampliação 20X. | 53 |
| Tabela 10 – Acerto obtido na classificação de gênero com amostras de SAM e imagens do microscópio Olympus e ampliação 20X. | 54 |
| Tabela 11 – Acerto obtido na classificação de genótipo com ambos os gêneros e imagens do microscópio Leica e ampliação 20X. | 54 |
| Tabela 12 – Acerto obtido na classificação de genótipo com apenas fêmeas e imagens do microscópio Leica e ampliação 20X. | 55 |
| Tabela 13 – Acerto obtido na classificação de gênero com todos os genótipos e imagens do microscópio Leica e ampliação 20X. | 55 |
| Tabela 14 – Acerto obtido na classificação de gênero com amostras de SAM e imagens do microscópio Leica e ampliação 20X. | 56 |
| Tabela 15 – Acerto obtido na classificação de genótipo com ambos os gêneros e imagens do microscópio Leica e ampliação 40X. | 56 |
| Tabela 16 – Acerto obtido na classificação de genótipo com apenas fêmeas e imagens do microscópio Leica e ampliação 40X. | 57 |

| | |
|---|----|
| Tabela 17 – Acerto obtido na classificação de gênero com todos os genótipos e imagens do microscópio Leica e ampliação 40X. | 57 |
| Tabela 18 – Acerto obtido na classificação de gênero com amostras de SAM e imagens do microscópio Leica e ampliação 40X. | 58 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|---------------|---|
| <i>BioCAT</i> | Ferramenta de Classificação e Anotação de Bioimagem, <i>Bioimage Classification and Annotation Tool</i> |
| <i>DNA</i> | Ácido Desoxirribonucleico, <i>Deoxyribonucleic Acid</i> |
| <i>Egfr</i> | Receptor do fator de crescimento epidérmico, <i>Epidermal growth factor receptor</i> |
| <i>GEFEML</i> | Extração de Característica Genética e Evolucionária - Aprendizado de Máquina, <i>Genetic and Evolutionary Feature Extraction – Machine Learning</i> |
| <i>LBP</i> | Padrão Binário Local, <i>Local Binary Pattern</i> |
| <i>LDA</i> | Análise por Discriminante Linear, <i>Linear Discrimination Analysis</i> |
| <i>LVQ</i> | Quantização de Vetor de Aprendizado, <i>Learning Vector Quantization</i> |
| <i>MLBP</i> | Padrão Binário Local Modificado, <i>Modified Local Binary Pattern</i> |
| <i>MLP</i> | <i>Perceptron</i> Multicamadas, <i>MultiLayer Perceptron</i> |
| <i>PD</i> | Distância de Procrustes, <i>Procrustes distance</i> |
| <i>QDA</i> | Análise por Discriminante Quadrático, <i>Quadratic Discrimination Analysis</i> |
| <i>RBF</i> | Função de Base Radial, <i>Radial Basis Function</i> |
| <i>RGB</i> | Vermelho, Verde, Azul, <i>Red, Green, Blue</i> |
| <i>SOM</i> | Mapas Auto-Organizáveis, <i>Self-Organizing Maps</i> |
| <i>SVM</i> | Máquinas de Vetor de Suporte, <i>Support-Vector Machines</i> |
| <i>kNN</i> | <i>K</i> Vizinhos Mais Próximos, <i>K Nearest Neighbors</i> |
| DF | Dimensão Fractal |
| EE | Elemento Estruturante |
| OOB | Fora Do Saco, <i>Out-Of-Bag</i> |
| RD | Retinopatia Diabética |
| RF | Floresta Aleatória, <i>Random Forest</i> |
| RM | Ressonância Magnética |
| RNA | Rede Neural Artificial |
| TW | Transformada <i>Wavelet</i> |
| TWC | Transformada <i>Wavelet</i> Contínua, <i>Continuous Wavelet Transform</i> |
| TWD | Transformada <i>Wavelet</i> Discreta, <i>Discrete Wavelet Transform</i> |
| TWE | Transformada <i>Wavelet</i> Estacionária |
| UV | Ultravioleta |

SUMÁRIO

| | | |
|--------------|--|-----------|
| 1 | INTRODUÇÃO | 15 |
| 1.1 | Trabalhos relacionados | 16 |
| 1.2 | Objetivos do Trabalho | 17 |
| 1.3 | Organização da Dissertação | 17 |
| 2 | FUNDAMENTAÇÃO TEÓRICA | 19 |
| 2.1 | Introdução | 19 |
| 2.2 | Reconhecimento de padrões em imagens | 19 |
| 2.3 | Base de imagens e variação de genótipo | 19 |
| 2.4 | Filtro da mediana | 21 |
| 2.5 | Transformada <i>Wavelet</i> | 21 |
| 2.6 | Filtro de Canny | 26 |
| 2.7 | Dimensão fractal | 28 |
| 2.8 | Classificadores de Padrões | 29 |
| 2.8.1 | <i>Floresta Aleatória</i> | 30 |
| 2.8.2 | <i>Máquinas de Vetor de Suporte</i> | 30 |
| 2.8.3 | <i>Perceptron Multicamadas</i> | 32 |
| 2.8.4 | <i>Análise por Discriminante Linear e Análise por Discriminante Quadrático</i> | 33 |
| 2.8.5 | <i>K Vizinhos Mais Próximos</i> | 34 |
| 2.9 | Considerações finais | 35 |
| 3 | METODOLOGIA | 36 |
| 3.1 | Visão geral da metodologia | 36 |
| 3.1.1 | <i>Pré-processamento</i> | 36 |
| 3.1.2 | <i>Extração das características</i> | 36 |
| 3.1.3 | <i>Classificação e análise dos resultados</i> | 36 |
| 3.1.4 | <i>Detalhes da metodologia</i> | 38 |
| 3.2 | Implementação | 39 |
| 3.3 | Considerações finais | 39 |
| 4 | RESULTADOS | 40 |
| 4.1 | Genótipo | 40 |
| 4.2 | Gênero | 41 |

| | | |
|------------|---|-----------|
| 4.3 | Comparação com literatura | 43 |
| 5 | CONCLUSÕES E TRABALHOS FUTUROS | 45 |
| 5.1 | Contribuições | 45 |
| 5.2 | Perspectivas de trabalhos futuros | 46 |
| 5.3 | Trabalhos aceitos | 46 |
| | REFERÊNCIAS | 47 |
| | APÊNDICES | 52 |
| | APÊNDICE A – Resultados adicionais | 52 |

1 INTRODUÇÃO

O número de espécimes e a quantidade de informações que podem ser extraídas limitam muitas tarefas em áreas biológicas, como estudos sobre seleção artificial em moscas (HOULE *et al.*, 2003; WEBER; DIGGINS, 1990). Traços comportamentais e físicos, como impressão digital, íris ou forma de andar, podem ser usados para identificar indivíduos, humanos ou não (PAYNE *et al.*, 2013; AHMAD *et al.*, 2014; SONNENSCHHEIN *et al.*, 2015a). Ferramentas biométricas podem extrair informações fenômicas de imagens biológicas, como indicam Sonnenschein *et al.* (2015a). Neste trabalho, os autores explicam que um dos objetivos da fenômica (*phenomics*) é estudar as causas e consequências da variação do fenótipo em seres vivos. Segundo eles, características extraídas computacionalmente de imagens ou selecionadas manualmente podem classificar essas informações em grupos, como gênero ou genótipo.

Além de humanos, indivíduos de outras espécies podem ser identificados com técnicas biométricas, como exemplo pode-se citar as moscas da espécie *Drosophila melanogaster*, a mosca da fruta (PAYNE *et al.*, 2013; AHMAD *et al.*, 2014). Payne *et al.* (2013) explicam que essas moscas se tornaram um tipo de organismo modelo genético por mais de 100 anos devido ao ciclo de vida curto, entre 10 e 14 dias, e distribuição quase global como comensal do homem. Eles declaram que essas características tornaram as moscas da fruta uma ferramenta biomédica importante para identificação de mutações. Especialistas, como biólogos e entomologistas, podem analisar os impactos na natureza causados pelo desenvolvimento de diferentes insetos a partir de estudos com essas moscas (AHMAD *et al.*, 2014). Embora sejam espécies diferentes, proteínas e genes encontradas em humanos dão forma a asas de espécimes de *Drosophila melanogaster* (AHMAD *et al.*, 2014).

Visualmente, as asas de machos e fêmeas dessas moscas são tão similares que até um especialista encontra dificuldade em classificá-las (PAYNE *et al.*, 2013). Para resolver esse problema, Sonnenschein *et al.* (2015a) citam técnicas como reconhecimento de padrões computacionais e morfometria geométrica. Reconhecimento de padrões atribui uma categoria a uma imagem a partir de um classificador treinado (ZHOU *et al.*, 2013). Métodos de reconhecimento facial adaptados têm alcançado resultados satisfatórios na discriminação de imagens de asas de mosca (AHMAD *et al.*, 2014; PAYNE *et al.*, 2013; SONNENSCHHEIN *et al.*, 2015a). A morfometria geométrica requer a marcação de pontos de referência na imagem, onde é necessário um conhecimento prévio da homologia e potencial informativo do sistema biológico (SONNENSCHHEIN *et al.*, 2015a). Esse método remove a variação em tamanho, localização e

orientação, tratando da informação geométrica resultante (MITTEROECKER; GUNZ, 2009; KLINGENBERG, 2011; ZELDITCH *et al.*, 2012; SONNENSCHHEIN *et al.*, 2015a). Uma forma comum dessa análise é fazer uma correspondência um-a-um das espécimes a partir da posição relativa de pontos de referência (KLINGENBERG, 2011).

1.1 Trabalhos relacionados

Esta Seção apresenta trabalhos relacionados a temas como softwares para processamento digital de imagens de asas de moscas e classificadores utilizados.

Estudos envolvendo softwares para classificação de imagens de asas de moscas são recentes. Payne *et al.* (2013) utilizaram a técnica Extração de Característica Genética e Evolucionária - Aprendizado de Máquina, *Genetic and Evolutionary Feature Extraction – Machine Learning (GEFEML)*, desenvolvida por Shelton *et al.* (2012), e alcançaram 73,16 % de acerto. Esse método é baseado em Padrão Binário Local, *Local Binary Pattern (LBP)* e extratores de características evolucionários e foi implementado para mitigar ataques de repetição (*replay*)¹ em sistemas biométricos. Ahmad *et al.* (2014) compararam os extratores *LBP* e Padrão Binário Local Modificado, *Modified Local Binary Pattern (MLBP)* alcançando 90% e 89,5% de sucesso com os classificadores Floresta Aleatória, *Random Forest (RF)* e Máquinas de Vetor de Suporte, *Support-Vector Machines (SVM)*, respectivamente. *LBP* extrai padrões texturais das imagens como atributos (SHELTON *et al.*, 2012) e *MLBP* é uma versão modificada que combina sinal e magnitude para um melhor desempenho na extração das características (AHMAD *et al.*, 2014).

Sonnenschein *et al.* (2015a) compararam os softwares Ferramenta de Classificação e Anotação de Bioimagem, *Bioimage Classification and Annotation Tool (BioCAT)* (ZHOU *et al.*, 2013) e WINGMACHINE (HOULE *et al.*, 2003). O primeiro obteve taxa de sucesso acima de 80% para gênero, porém taxas mais baixas para genótipo, no máximo 52%. O segundo atingiu resultados melhores: mais de 90% para gênero e mais de 80% para genótipo. Entretanto necessita da marcação de pontos de referência na imagem, o que requer um especialista da área (HOULE *et al.*, 2003). *BioCAT* combina extratores de características e classificadores, além de ferramentas de anotação e suporte a imagens em duas ou três dimensões (ZHOU *et al.*, 2013). FijiWings (DOBENS; DOBENS, 2013) e MorphoJ (KLINGENBERG, 2011) são outras aplicações para análise de asas, desenvolvidas em Java e de código fonte aberto.

¹ Repetição ocorre quando dados biométricos são capturados e usados para acesso ilegal (RATHA *et al.*, 2001; ROBERTS, 2007; SHELTON *et al.*, 2012).

FijiWings é baseado na plataforma Fiji (SCHINDELIN *et al.*, 2012) que nada mais é que o software ImageJ (SCHINDELIN *et al.*, 2015) com extensões adicionais, aumentando as opções de ferramentas para processamento digital de imagens. MorphoJ foi desenvolvido para análise usando morfometria geométrica, necessitando da marcação de pontos de referência (KLINGENBERG, 2011).

Após a extração das características, é feita a classificação. Alguns métodos comuns são RF e SVM (AHMAD *et al.*, 2014; SONNENSCHHEIN *et al.*, 2015a), *K* Vizinhos Mais Próximos, *K Nearest Neighbors (kNN)* (MURTY *et al.*, 2009; SONNENSCHHEIN *et al.*, 2015a), Análise por Discriminante Linear, *Linear Discrimination Analysis (LDA)* (BRUNO *et al.*, 2008; SONNENSCHHEIN *et al.*, 2015a) e redes neurais artificiais (SONNENSCHHEIN *et al.*, 2015a).

1.2 Objetivos do Trabalho

O objetivo geral deste trabalho é extrair características, de forma automática, em imagens segmentadas de asas de mosca para classificação de gênero e variação de genótipos de indivíduos da espécie *Drosophila melanogaster*, a partir da técnica de dimensão fractal.

Como objetivos específicos, tem-se:

- Analisar a combinação de técnicas eficientes de extração de características presentes na literatura;
- Validar a metodologia ao comparar as taxas de acerto média e máxima de métodos de classificação, variando grupos de treinamento e teste;
- Avaliar se a metodologia proposta alcança resultados satisfatórios sem a necessidade da interação humana.

1.3 Organização da Dissertação

Nesse primeiro capítulo é introduzida a motivação para o estudo, análise e classificação de asas de mosca *Drosophila melanogaster*, também são comentados trabalhos da literatura relacionados ao tema e os objetivos da dissertação. No Capítulo 2 são apresentados o problema da classificação de asas de mosca e os conceitos de transformada *wavelet* não decimada, filtro de Canny e dimensão fractal. No Capítulo 3 é detalhada a metodologia proposta: aquisição das imagens, extração das componentes da transformada *wavelet*, segmentação e cálculo da dimensão fractal, finalizando com as técnicas de classificação.

O Capítulo 4 trata da análise dos resultados obtidos para a classificação de gênero e genótipo, variando a população utilizada e os grupos de treinamento e teste, também é realizada uma comparação com trabalhos da literatura. As conclusões obtidas e propostas de trabalhos futuros são apresentadas no Capítulo 5. O Apêndice A mostra os resultados adicionais obtidos com outros grupos de imagens da base utilizada para validação da metodologia proposta.

2 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo será apresentado o problema de variação genotípica em asas de moscas. Em seguida, são mostrados breves resumos de reconhecimento de padrões em imagens e das definições de mediana, Transformada *Wavelet* Estacionária (TWE), filtro de Canny e dimensão fractal. O capítulo é encerrado com os conceitos dos métodos de classificação utilizados na metodologia proposta.

2.1 Introdução

Segundo Debat *et al.* (2009), variação fenotípica é o elemento principal da seleção natural. Eles dizem que fatores ecológicos combinados com desenvolvimento geram o fenótipo, que é o resultado tanto da evolução quanto ontogênese e cujas variações se refletem no desenvolvimento e evolução. Conforme esse trabalho, asas de mosca do gênero *Drosophila* são modelos acessíveis para estudo dessa variação. Para Houle *et al.* (2003) e Debat *et al.* (2009), o estudo de asas de *Drosophila* é importante porque essas estruturas têm várias funções, como voo e cortejo, além de fácil manipulação.

2.2 Reconhecimento de padrões em imagens

Reconhecimento de padrões em imagens extrai características ou padrões, como forma ou textura, de uma imagem para identificá-la como uma classe ou grupo previamente conhecido. A imagem é adquirida do meio físico com algum equipamento, como um microscópio. Em seguida a imagem é pré-processada. Etapas comuns são filtragem para remoção de ruído, decomposição e segmentação, para cálculo de atributos de contorno. As características, então, são extraídas das imagens binarizadas. Os padrões são usados para treinamento de classificadores e testados para a validação da abordagem proposta.

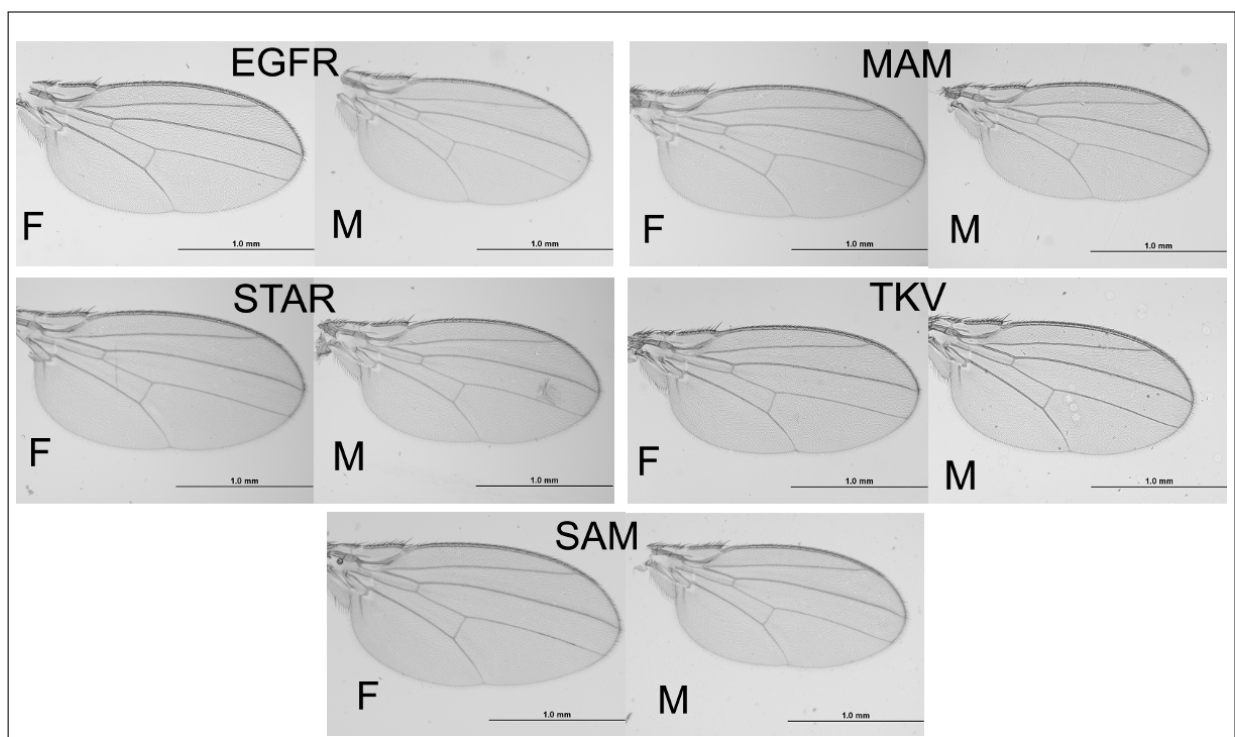
2.3 Base de imagens e variação de genótipo

As imagens utilizadas na metodologia proposta são das asas esquerda e direita, de amostras de machos e fêmeas, de espécimes selvagens de *Drosophila melanogaster* da região de Samarkand¹, analisadas por Sonnenschein *et al.* (2015a) e disponíveis em Sonnenschein *et al.*

¹ Os indivíduos sem mutação são indicados por *SAM* no decorrer do texto.

(2015b). . Há também a presença de indivíduos com as seguintes variações no genótipo: Receptor do fator de crescimento epidérmico, *Epidermal growth factor receptor (Egfr)*, *mastermind (mam)*, *Star (S)* e *thickveins (tkv)*. Segundo (SONNENSCHNEIN *et al.*, 2015a), todas são mutações heterozigóticas com perda de função e qualitativamente indistinguíveis, se comparados com selvagens. Eles afirmam, ainda, que quando homozigóticos, *Egfr*, *mam* e *S* são letais, enquanto *tkv* gera defeito qualitativo. As imagens dessa base foram adquiridas a partir dos microscópios Olympus e Leica, ambos com ampliação de 20X e 40X. A Figura 1 compara as imagens de variações de genótipo com indivíduo selvagem. Há 2270 imagens nos grupos do microscópio Leica e 2269 nos grupos do Olympus e entre 200 e 260 imagens por classe (variação de genótipo) (SONNENSCHNEIN *et al.*, 2015a).

Figura 1 – Exemplares de moscas machos (M) e fêmeas (F) selvagens e com variação de genótipo. Imagens do microscópio Olympus e ampliação de 40X

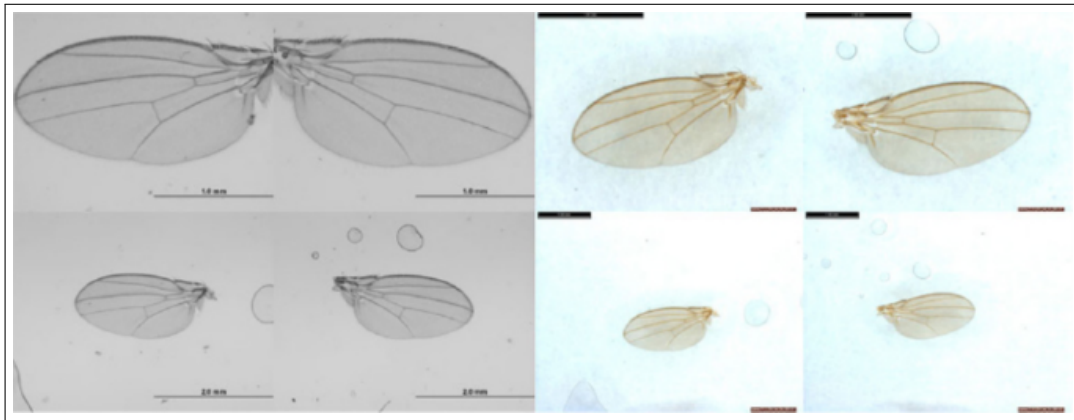


Fonte – O próprio autor com imagens de Sonnenschein *et al.* (2015b).

As Figuras 2 e 3 mostram as variações na aquisição das imagens e a diferença entre asas mutantes e não mutantes medida em Distância de Procrustes, *Procrustes distance (PD)*², respectivamente. A baixa variabilidade entre as imagens é o principal fator que levou a seleção dessa base para a metodologia proposta.

² Similar à distância Euclidiana entre vetores (SONNENSCHNEIN *et al.*, 2015a).

Figura 2 – Asas da mesma mosca fêmea tiradas com microscópios Olympus (direita) e Leica (esquerda) com ampliação de 40X (cima) e 20X (baixo)



Fonte – Sonnenschein *et al.* (2015a).

2.4 Filtro da mediana

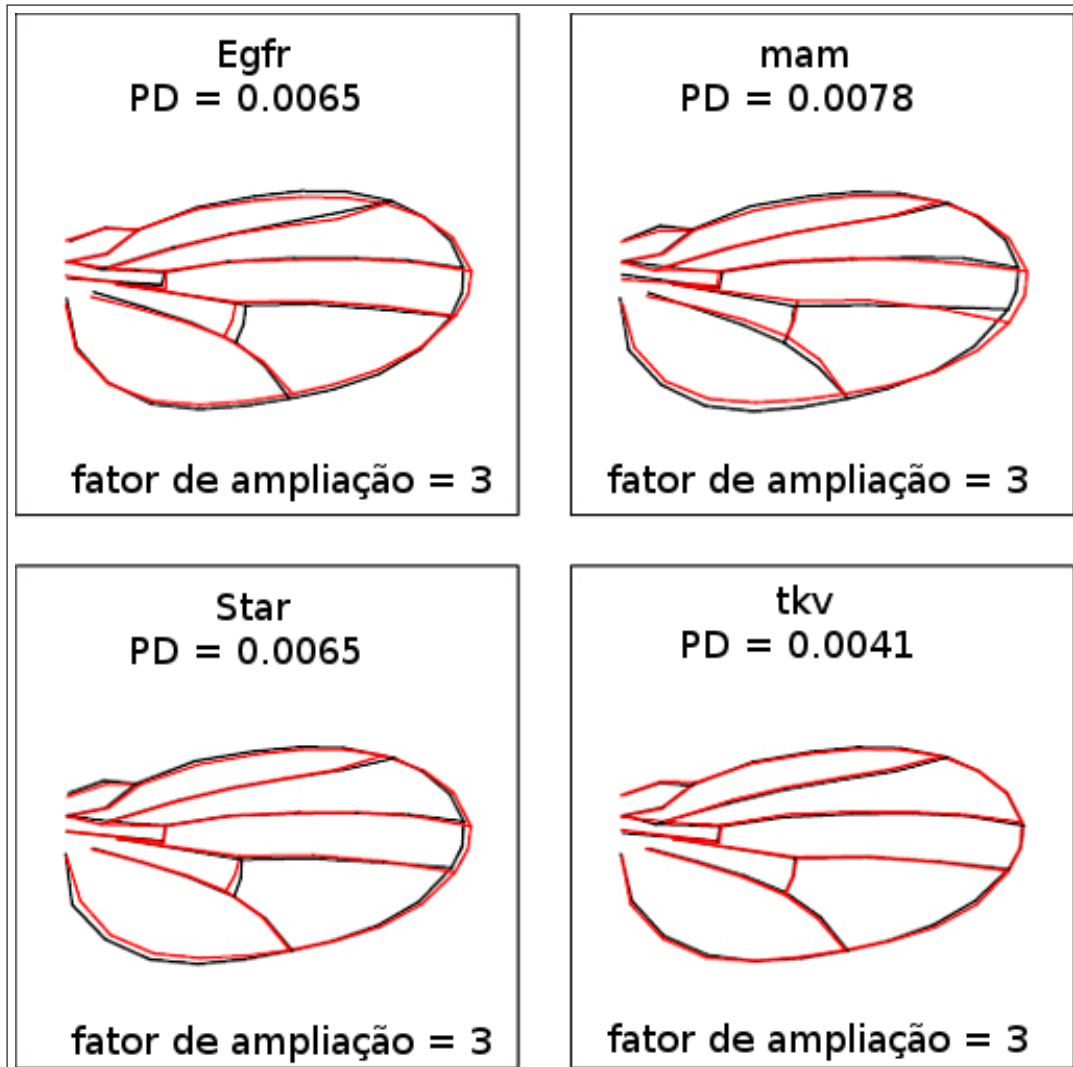
Segundo Costa e Jr. (2009), o filtro da mediana é utilizado para remover ruído, como o sal e pimenta³, baseado na vizinhança do pixel. Eles explicam que o pixel de referência e os vizinhos cobertos pelo operador são ordenados e o valor da mediana da sequência é selecionado. A Figura 4 mostra ua filtragem operador ou Elemento Estruturante (EE) do tipo disco com raio de 7 pixels. A Figura 5 compara a segmentação feita com o operador de Canny a partir de uma imagem tratada com mediana e com gaussiana. Com esse último filtro, poucos detalhes foram mantidos na imagem binária, não sendo útil para o cálculo da dimensão fractal.

2.5 Transformada *Wavelet*

Transformada *Wavelet* (TW) é uma ferramenta usada na análise de sinais de qualquer tipo como variação de dados financeiros, sinais elétricos e mesmo em imagens, considerando-as como sinais em duas dimensões. Figueiredo *et al.* (2015) detectaram lesões em retina causadas por Retinopatia Diabética (RD). Eles citam que RD afeta os pequenos vasos sanguíneos na retina, consequência da diabete, ameaçando a visão do indivíduo. Como forma de prevenção desse problema, diagnóstico precoce e tratamento são essenciais (FIGUEIREDO *et al.*, 2015). Khalid *et al.* (2016) segmentaram lesões na pele causadas por câncer de pele. Eles explicam que melanoma, tipo mais comum desse câncer, ocorre com exposição interminente de partes da pele ao sol, que danifica o Ácido Desoxirribonucleico, *Deoxyribonucleic Acid (DNA)* das células com

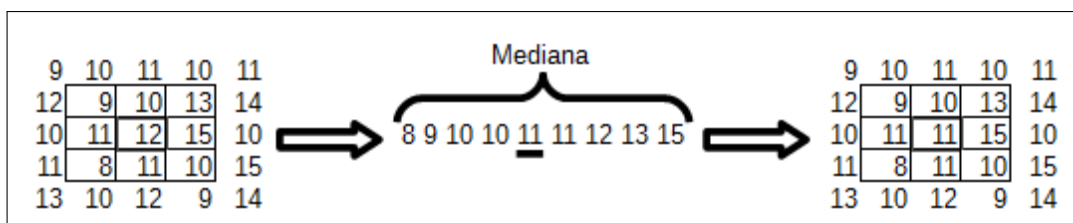
³ Segundo Costa e Jr. (2009), ruído sal e pimenta são “pontos isolados com diferentes níveis de cinza dispersos sobre regiões uniformes”.

Figura 3 – Vetores médios ampliados das mutações (vermelho) em relação ao tipo selvagem SAM (preto) e magnitude da diferença entre eles em PD (SONNENSCHNEIN *et al.*, 2015a)



Fonte – Adaptada de Sonnenschein *et al.* (2015a).

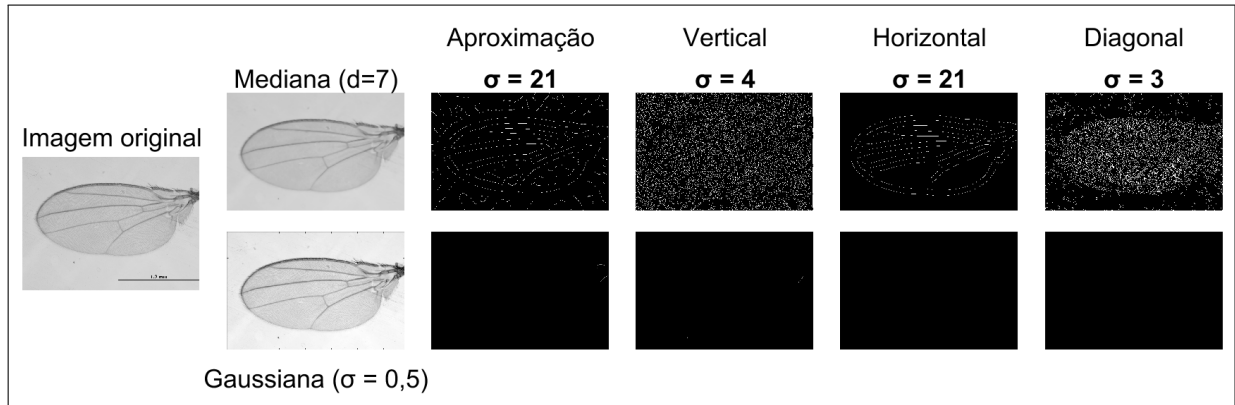
Figura 4 – Exemplo de processamento com filtro mediana



Fonte – O próprio autor com informações de Costa e Jr. (2009).

raios Ultravioleta (UV). Bankhead *et al.* (2012) segmentaram vasos de retina com método não supervisionado, alcançando acurácia de 93,71%. Demirhan e Güler (2011) identificaram tecidos

Figura 5 – Comparação da segmentação das componentes de segundo nível da TWE com operador de Canny de uma imagem tratada com filtro da mediana e da gaussiana. O valor $d = 7$ significa EE do tipo disco com raio 7



Fonte – O próprio autor com imagem de exemplo de Sonnenschein *et al.* (2015b).

orgânicos em imagens de Ressonância Magnética (RM) do cérebro, segmentando com uma combinação de rede Mapas Auto-Organizáveis, *Self-Organizing Maps (SOM)* e Quantização de Vetor de Aprendizado, *Learning Vector Quantization (LVQ)*. Nguyen *et al.* (2012) detectaram rachaduras em concreto associando TW a um filtro de Savitzky-Golay para identificação de borda. Algumas famílias de *wavelet* comuns são a B-spline cúbica (BANKHEAD *et al.*, 2012; FIGUEIREDO *et al.*, 2015; NGUYEN *et al.*, 2012), Daubechie (DEMIRHAN; GüLER, 2011) e Morlet (U. *et al.*, 2005).

Ebadi e Shafri (2015) explicam que TW é uma ferramenta para estudo de detalhes ocultos em certas escalas nos componentes da *wavelet* localizados em espaço e escala, por sua natureza de representação combinada de tempo-frequência. Eles falam ainda que essa transformada pode analisar séries contínuas ou discretas de escalas, recebendo o nome de Transformada *Wavelet* Contínua, *Continuous Wavelet Transform (TWC)* e Transformada *Wavelet* Discreta, *Discrete Wavelet Transform (TWD)*, respectivamente. Mallat (1989) desenvolveu uma estrutura rápida para decomposição e reconstrução da TWD que pode ser definida como (EBADI; SHAFRI, 2015):

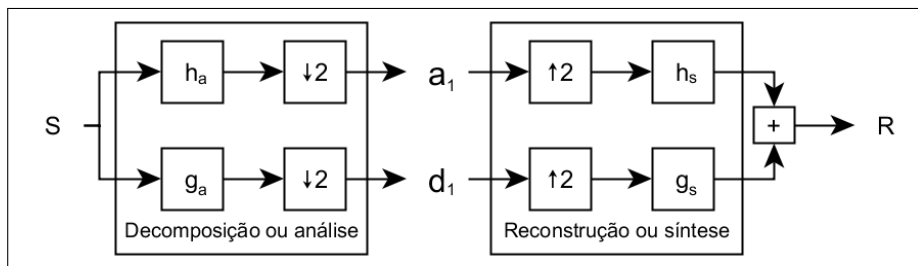
- A decomposição ou análise de um sinal S é feita convoluindo com os filtros h_a (passa baixa) e g_a (passa alta) gerando os coeficientes de aproximação a_1 e detalhes d_1 , respectivamente. A convolução é seguida de decimação (subamostragem) diádica ($\downarrow 2$), onde elementos de índice par são preservados. Com isso, o tamanho é reduzido pela metade do original;

- A síntese ou reconstrução é feita com superamostragem ($\uparrow 2$), inserindo zeros nos componentes a_1 e d_1 e passando filtros de reconstrução h_s (passa baixa) e g_s (passa alta), repetivamente. Ao final são somados para obter o sinal de saída R .

A Figura 6 mostra o processo de decomposição e reconstrução de uma TWE de um nível com dois bancos de filtros. O número de níveis de composição é calculado a partir do tamanho L do sinal, onde o maior valor J é dado por (EBADI; SHAFRI, 2015):

$$J = \log_2 L \quad (2.1)$$

Figura 6 – Representação de uma TWD de 1 nível com banco de 2 filtros



Fonte – O próprio autor com informações de Ebadi e Shafri (2015).

A TWD não é invariante a translação, o que afeta sinais durante o processo de remoção de ruído, segundo Ebadi e Shafri (2015). Eles explicam que pra resolver esse problema foi desenvolvida a TWE, que remove a subamostragem (decimação) e superamostragem da estrutura da TWD, por isso recebe também o nome de transformada *wavelet* não decimada. TWE foi desenvolvida por (UNSER, 1995) e é mostrada na Figura 7 para uma decomposição de três níveis. A análise em multiresolução consistente é mantida com superamostragem dos filtros dada por:

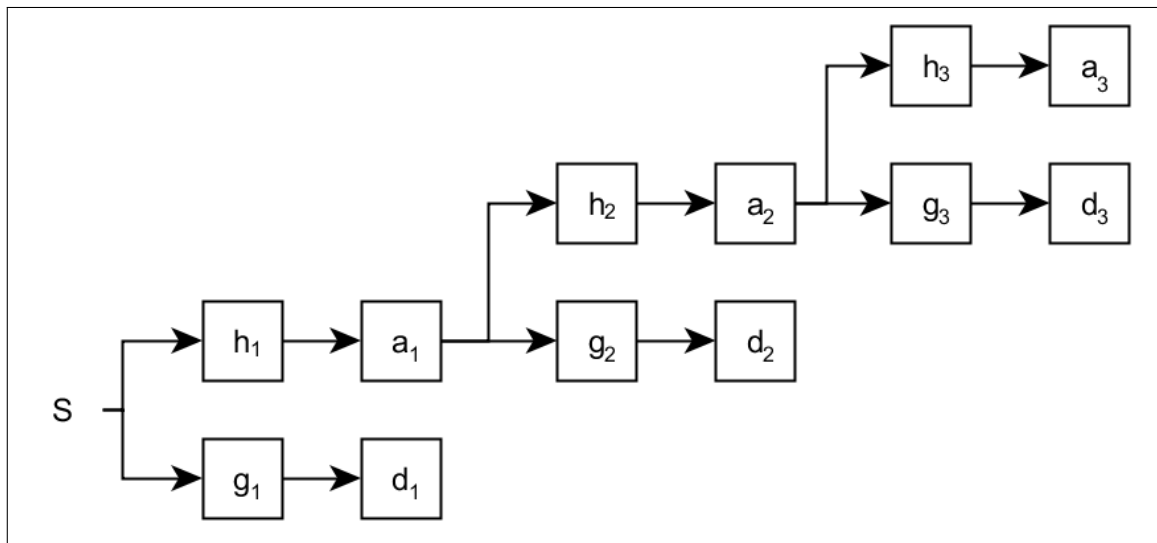
$$h_j = h_a \uparrow 2^{j-1} \quad (2.2)$$

$$g_j = g_a \uparrow 2^{j-1}, \quad j=1,2,\dots,N \quad (2.3)$$

onde j é o nível da transformada e N é o número de níveis da decomposição (UNSER, 1995; EBADI; SHAFRI, 2015; DEMIRHAN; GÜLER, 2011).

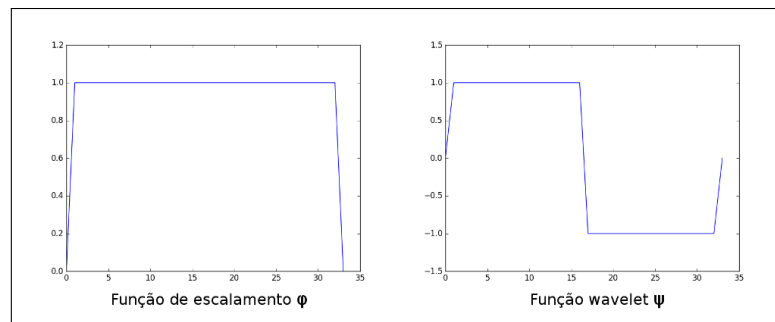
A Figura 8 mostra as funções ϕ (escalamento) e ψ (a função *wavelet*) da TW de Haar para um sinal de entrada do tipo Dirac.

Figura 7 – Decomposição de TWE em 3 níveis



Fonte – O próprio autor com informações de Ebadi e Shafri (2015).

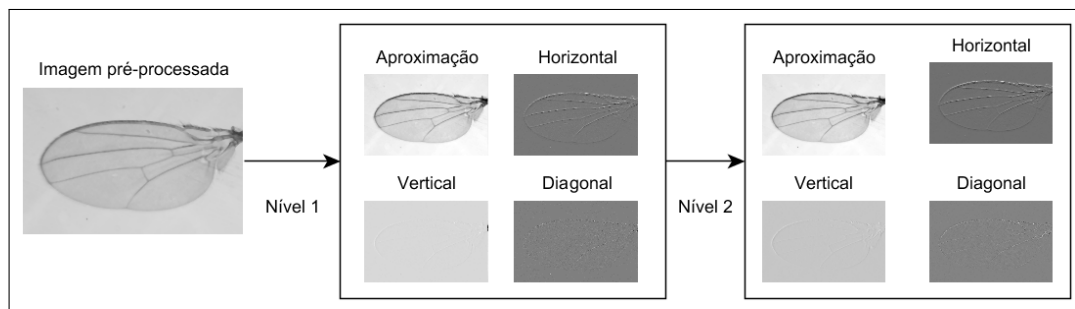
Figura 8 – Funções de escalamento e *wavelet* da transformada de Haar para um sinal Dirac (STARCK *et al.*, 2007)



Fonte – O próprio autor com dados de Grgic *et al.* (1999), Starck *et al.* (2007).

Essa Transformada é conceitualmente simples e rápida (MURTY *et al.*, 2009). A função da TW de Haar não é considerada útil em certas ocasiões por não ser suave, porém pode ser utilizada em remoção de ruído, como ruído de Poisson (GRGIC *et al.*, 1999; STARCK *et al.*, 2007). A Figura 9 mostra dois níveis de decomposição da TWE de Haar utilizada na metodologia proposta.

Figura 9 – Decomposição com TWE de Haar de 2 níveis



Fonte – O próprio autor com imagem de Sonnenschein *et al.* (2015b).

2.6 Filtro de Canny

Após o tratamento das imagens, uma etapa comum para a detecção de objetos é a segmentação. Uma técnica utilizada é o filtro de Canny. Xavierarockiaraj *et al.* (2012) utilizaram esse método na detecção de tumor cerebral em imagens de RM.

O filtro ou operador foi desenvolvido por Canny (1986) para detecção de bordas. Existem outros métodos como Sobel, Roberts e Prewitt mas em imagens com pouco ruído o operador Canny obtém melhores resultados (ÖZTÜRK; AKDEMIR, 2015). O algoritmo de detecção de Canny segue os seguintes passos (CANNY, 1986; ÖZTÜRK; AKDEMIR, 2015):

1. Remoção do ruído da imagem com uma filtragem gaussiana;
2. Cálculo das derivadas em duas dimensões (G_x e G_y) da imagem com algum operador. Na implementação foi utilizado Sobel (SKIMAGE..., 2017), mostrado na Figura 10. O gradiente da intensidade é calculado com a equação $|G| = |G_x| + |G_y|$;
3. Determinação da direção das bordas, escaneando os pixels em diferentes ângulos (0° , 45° , 90° e 135°);
4. Supressão de pontos não-máximos. Pixels não marcados como borda são reduzidos a 0;
5. Aplicação de limiar para binarização da imagem. São considerados borda os valores acima do limiar ou dentro do limiar mas conectados à borda.

A variação do parâmetro σ do filtro gaussiano influencia no nível da detecção das bordas, como mostra a Figura 11. A Figura 12 compara filtros de Roberts, Prewitt, Sobel e Laplaciano com filtro de Canny, não alcançando resultados desejáveis (ÖZTÜRK; AKDEMIR,

Figura 10 – Operador de Sobel na aplicação do filtro de Canny

| | | |
|----|---|---|
| -1 | 0 | 1 |
| -2 | 0 | 2 |
| -1 | 0 | 1 |

Gx

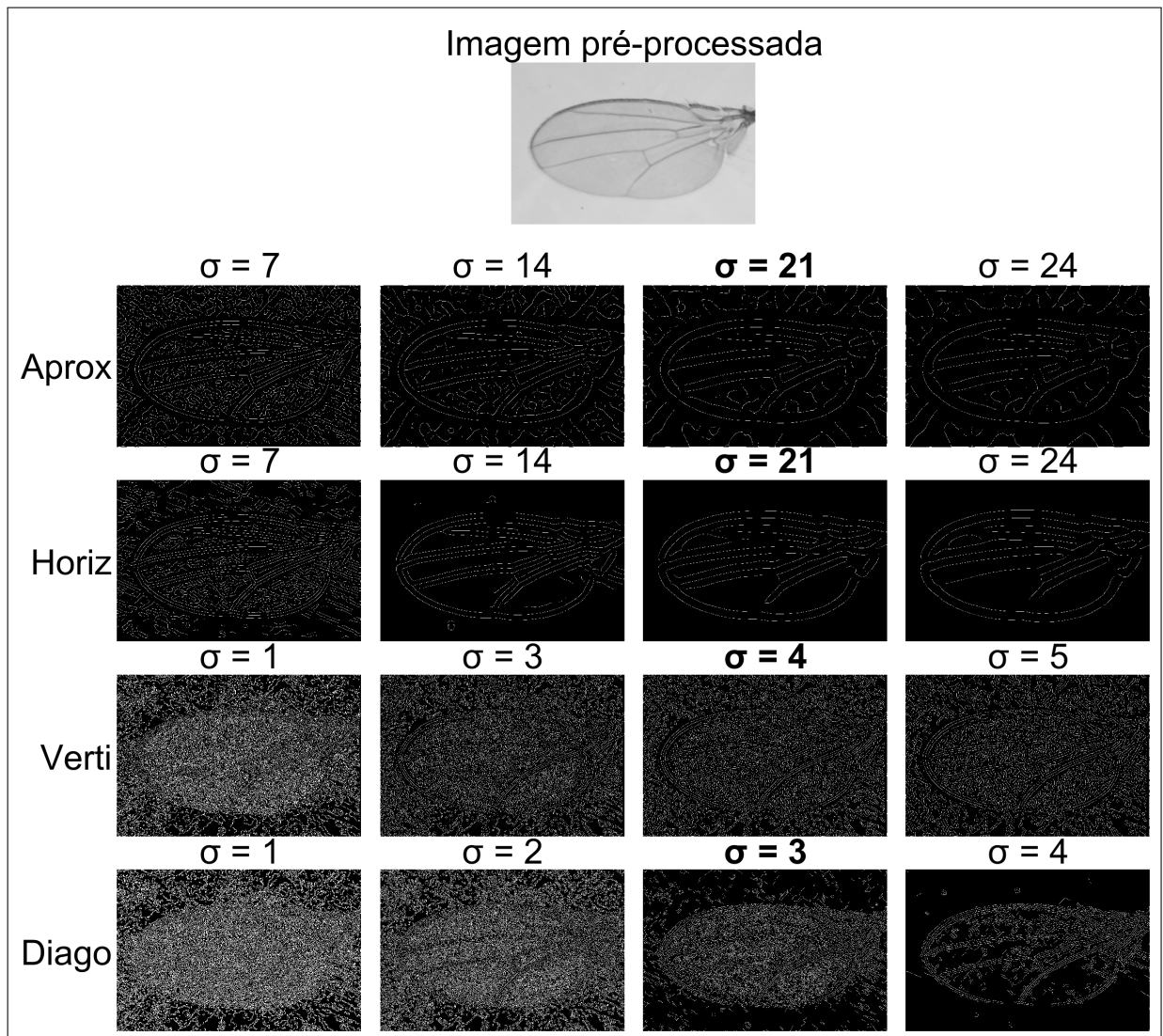
| | | |
|----|----|----|
| 1 | 2 | 1 |
| 0 | 0 | 0 |
| -1 | -2 | -1 |

Gy

Fonte – O próprio autor com informações de Öztürk e Akdemir (2015)

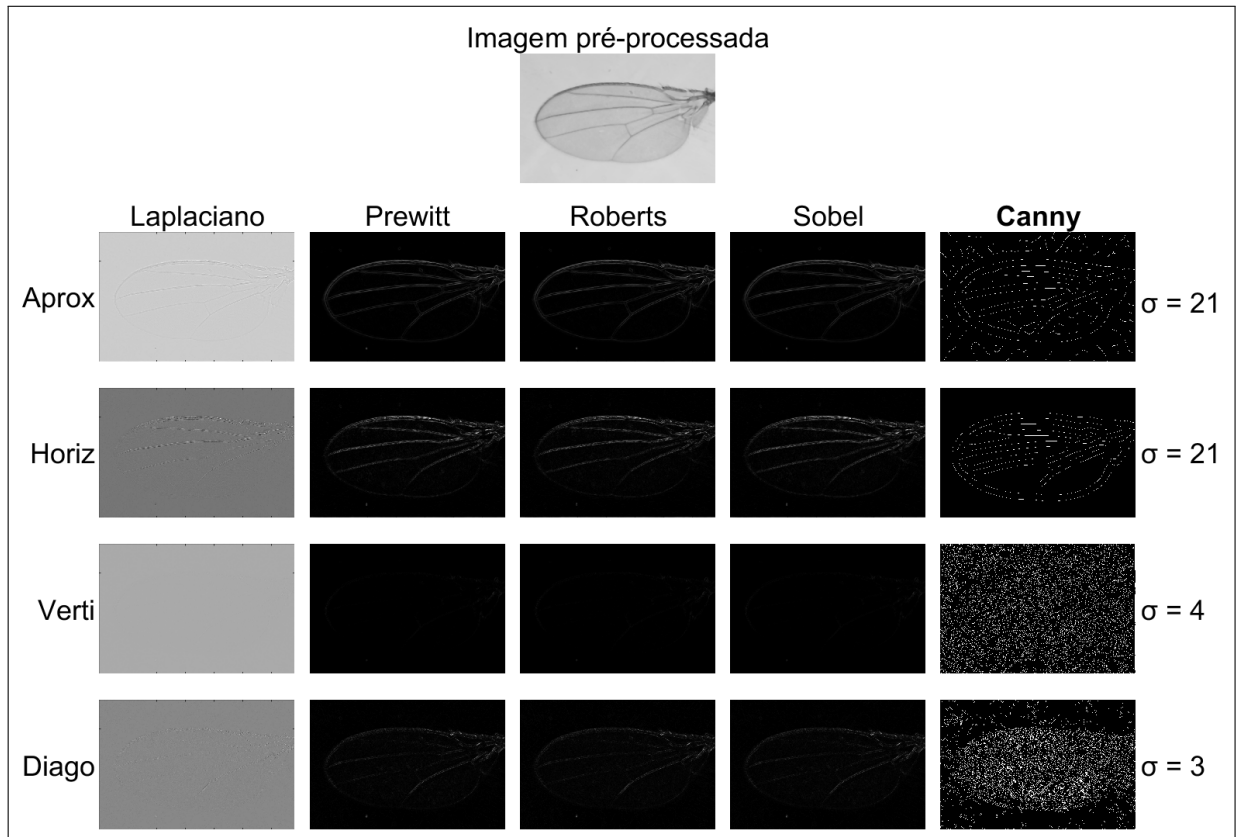
2015). Nessa Figura, pode-se perceber que as componentes de aproximação e horizontal mantêm a maior quantidade de detalhes e menos ruído que as componentes vertical e diagonal. O operador de Canny obteve melhor resultado na identificação do contorno da asa e das nervuras.

Figura 11 – Segmentação das componentes de segundo nível da TWE com variação de σ do filtro de Canny. Valores destacados foram utilizados na metodologia proposta



Fonte – O próprio autor com imagem de exemplo de Sonnenschein *et al.* (2015b).

Figura 12 – Comparação de métodos de segmentação com componentes de segundo nível da TWE



Fonte – O próprio autor com imagem de exemplo de Sonnenschein *et al.* (2015b).

2.7 Dimensão fractal

Após a segmentação, é feita a extração de características. Dimensão Fractal (DF) é um exemplo desses atributos. Bruno *et al.* (2008) comparam dois métodos de estimação de DF, *box-counting* (contagem de caixas) e Minkowski–Sausage multiescala, na classificação de folhas de plantas de biomas brasileiros a partir das nervuras. Polychronaki *et al.* (2010) estudaram ataque epilético a partir de sinais de eletroencefalograma, comparando três técnicas de cálculo de DF: Katz, Higuichi e *kNN*.

Fractal é um conjunto de pontos que lembra o todo ao diminuir a escala (U. *et al.*, 2005), como mostra a Figura 13. Na geometria euclidiana tradicional, a dimensão é um número natural que representa a quantidade de direções que cada diferencial da forma ocupa no espaço (BRUNO *et al.*, 2008; U. *et al.*, 2005). Por exemplo, um ponto tem dimensão 0, uma linha tem 1, um plano tem 2 e um sólido tem 3 dimensões (BRUNO *et al.*, 2008). Na DF, esse valor é fracionário, intermediário entre 2 dimensões euclidianas (BRUNO *et al.*, 2008; U. *et al.*, 2005).

DF é discutida desde o século 19, por matemáticos como Hausdorff, Koch, Sierpinski e Singer (BRUNO *et al.*, 2008). A Figura 13 mostra um exemplo de fractal. Existem vários métodos de cálculo dessa dimensão. Uma técnica é conhecida como dimensão de Hausdorff e é definida como:

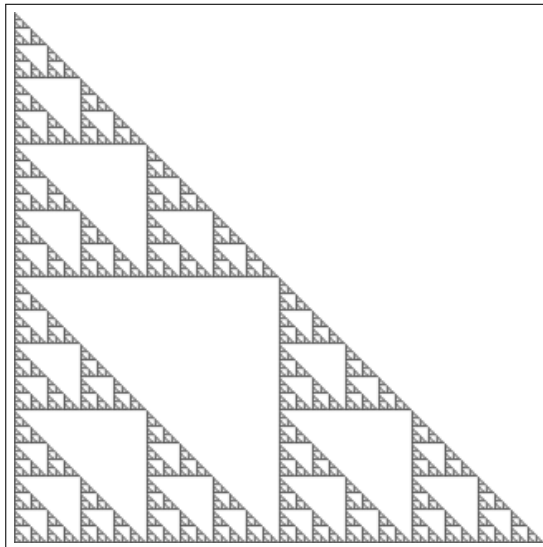
$$d_f = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}, \quad (2.4)$$

em que $N(\varepsilon)$ é o número de cubos de lado ε para preencher o espaço (BRUNO *et al.*, 2008). Uma aproximação dessa dimensão é feita com o método *box-counting*, calculado por (BRUNO *et al.*, 2008; VOSS, 1988):

$$d \sim -\frac{\log N(\varepsilon)}{\log \varepsilon}, \quad (2.5)$$

onde $N(\varepsilon)$ é o número de caixas de tamanho ε para preencher a imagem.

Figura 13 – Exemplo de fractal: triângulo de Sierpinski



Fonte – <<http://www.alceufc.com/computer/vision,/image/processing,/matlab,/fractal/2013/11/10/fractal-dimension-image.html>>

2.8 Classificadores de Padrões

Esta Seção conceitua os classificadores utilizados na metodologia proposta. Após a extração de características, é formado um vetor de padrões $\mathbf{x} = [x_1, x_2, \dots]$. O conjunto desses vetores é particionado em duas partes: uma para treinamento de um classificador e outra para o teste. Esse particionamento pode ser repetido, alterando os dados de cada grupo. Cada execução é chamada de realização.

2.8.1 Floresta Aleatória

RF foi desenvolvido por Breiman (2001) e utiliza a seleção dos melhores atributos em cada nó, diferente de árvores comuns que utilizam melhores divisões dos dados (LIAW; WIENER, 2002). Além de obter melhores resultados comparados com outros classificadores, utiliza apenas 2 parâmetros: o número de árvores e o número de variáveis para cada subconjunto aleatório em cada nó (BREIMAN, 2001; LIAW; WIENER, 2002).

O algoritmo de RF pode ser descrito como (LIAW; WIENER, 2002):

1. Selecionar $n_{\text{árvore}}$ amostras dos dados iniciais para *bootstrap*;
2. Criar uma árvore não-podada para cada amostra de *bootstrap*, selecionar m atributos aleatoriamente e escolher a melhor divisão de dados (*split*) baseado nesses atributos;
3. Predizer os novos dados a partir da maioria de votos das $n_{\text{árvore}}$.

Segundo Ho (1998), “em *bootstrapping*, subconjuntos de amostras de treinamento são independentemente e aleatoriamente selecionadas, com reposição, de acordo com uma distribuição de probabilidade uniforme”. A estimativa do erro, baseada nos dados de treinamento, pode ser calculada como (LIAW; WIENER, 2002):

1. Predizer os dados não selecionados a cada iteração de *bootstrap* utilizando a árvore criada nessa etapa. Esses dados não selecionados recebem o nome de Fora Do Saco, *Out-Of-Bag* (OOB) (BREIMAN, 2001);
2. Agregar os valores dessas previsões e calcular a taxa de erro.

2.8.2 Máquinas de Vetor de Suporte

SVM foi desenvolvido por Cortes e Vapnik (1995) e possui várias formulações. A metodologia proposta utiliza *SVM* de margem flexível, que resolve o seguinte problema primal:

$$\min_{\mathbf{w}, b, \xi} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \quad (2.6)$$

$$\text{sujeito a } y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \xi \geq 0, \quad i = 1, \dots, l, \quad (2.7)$$

no qual $C > 0$ é um parâmetro de regularização, $y_i \in \{-1, 1\}$ ($y \in R^1$), $\mathbf{x}_i \in R^n$ e $\phi(\mathbf{x}_i)$ mapeia \mathbf{x}_i em um espaço de altíssima dimensão, ξ manipula os dados não separáveis e o vetor de variáveis \mathbf{w}

é calculado como:

$$\mathbf{w} = \sum_{i=1}^l y_i \alpha_i \phi(\mathbf{x}_i), \quad (2.8)$$

onde α satisfaz o dual:

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \quad (2.9)$$

$$\text{sujeito a } y^T \alpha = 0, \quad 0 \leq \alpha_i \leq C, \quad (2.10)$$

no qual $e = [1, \dots, l]^T$ e Q é uma matriz $l \times l$ positiva semidefinida, em que $Q_{ij} \equiv y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ e $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ é a função *kernel* (AHMAD *et al.*, 2014; CHANG; LIN, 2011; CORTES; VAPNIK, 1995; BOSER *et al.*, 1992).

A função de decisão é dada por:

$$\text{sgn}(\mathbf{w}^T \phi(\mathbf{x}) + b) = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \right), \quad (2.11)$$

onde b é o viés (*bias*), que pode ser fixado *a priori* e mantido constante durante treinamento (CHANG; LIN, 2011; BOSER *et al.*, 1992).

A Figura 14 mostra um exemplo de classificação com *SVM* linear com duas classes, quadrados e círculos, onde os elementos destacados são os vetores de suporte. A linha fechada representa o hiperplano ótimo de separação e a distância entre as linhas tracejadas é a margem ótima (CORTES; VAPNIK, 1995).

Alguns exemplos de função *kernel* são :

- Linear

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j; \quad (2.12)$$

- Polinomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \quad \gamma \geq 0; \quad (2.13)$$

- Função de Base Radial, *Radial Basis Function (RBF)*

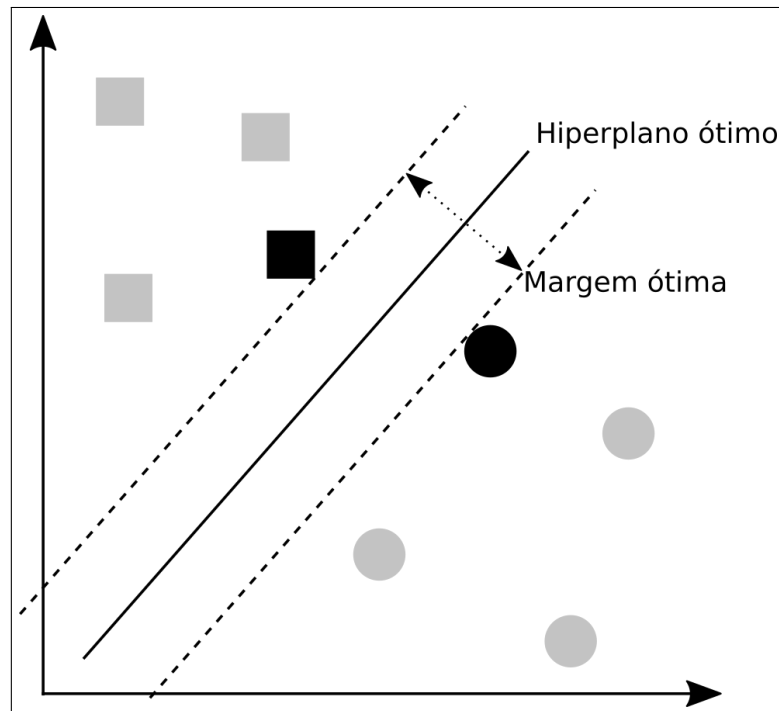
$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \gamma \geq 0; \quad (2.14)$$

- Sigmóide

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r), \quad (2.15)$$

onde γ , r e d são parâmetros das funções *kernel* (AHMAD *et al.*, 2014).

Figura 14 – Vetores de suporte, em preto, em uma classificação com SVM linear



Fonte – O próprio autor com informações de Cortes e Vapnik (1995).

2.8.3 Perceptron Multicamadas

Perceptron Multicamadas, MultiLayer Perceptron (MLP) é um tipo de Rede Neural Artificial (RNA) para aprendizado supervisionado (CRONE *et al.*, 2006). A arquitetura é mostrada na Figura 15 e é formada por nós u_j ligados a pesos w_{ij} que conectam a camada anterior à próxima (CRONE *et al.*, 2006; REED; MARKS, 1998). O funcionamento da rede pode ser resumido como (CRONE *et al.*, 2006; YAN *et al.*, 2006):

1. Os dados são lidos na camada de entrada;
2. Cada nó j da camada oculta multiplica a saída da camada anterior o_i por um peso treinável w_{ij} e soma com um viés b de peso θ_j ;
3. O resultado do somatório passa por uma função de ativação $f_j(\sum_{i=1}^p w_{ij}x_i + \theta_j)$. Um exemplo é a função retificadora linear $f(x) = \max(x, 0)$ (DOCUMENTATION..., 2017);
4. A saída da função de ativação é enviada para a próxima camada oculta ou camada de saída, que calcula a classe da amostra.

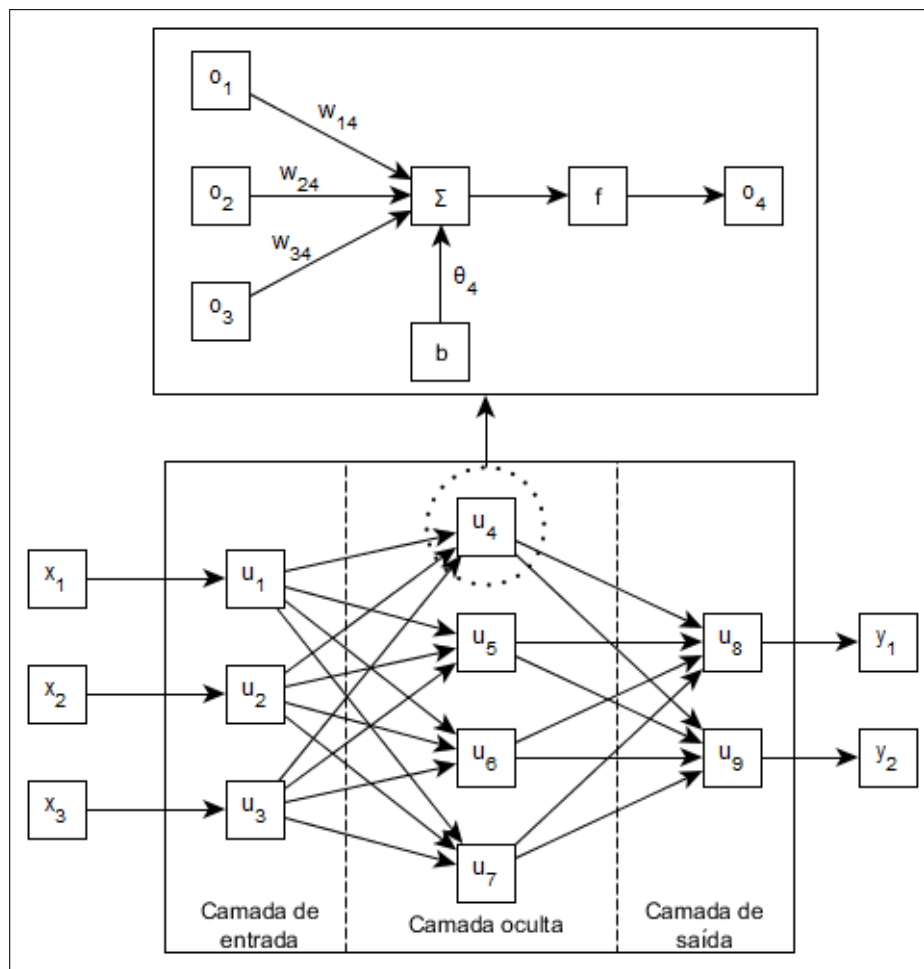
Segundo Crone *et al.* (2006) e Reed e Marks (1998), cada nó na camada oculta do *MLP* cria um hiperplano no espaço, dividindo o espaço em duas partes. Os nós na camada de saída combinam

essas regiões em conjuntos poliédricos, criando uma superfície de decisão, que pode ser binária ou não (CRONE *et al.*, 2006). Os pesos inicialmente são aleatórios e são adaptados a cada iteração por uma função objetivo $E(t)$ e uma regra de aprendizado $\eta(t)$, que pode ser calculada como:

$$\eta(t) = \eta_0 + \lambda E(t - 1), \quad (2.16)$$

onde t é a iteração, η_0 é um valor predefinido e $\lambda > 0$ (CRONE *et al.*, 2006; YAN *et al.*, 2006).

Figura 15 – Rede MLP com 1 camada oculta. o_i é a ativação do nó i e f é a função de ativação



Fonte – O próprio autor com informações de Crone *et al.* (2006) e Yan *et al.* (2006).

2.8.4 Análise por Discriminante Linear e Análise por Discriminante Quadrático

Segundo Kim *et al.* (2011), o vetor de variáveis \mathbf{X} é tido como multivariado normalmente distribuído com vetor média μ_i e, no caso do LDA, matriz de covariância Σ ou, no caso do Análise por Discriminante Quadrático, *Quadratic Discrimination Analysis (QDA)*, matriz de

covariância da classe Σ_i . A classificação é feita com uma função discriminante $g(\mathbf{X})$, dada por:

$$g_i(\mathbf{X}) = \mathbf{X}^T \Sigma^{-1} \mu_i - \frac{1}{2} \mu_i^T \Sigma^{-1} \mu_i + \log(\pi_i), \quad (2.17)$$

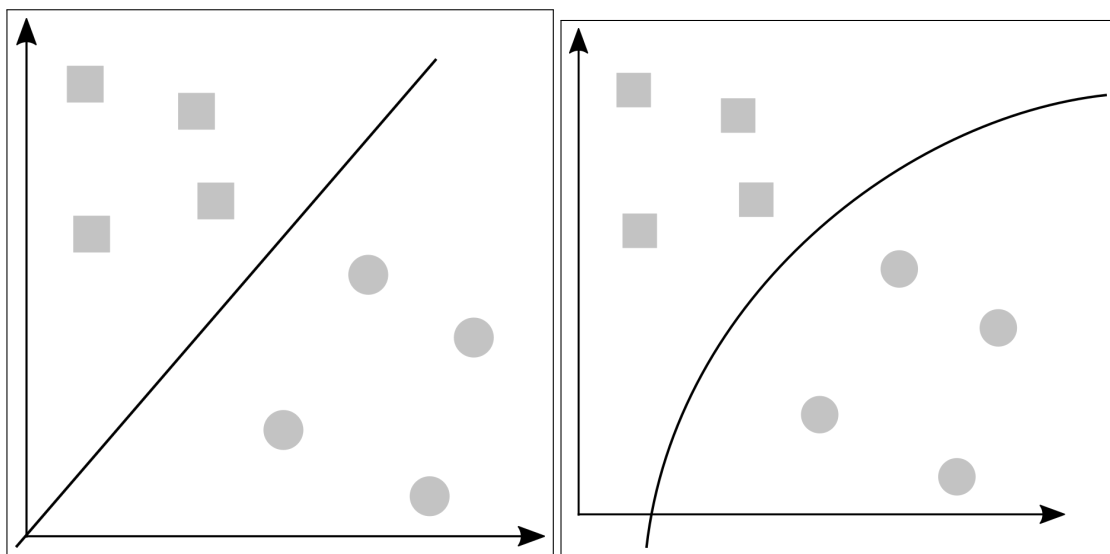
$$g_i(\mathbf{X}) = -\frac{1}{2} (\mathbf{X} - \mu_i)^T \Sigma_i^{-1} (\mathbf{X} - \mu_i) - \frac{1}{2} \log(|\Sigma_i|) + \log(\pi_i), \quad (2.18)$$

onde π_i é a probabilidade *a priori* da i -ésima classe e T indica transposição (KIM *et al.*, 2011; GROUVEN *et al.*, 1996). As Equações 2.17 e 2.18 se referem ao método *LDA* e *QDA*, respectivamente. A densidade condicional $f_i(\mathbf{X})$ da classe i de \mathbf{X} é dada por:

$$f_i(\mathbf{X}) = \frac{1}{(2\pi)^{P/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{X} - \mu_i)^T \Sigma_i^{-1} (\mathbf{X} - \mu_i) \right], \quad (2.19)$$

onde P é o número de atributos, vale 1 para *LDA* ou 2 para *QDA* (KIM *et al.*, 2011). A Figura 16 apresenta exemplo de classificação com os métodos *LDA* e *QDA*.

Figura 16 – Separação de amostras utilizando *LDA* (superfície de separação linear) e *QDA* (superfície de separação quadrática)

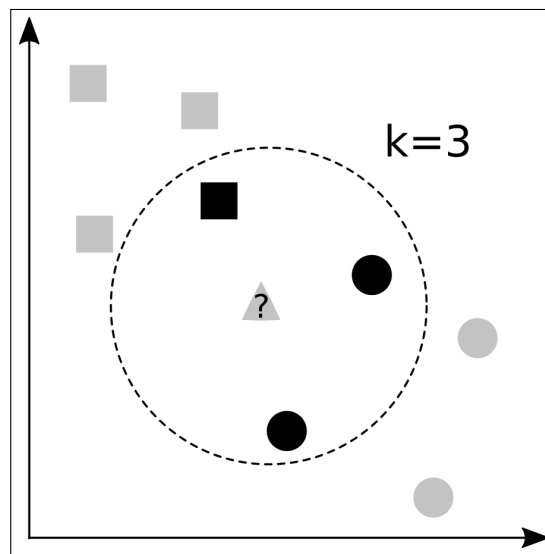


Fonte – O próprio autor.

2.8.5 *K Vizinhos Mais Próximos*

A técnica *kNN* utiliza o número de amostras mais próximas da amostra de teste para identificação da classe (COSTA; JR., 2009). A classe selecionada é baseada na classe com maior frequência entre as k amostras com menor distância do elemento testado, como mostra a Figura 17 (COSTA; JR., 2009). Nessa Figura, o elemento triângulo deverá ser testado. Usando $k = 3$, são selecionados 2 círculos e 1 quadrado (mais próximos no gráfico). A classe mais frequente para esse exemplo é círculo, logo o triângulo é classificado como um círculo.

Figura 17 – Exemplo de abordagem kNN com $k = 3$ vizinhos mais próximos destacados



Fonte – O próprio autor com informações de Costa e Jr. (2009).

2.9 Considerações finais

Este capítulo abordou os fundamentos necessários para o desenvolvimento da metodologia proposta: o problema da variação genotípica das asas e como esse afeta a forma das asas e foram apresentados os conceitos de TWE, filtro de Canny, DF e os classificadores.

O próximo Capítulo apresenta os detalhes da implementação da metodologia proposta, que compreende as etapas de pré-processamento, extração de características e classificação das espécies a partir de características das asas.

3 METODOLOGIA

Este Capítulo detalha a metodologia proposta para extração automática e classificação de asas de mosca. Essa abordagem se inicia com a aquisição das imagens, tratamento, extração das características, classificação e análise dos resultados. Em seguida são apresentados detalhes da implementação do algoritmo.

3.1 Visão geral da metodologia

A metodologia proposta é mostrada na Figura 18 e é descrita nas próximas Subseções.

3.1.1 Pré-processamento

Inicialmente é feita a leitura da imagem em nível de cinza. As imagens do microscópio Olympus já estão em nível de cinza, enquanto as do Leica estão em Vermelho, Verde, Azul, *Red*, *Green*, *Blue* (*RGB*) e precisam de conversão. Em seguida é removido o ruído com a aplicação de filtro de mediana com EE do tipo disco com raio de 7 pixels. A imagem é cortada para remoção da marcação de 1mm: a partir do pixel do topo à esquerda até uma altura de 900 pixels para as imagens do microscópio Olympus e do pixel número 51 até 950 para as do Leica.

3.1.2 Extração das características

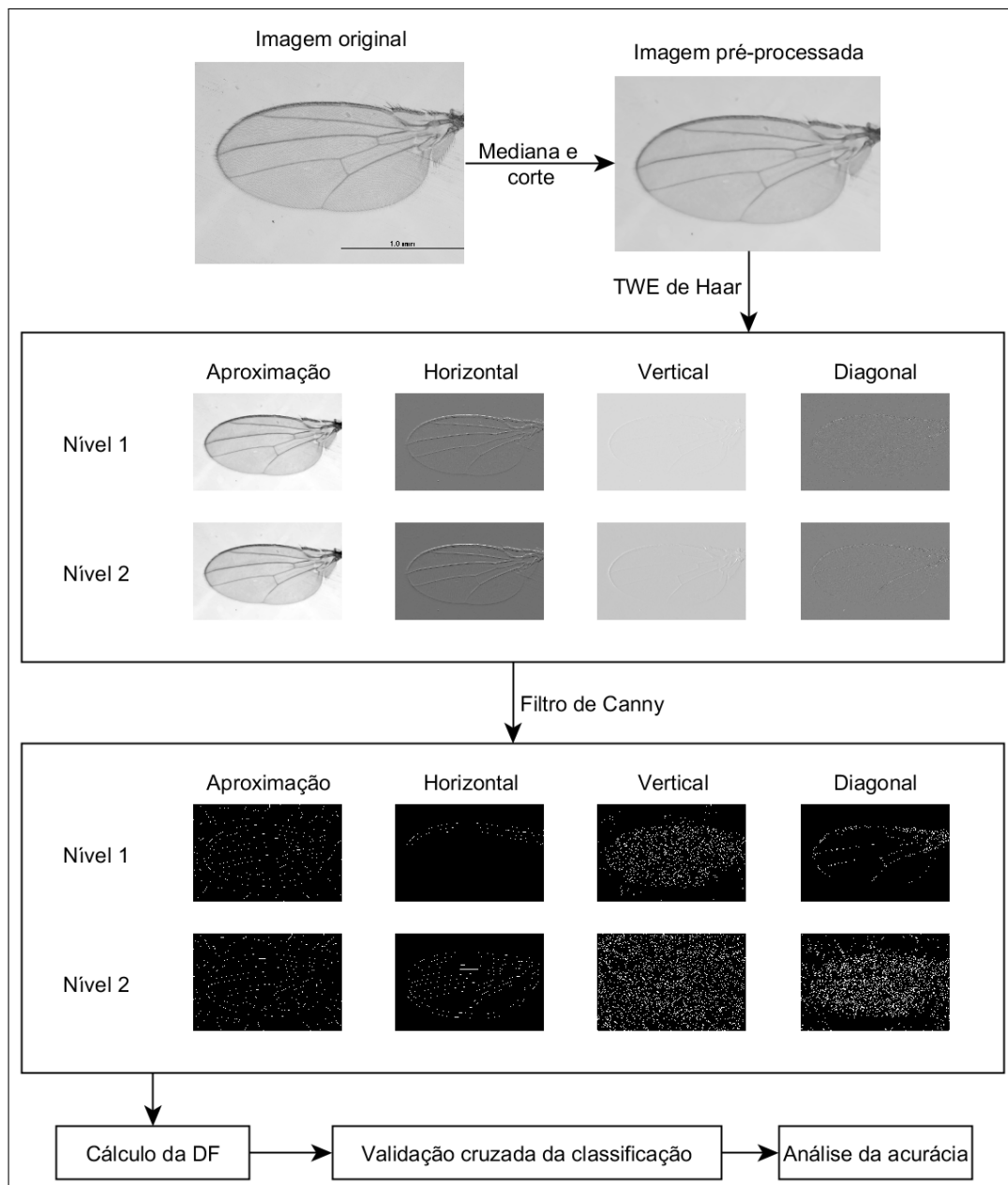
A imagem pré-processada é decomposta com TWE de Haar de 2 níveis. Cada nível possui 4 componentes, aproximação, horizontal, vertical e diagonal, totalizando 8 saídas. Para cada componente são detectadas as bordas a partir do filtro de Canny com $\sigma = 21$ para componentes aproximação e horizontal, $\sigma = 4$ para componente vertical e $\sigma = 3$ para componente diagonal. A dimensão fractal é extraída de cada imagem binarizada resultante.

3.1.3 Classificação e análise dos resultados

Os atributos são combinados e é feita uma seleção aleatória para formação de grupos de treinamento e teste. Para genótipo, os dados são divididos nos seguintes grupos:

- **A:** 360 amostras mutantes (sem número fixo para cada variação de genótipo) e 90 *SAM* com 20% de taxa de teste;
- **B:** 90 amostras de cada genótipo com 20% de taxa de teste;

Figura 18 – Definição da metodologia proposta.



Fonte – O próprio autor com imagem de Sonnenschein *et al.* (2015b).

- **C:** 150 amostras de cada genótipo com 50% de taxa de teste;
- **D:** 200 amostras de cada genótipo com 50% de taxa de teste.

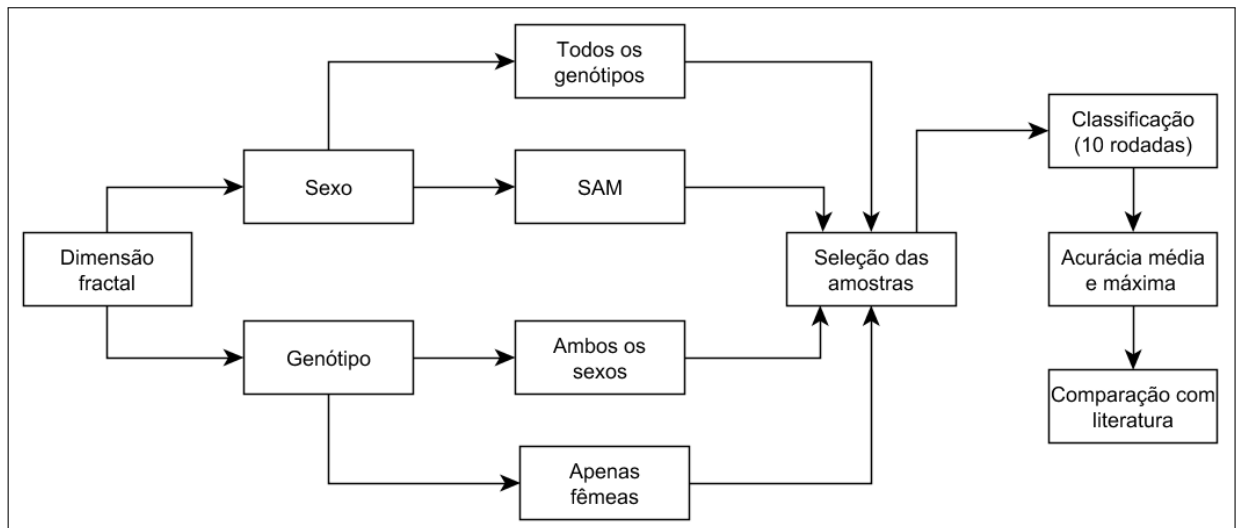
Após isso, são retiradas as amostras de machos e grupos são selecionados novamente. Para gênero, os dados são divididos nos seguintes grupos:

- **E:** 90 amostras de cada gênero (180 no total) por genótipo com 20% de taxa de teste;
- **F:** 100 amostras de cada gênero (200 no total) por genótipo com 50% de taxa de teste;

- **G**: 150 amostras de cada gênero (300 no total) por genótipo com 50% de taxa de teste;
- **H**: 200 amostras de cada gênero (400 no total) por genótipo com 50% de taxa de teste.

Após são selecionadas amostras de apenas *SAM* e repetido o processo. Cada grupo tem 10 realizações com cada método listado na Seção 2.8 e a taxa de acerto obtida é comparada entre os métodos da metodologia proposta e da literatura. Na Figura 19 é mostrado o fluxograma da validação da metodologia proposta.

Figura 19 – Fluxograma da análise dos resultados obtidos.



Fonte – O próprio autor.

3.1.4 Detalhes da metodologia

A marcação de 1mm foi removida por não ser necessária para a classificação e não interferir no cálculo da dimensão fractal. A imagem é cortada em 900 pixels devido ao algoritmo de TWE implementado em (PYWAVELETS..., 2017). Como parâmetros dos classificadores foram selecionados: RF com 15 árvores, SVM com *kernel* polinomial de grau 12, MLP com 20 neurônios na camada oculta e ativação com função retificadora linear e kNN com 5 vizinhos. Os valores do σ para o filtro de Canny e dos classificadores foram encontrados com o método de tentativas exaustivas. Os grupos de treinamento e teste variaram as taxas de amostragem para uma melhor análise da acurácia. Os trabalhos comparados no Capítulo 4 utilizaram abordagens semelhantes.

3.2 Implementação

O sistema foi desenvolvido em Python (ROSSUM, 1995) utilizando as bibliotecas scikit-image (WALT *et al.*, 2014) para processamento das imagens e segmentação com filtro de Canny, numpy (WALT *et al.*, 2011) para cálculos matemáticos e manipulação de dados, pywavelets (PYWAVELETS... , 2017) para computação das TWE, Cython (BEHNEL *et al.*, 2011) para cálculo da dimensão fractal, scikit-learn (PEDREGOSA *et al.*, 2011) para classificação dos dados, multiprocessing para execução paralela da extração das características e módulos padrões para atividades comuns, como listagem de arquivos. Algumas funções do scikit-image (SKIMAGE... , 2017) funcionam apenas com imagens em nível de cinza, por isso a conversão. A implementação da dimensão fractal foi adaptado para Python de código em Matlab disponível online¹ e convertido para Cython.

Cython compila o código diretamente para C e a paralelização aumentam o desempenho do algoritmo (BEHNEL *et al.*, 2011). Como Python é uma linguagem interpretada, a implementação de cálculos matemáticos mais complexos é mais lenta, enquanto C é compilada para linguagem de máquina, dando um aumento na velocidade da execução. A linguagem e as bibliotecas utilizadas são gratuitas e de código-fonte aberto.

3.3 Considerações finais

Neste capítulo foram apresentadas as etapas para extração de características e classificação da metodologia proposta. Os detalhes e restrições da implementação são descritos em seguida. Os resultados obtidos com combinação de vários métodos de classificação e variação nos grupos de treinamento e teste são mostrados e analisados no capítulo seguinte.

¹ <<http://www.alceufc.com/computer/vision,/image/processing,/matlab,/fractal/2013/11/10/fractal-dimension-image.html>>

4 RESULTADOS

Este capítulo apresenta os resultados obtidos com o método de extração de características proposto nesse trabalho, comparando técnicas de classificação e particionamento de dados.

Na próxima Seção são mostrados os resultados obtidos na discriminação para genótipo e, em seguida, para gênero. Finaliza-se o capítulo com a análise dos resultados em comparação com outras abordagens de classificação de asas de mosca disponíveis na literatura. As imagens utilizadas foram adquiridas com microscópio Olympus e com ampliação de 40X (SONNENSCHNEIN *et al.*, 2015b), as mesmas utilizadas em Sonnenschein *et al.* (2015a) e semelhantes às utilizadas em Payne *et al.* (2013) e Ahmad *et al.* (2014). Resultados adicionais obtidos com imagens com ampliação em 20X e obtidas com um microscópio Leica com ampliação 20X e 40X (SONNENSCHNEIN *et al.*, 2015b) são mostrados no Apêndice A.

4.1 Genótipo

A Tabela 1 mostra os resultados obtidos com a classificação de genótipo e ambos os gêneros e Tabela 2 utiliza a mesma divisão das amostras, porém com apenas espécies fêmeas.

Tabela 1 – Acerto obtido nos testes da metodologia proposta com amostras de ambos os gêneros.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 49,67 ± 4,74 | 57,78 | B | 52,44 ± 3,70 | 56,67 |
| <i>SVM</i> | | 49,56 ± 5,11 | 57,78 | | 50,44 ± 4,63 | 58,89 |
| <i>MLP</i> | | 42,44 ± 4,53 | 48,89 | | 46,78 ± 4,17 | 52,22 |
| <i>LDA</i> | | 47,00 ± 3,63 | 52,22 | | 51,22 ± 3,61 | 56,67 |
| <i>QDA</i> | | 52,33 ± 2,98 | 57,78 | | 55,67 ± 4,00 | 60,00 |
| <i>KNN</i> | | 48,89 ± 2,92 | 54,44 | | 51,78 ± 3,96 | 57,78 |
| <i>RF</i> | C | 55,36 ± 1,69 | 58,40 | D | 57,28 ± 1,93 | 60,40 |
| <i>SVM</i> | | 53,84 ± 1,81 | 56,00 | | 55,64 ± 2,24 | 60,20 |
| <i>MLP</i> | | 48,27 ± 1,82 | 50,93 | | 47,70 ± 2,42 | 51,60 |
| <i>LDA</i> | | 52,45 ± 2,09 | 55,73 | | 52,80 ± 1,64 | 54,80 |
| <i>QDA</i> | | 57,57 ± 2,18 | 60,80 | | 57,92 ± 1,47 | 60,20 |
| <i>KNN</i> | | 52,37 ± 2,33 | 57,07 | | 53,64 ± 1,99 | 58,00 |

Fonte – O próprio autor.

Com ambos os gêneros, conforme Tabela 1, o classificador *RF* se mostrou mais eficiente que o método *SVM*. O método *MLP* obteve pior resultado, seguido do *LDA* e *kNN*. *QDA* alcançou melhores resultados que *LDA* e que as outras técnicas. A acurácia máxima

(máx.) obtida entre os testes foi de 60,80% com *QDA* e grupo C. Taxas baixas de acerto eram previstas (SONNENSCHNEIN *et al.*, 2015a), pois a semelhança da estrutura física das asas entre as variações genóticas dificulta a classificação por técnicas que utilizem características visuais das asas, como a forma (ver Figura 3).

Tabela 2 – Acerto obtido nos testes da metodologia proposta com amostras do gênero feminino apenas.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx.(%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|-------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 66,11 ± 4,10 | 72,22 | B | 60,78 ± 1,96 | 63,33 |
| <i>SVM</i> | | 67,00 ± 4,32 | 74,44 | | 65,89 ± 5,29 | 75,56 |
| <i>MLP</i> | | 57,33 ± 6,11 | 64,44 | | 54,78 ± 6,18 | 65,56 |
| <i>LDA</i> | | 66,78 ± 5,75 | 76,67 | | 65,11 ± 4,60 | 71,11 |
| <i>QDA</i> | | 67,11 ± 5,37 | 76,67 | | 68,22 ± 2,58 | 71,11 |
| <i>KNN</i> | | 63,67 ± 5,59 | 73,33 | | 62,11 ± 5,17 | 70,00 |
| <i>RF</i> | C | 65,49 ± 2,27 | 67,73 | D | 65,52 ± 2,73 | 70,40 |
| <i>SVM</i> | | 70,51 ± 1,90 | 74,13 | | 70,48 ± 1,10 | 72,40 |
| <i>MLP</i> | | 58,21 ± 2,37 | 61,60 | | 57,50 ± 1,71 | 59,40 |
| <i>LDA</i> | | 67,01 ± 2,35 | 70,40 | | 67,38 ± 1,30 | 69,40 |
| <i>QDA</i> | | 70,43 ± 0,98 | 71,47 | | 70,36 ± 1,78 | 73,20 |
| <i>KNN</i> | | 63,52 ± 2,59 | 67,47 | | 63,14 ± 1,96 | 66,40 |

Fonte – O próprio autor.

Pode-se observar uma melhora significativa das taxas de acerto com as amostras de gênero feminino apenas. *SVM* atingiu taxas de acerto maior que *RF* e em alguns casos melhores que *QDA*, enquanto os outros classificadores mantiveram resultados consistentes, ver Tabela 2. O aumento do número de amostras para treinamento e a porcentagem de teste aumentou a taxa de acerto média na maioria dos testes com todos os genótipos. A utilização de apenas fêmeas aumentou a taxa de sucesso, mais expressivamente com *LDA* no grupo A, onde houve um incremento em torno de 20% de acurácia média. A maior taxa de acerto foi de 76,67 % com *LDA* e *QDA* no grupo A.

4.2 Gênero

A Tabela 3 mostra os resultados obtidos com a classificação de gênero e todos os genótipos e a Tabela 4 utiliza a mesma divisão das amostras, porém com apenas espécimes de *SAM*.

Com todos os genótipos, *kNN* e *MLP* obtiveram os menores resultados, seguidos de *RF*. *SVM* e *LDA* alcançaram as taxas médias de acerto mais elevadas, *QDA* obteve acurácias

Tabela 3 – Acerto obtido nos testes da metodologia proposta com todos genótipos.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | E | 87,28 ± 1,69 | 89,44 | F | 86,76 ± 1,33 | 89,00 |
| <i>SVM</i> | | 88,28 ± 1,45 | 90,00 | | 88,14 ± 1,10 | 89,60 |
| <i>MLP</i> | | 87,11 ± 1,48 | 89,44 | | 87,62 ± 1,13 | 89,00 |
| <i>LDA</i> | | 87,94 ± 1,76 | 90,56 | | 88,00 ± 0,81 | 89,20 |
| <i>QDA</i> | | 88,00 ± 2,47 | 91,67 | | 87,38 ± 1,27 | 89,20 |
| <i>KNN</i> | | 87,06 ± 1,78 | 90,00 | | 86,68 ± 1,16 | 89,00 |
| <i>RF</i> | G | 86,91 ± 1,21 | 89,20 | H | 87,86 ± 0,57 | 88,80 |
| <i>SVM</i> | | 88,52 ± 0,64 | 89,20 | | 88,15 ± 0,79 | 89,20 |
| <i>MLP</i> | | 87,71 ± 1,12 | 89,33 | | 88,15 ± 0,59 | 89,20 |
| <i>LDA</i> | | 88,03 ± 0,86 | 89,33 | | 87,76 ± 0,48 | 88,60 |
| <i>QDA</i> | | 88,60 ± 1,21 | 90,40 | | 88,67 ± 0,66 | 90,30 |
| <i>KNN</i> | | 86,75 ± 0,83 | 87,87 | | 87,49 ± 0,92 | 88,50 |

Fonte – O próprio autor.

máximas maiores em relação aos outros métodos. A taxa de sucesso mais alta foi de 91,67 % com *QDA* na discriminação do grupo E. Embora a semelhança física tenha prejudicado a classificação por genótipo, isso não influenciou os resultados de discriminação por gênero.

Tabela 4 – Acerto obtido nos testes da metodologia proposta com amostras de *SAM*.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | E | 91,39 ± 5,47 | 97,22 | F | 91,10 ± 2,77 | 96,00 |
| <i>SVM</i> | | 94,44 ± 2,93 | 97,22 | | 95,50 ± 2,64 | 99,00 |
| <i>MLP</i> | | 91,94 ± 4,80 | 97,22 | | 94,00 ± 2,45 | 98,00 |
| <i>LDA</i> | | 95,28 ± 3,22 | 100,00 | | 94,30 ± 2,41 | 98,00 |
| <i>QDA</i> | | 93,33 ± 3,51 | 97,22 | | 91,10 ± 2,56 | 94,00 |
| <i>KNN</i> | | 90,83 ± 3,94 | 97,22 | | 91,20 ± 2,66 | 95,00 |
| <i>RF</i> | G | 92,33 ± 2,42 | 95,33 | H | 93,45 ± 1,50 | 96,00 |
| <i>SVM</i> | | 93,73 ± 0,90 | 94,67 | | 95,55 ± 0,93 | 97,00 |
| <i>MLP</i> | | 92,67 ± 1,94 | 95,33 | | 94,45 ± 1,26 | 96,50 |
| <i>LDA</i> | | 94,13 ± 1,12 | 96,00 | | 95,80 ± 1,06 | 98,00 |
| <i>QDA</i> | | 92,07 ± 2,05 | 94,67 | | 94,90 ± 0,94 | 96,00 |
| <i>KNN</i> | | 91,80 ± 1,66 | 94,67 | | 93,85 ± 1,76 | 97,00 |

Fonte – O próprio autor.

Com amostras exclusivamente de *SAM*, os resultados foram semelhantes, embora *QDA* tenha diminuído a taxa de acerto em relação aos outros classificadores. O *LDA* obteve os melhores resultados médios e máximos, alcançando 100 % de acerto com o grupo E.

4.3 Comparação com literatura

A comparação entre os resultados alcançados neste trabalho e aqueles encontrados em outros trabalhos da literatura são mostrados nas Tabelas 5, que utiliza os valores destacados das Tabelas 1 e 2, e 6, que utiliza os valores destacados das Tabelas 3 e 4. Para a discriminação por genótipo, a abordagem proposta superou os resultados obtidos pelo trabalho comparado. Sonnenschein *et al.* (2015a) conseguiu taxas de acerto maiores ao usar pontos de referências (*landmarks* e *semi-landmarks*), o que requer a interação humana, contrário da metodologia proposta, que visa uma técnica automática. Os valores são mais baixos que as taxas atingidas pela classificação por gênero pela semelhança entre as asas de genótipos diferentes (ver Figura 3). A utilização de apenas amostras fêmeas, abordagem utilizada em Sonnenschein *et al.* (2015a), aumentou expressivamente a acurácia, acima de 8% nos classificadores comparados na Tabela 5.

Tabela 5 – Comparação dos resultados obtidos para classificação de genótipo com trabalho na literatura.

| Metodologia | Características | Classificador | Acurácia (%) |
|------------------------------------|------------------|------------------|--------------------------|
| Sonnenschein <i>et al.</i> (2015a) | Hessianos | RF (10) | 52,00 |
| | | SVM (linear) | 43,33 |
| Proposta | Dimensão fractal | RF (15) | 57,28* |
| | | | 65,52[†] |
| | | SVM (polinomial) | 55,64* |
| | | | 70,48[†] |

Fonte – O próprio autor.

Nota – * representa acerto nos testes com ambos os gêneros e [†], com apenas fêmeas.

A metodologia proposta também superou outras técnicas na discriminação de gênero (Tabela 6). Ahmad *et al.* (2014) atingiu 94% de acerto máximo com *SVM* e *MLBP*, excedido para 97% com a abordagem proposta e *SVM* nos testes com grupo H e apenas *SAM*. Sonnenschein *et al.* (2015a) alcançou taxas elevadas utilizando ponto de referência. Payne *et al.* (2013) utilizaram 300 amostras para treinamento, 200 para validação e 156 para teste, distribuídas uniformemente entre machos e fêmeas. Ahmad *et al.* (2014) utilizou dois grupos de teste, sendo comparado os resultados com o grupo 1, com 300 amostras, 150 machos e 150 fêmeas. Sonnenschein *et al.* (2015a), para genótipo, utilizou 70 amostras de cada classe para treinamento e 2 grupos com 15 amostras cada para teste (SONNENSCHNEIN *et al.*, 2015b). Para gênero, Sonnenschein *et al.* (2015a) utilizou 68 machos e 71 fêmeas de *SAM* para treinamento e 2 grupos de 15 machos e 15

Tabela 6 – Comparação dos resultados obtidos para classificação de gênero com trabalhos na literatura.

| Metodologia | Características | Classificador | Acurácia (%) |
|------------------------------------|--------------------|---------------------|--------------------------|
| Payne <i>et al.</i> (2013) | GEFE _{ML} | distância Manhattan | 73,16 |
| Ahmad <i>et al.</i> (2014) | MLBP LBP | RF | 90,00 |
| | | SVM (linear) | 89,50 |
| Sonnenschein <i>et al.</i> (2015a) | Hessianos | RF (10) | 85 |
| | | SVM (linear) | 81,70 |
| Proposta | Dimensão fractal | RF (15) | 87,86* |
| | | | 93,45[†] |
| | | SVM (polinomial) | 88,15* |
| | | | 95,55[†] |

Fonte – O próprio autor.

Nota – * representa acerto nos testes com todos os genótipos e [†], com apenas SAM.

fêmeas de SAM para teste (SONNENSCHNEIN *et al.*, 2015b). As acurácias mostradas na Tabela 6 para Payne *et al.* (2013) e Ahmad *et al.* (2014) são médias obtidas nos respectivos trabalhos. Na metodologia proposta, o treinamento e teste foram executados 10 vezes, onde a acurácia média é mostrada na Tabela 5 e 6.

5 CONCLUSÕES E TRABALHOS FUTUROS

Esta dissertação propõe uma metodologia para classificação do gênero e genótipo de moscas da espécie *Drosophila melanogaster* partir de características extraídas de imagens de asas.

A classificação por gênero se mostrou eficiente, alcançando mais de 90%, enquanto a classificação por genótipo obteve resultados mais baixos mas melhores que para os trabalhos comparados. Na classificação por gênero, primeiro foram utilizados todos os genótipos e, em seguida, com apenas amostras de *SAM*, o que melhorou o resultado. Na classificação por genótipo, primeiro foram utilizados ambos os gêneros e, posteriormente, apenas fêmeas, onde houve um aumento no desempenho. A semelhança na estrutura física das asas dificulta a discriminação das imagens por meios visuais, o que explica as baixas taxas de acerto para genótipo. No entanto, a metodologia proposta se mostrou robusta a um certo nível de translação, rotação, escalamento, oclusão e ruído quando se é utilizado um microscópio e zoom para ambos treinamento e teste. Ademais, é utilizado um número reduzido de atributos. Classificadores lineares obtiveram taxas de acerto mais alta na classificação por gênero, na proporção em que técnicas quadráticas foram melhores com genótipo. Esse resultado pode ter sido influenciado pela seleção manual exaustiva de parâmetros dos classificadores.

5.1 Contribuições

A classificação automatizada de imagens de asas de moscas *Drosophila melanogaster* é um estudo recente e ainda em desenvolvimento na literatura, em que requer iteração humana em muitos casos. Como contribuições desse trabalho, temos:

- Combinação de transformada wavelet não-decimada com filtro de Canny e dimensão fractal para a discriminação de gênero e genótipo;
- Comparação da taxa de acerto de classificadores de diferentes abordagens, como rede neural, árvores, discriminantes lineares e quadráticos, variando taxas de treinamento e teste;
- Proposta de uma nova metodologia de classificação automática com número reduzido de atributos.

5.2 Perspectivas de trabalhos futuros

Como forma de desenvolvimento do estudo iniciado nesse trabalho, perspectivas de trabalhos futuros envolvem:

- Analisar a metodologia proposta com famílias de *Wavelet* mais complexas, como Daubechie e chapéu mexicano;
- Comparar métodos de classificação mais robustos, como técnicas de aprendizado profundo;
- Validar a abordagem proposta com outras bases de imagens da literatura.

5.3 Trabalhos aceitos

Durante a pesquisa para a presente dissertação, os seguintes trabalhos foram aceitos em eventos da área científica:

- **NETO, F. G. M.**; LIMA, A. F.; MEDEIROS, F. N. S.; SOUZA, M. M. S.; JÚNIOR, I. C. de P. Classificação de espécies vegetais por atributos de forma e textura das folhas. In: ANAIS - ERIPI 2016. Teresina, 2016.
- FIGUEIREDO, J. C.; **NETO, F. G. M.**; PAULA, I. C. de. Contour-based feature extraction for image classification and retrieval. In: 2016 35th International Conference of the Chilean Computer Science Society (SCCC). Valparaiso, 2016. p. 1–7. Disponível em <<http://ieeexplore.ieee.org/document/7836058/>>.
- **NETO, F. G. M.**; JÚNIOR, R. F. P.; ROCHA, M. G. O.; SÁ JÚNIOR, J. J. de M.; DE PAULA JÚNIOR, I. C. Aprendizado profundo: conceitos, técnicas e estudo de caso com Java e análise de imagens. In: ERIPI 2017 Minicursos. Picos, 2017.
- **NETO, F. G. M.**; JÚNIOR, I. C. de P. Fractal Dimension-based Drosophila melanogaster Wing Genotype Classification. In: BRACIS 2017. Uberlândia, 2017.
- **NETO, F. G. M.**; BRAGA, Í. R.; HARBER, M. H.; JÚNIOR, I. C. de P. Drosophila melanogaster Gender Classification Based on Fractal Dimension. In: SIBGRAPI 2017. Niterói, 2017.

REFERÊNCIAS

- AHMAD, F.; ROY, K.; O'CONNOR, B.; SHELTON, J.; DOZIER, G.; DWORKIN, I. Fly wing biometrics using modified local binary pattern, svms and random forest. **International Journal of Machine Learning and Computing**, IACSIT Press, v. 4, n. 3, p. 279–285, 2014.
- BANKHEAD, P.; SCHOLFIELD, C. N.; MCGEOWN, J. G.; CURTIS, T. M. Fast retinal vessel detection and measurement using wavelets and edge location refinement. **PLoS ONE**, Public Library of Science, v. 7, n. 3, p. 1–12, 03 2012. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0032435>>.
- BEHNEL, S.; BRADSHAW, R.; CITRO, C.; DALCIN, L.; SELJEBOTN, D. S.; SMITH, K. Cython: The best of both worlds. **Computing in Science Engineering**, v. 13, n. 2, p. 31–39, March 2011. ISSN 1521-9615.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: ACM. **Proceedings of the fifth annual workshop on Computational learning theory**. [S.l.], 1992. p. 144–152.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. ISSN 1573-0565. Disponível em: <<http://dx.doi.org/10.1023/A:1010933404324>>.
- BRUNO, O. M.; PLOTZE, R. de O.; FALVO, M.; CASTRO, M. de. Fractal dimension applied to plant identification. **Information Sciences**, v. 178, n. 12, p. 2722 – 2733, 2008. ISSN 0020-0255. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0020025508000364>>.
- CANNY, J. A computational approach to edge detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-8, n. 6, p. 679–698, Nov 1986. ISSN 0162-8828.
- CHANG, C.-C.; LIN, C.-J. LIBSVM: A library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, v. 2, p. 27:1–27:27, 2011. Software disponível em <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, 1995. ISSN 1573-0565. Disponível em: <<http://dx.doi.org/10.1007/BF00994018>>.
- COSTA, L. da F.; JR., R. M. C. **Shape Classification and Analysis: Theory and Practice**. 2nd. ed. Boca Raton, FL, USA: CRC Press, Inc., 2009. ISBN 0849379296, 9780849379291.
- CRONE, S. F.; LESSMANN, S.; STAHLBOCK, R. The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. **European Journal of Operational Research**, v. 173, n. 3, p. 781 – 800, 2006. ISSN 0377-2217. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0377221705006739>>.
- DEBAT, V.; DEBELLE, A.; DWORKIN, I. Plasticity, canalization, and developmental stability of the drosophila wing: Joint effects of mutations and developmental temperature. **Evolution**, Blackwell Publishing Inc, v. 63, n. 11, p. 2864–2876, 2009. ISSN 1558-5646. Disponível em: <<http://dx.doi.org/10.1111/j.1558-5646.2009.00774.x>>.
- DEMIRHAN, A.; GÜLER İnan. Combining stationary wavelet transform and self-organizing maps for brain {MR} image segmentation. **Engineering Applications of Artificial Intelligence**, v. 24, n. 2, p. 358 – 367, 2011. ISSN 0952-1976. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0952197610001740>>.

DOBENS, A. C.; DOBENS, L. L. Fijiwings: An open source toolkit for semiautomated morphometric analysis of insect wings. **G3: Genes|Genomes|Genetics**, v. 3, n. 8, p. 1443–1449, 2013. Disponível em: <<http://www.g3journal.org/content/3/8/1443.abstract>>.

DOCUMENTATION of scikit-learn 0.18. 2017. Documentação. Disponível em: <<http://scikit-learn.org/stable/documentation.html>>.

EBADI, L.; SHAFRI, H. Z. M. A stable and accurate wavelet-based method for noise reduction from hyperspectral vegetation spectrum. **Earth Science Informatics**, v. 8, n. 2, p. 411–425, 2015. ISSN 1865-0481. Disponível em: <<http://dx.doi.org/10.1007/s12145-014-0168-0>>.

FIGUEIREDO, I.; KUMAR, S.; OLIVEIRA, C.; RAMOS, J.; ENGQUIST, B. Automated lesion detectors in retinal fundus images. **Computers in Biology and Medicine**, v. 66, p. 47 – 65, 2015. ISSN 0010-4825. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0010482515002851>>.

GRGIC, M.; RAVNJAK, M.; ZOVKO-CIHLAR, B. Filter comparison in wavelet transform of still images. In: **Industrial Electronics, 1999. ISIE '99. Proceedings of the IEEE International Symposium on**. [S.l.: s.n.], 1999. v. 1, p. 105–110 vol.1.

GROUVEN, U.; BERGEL, F.; SCHULTZ, A. Implementation of linear and quadratic discriminant analysis incorporating costs of misclassification. **Computer Methods and Programs in Biomedicine**, v. 49, n. 1, p. 55 – 60, 1996. ISSN 0169-2607. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0169260795017054>>.

HO, T. K. The random subspace method for constructing decision forests. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 8, p. 832–844, Aug 1998. ISSN 0162-8828.

HOULE, D.; MEZEY, J.; GALPERN, P.; CARTER, A. Automated measurement of drosophila wings. **BMC Evolutionary Biology**, v. 3, n. 1, p. 25, 2003. ISSN 1471-2148. Disponível em: <<http://dx.doi.org/10.1186/1471-2148-3-25>>.

KHALID, S.; JAMIL, U.; SALEEM, K.; AKRAM, M. U.; MANZOOR, W.; AHMED, W.; SOHAIL, A. Segmentation of skin lesion using cohen–daubechies–feauveau biorthogonal wavelet. **SpringerPlus**, v. 5, n. 1, p. 1603, 2016. ISSN 2193-1801. Disponível em: <<http://dx.doi.org/10.1186/s40064-016-3211-4>>.

KIM, K. S.; CHOI, H. H.; MOON, C. S.; MUN, C. W. Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions. **Current Applied Physics**, v. 11, n. 3, p. 740 – 745, 2011. ISSN 1567-1739. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1567173910004153>>.

KLINGENBERG, C. P. Morphoj: an integrated software package for geometric morphometrics. **Molecular Ecology Resources**, Blackwell Publishing Ltd, v. 11, n. 2, p. 353–357, 2011. ISSN 1755-0998. Disponível em: <<http://dx.doi.org/10.1111/j.1755-0998.2010.02924.x>>.

LIAW, A.; WIENER, M. Classification and regression by randomforest. **R news**, v. 2, n. 3, p. 18–22, 2002.

MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 7, p. 674–693, Jul 1989. ISSN 0162-8828.

MITTEROECKER, P.; GUNZ, P. Advances in geometric morphometrics. **Evolutionary Biology**, v. 36, n. 2, p. 235–247, Jun 2009. ISSN 1934-2845. Disponível em: <<https://doi.org/10.1007/s11692-009-9055-x>>.

MURTY, P. S. C.; REDDY, E. S.; BABU, I. R. Iris recognition system using fractal dimensions of haar patterns. **International Journal of Signal Processing, Image Processing and Pattern Recognition**, v. 2, n. 3, p. 75–84, 2009.

NGUYEN, H. N.; KAM, T. Y.; CHENG, P. Y. A novel automatic concrete surface crack identification using isotropic undecimated wavelet transform. In: **Intelligent Signal Processing and Communications Systems (ISPACS), 2012 International Symposium on**. [S.l.: s.n.], 2012. p. 766–771.

PAYNE, M.; TURNER, J.; SHELTON, J.; ADAMS, J.; CARTER, J.; WILLIAMS, H.; HANSEN, C.; DWORKIN, I.; DOZIER, G. Fly wing biometrics. In: **2013 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)**. [S.l.: s.n.], 2013. p. 42–46. ISSN 2325-4300.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

POLYCHRONAKI, G. E.; KTONAS, P. Y.; GATZONIS, S.; SIATOUNI, A.; ASVESTAS, P. A.; TSEKOU, H.; SAKAS, D.; NIKITA, K. S. Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. **Journal of Neural Engineering**, v. 7, n. 4, p. 046007, 2010. Disponível em: <<http://stacks.iop.org/1741-2552/7/i=4/a=046007>>.

PYWAVELETS - Discrete Wavelet Transforms in Python. 2017. Disponível em: <<https://github.com/PyWavelets/pywt>>.

RATHA, N. K.; CONNELL, J. H.; BOLLE, R. M. Enhancing security and privacy in biometrics-based authentication systems. **IBM Systems Journal**, v. 40, n. 3, p. 614–634, 2001. ISSN 0018-8670.

REED, R. D.; MARKS, R. J. **Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks**. Cambridge, MA, EUA: MIT Press, 1998. ISBN 0262181908.

ROBERTS, C. Biometric attack vectors and defences. **Computers & Security**, v. 26, n. 1, p. 14–25, 2007. ISSN 0167-4048. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016740480600215X>>.

ROSSUM, G. **Python Tutorial**. Amsterdam, Holanda: CWI (Centre for Mathematics and Computer Science), 1995.

SCHINDELIN, J.; ARGANDA-CARRERAS, I.; FRISE, E.; KAYNIG, V.; LONGAIR, M.; PIETZSCH, T.; PREIBISCH, S.; RUEDEN, C.; SAALFELD, S.; SCHMID, B.; TINEVEZ,

J.-Y.; WHITE, D. J.; HARTENSTEIN, V.; ELICEIRI, K.; TOMANCAK, P.; CARDONA, A. Fiji - an open source platform for biological image analysis. **Nature methods**, v. 9, n. 7, p. 676–682, jun. 2012. ISSN 1548-7105.

SCHINDELIN, J.; RUEDEN, C. T.; HINER, M. C.; ELICEIRI, K. W. The imagej ecosystem: An open platform for biomedical image analysis. **Molecular Reproduction and Development**, v. 82, n. 7-8, p. 518–529, 2015. ISSN 1098-2795. Disponível em: <<http://dx.doi.org/10.1002/mrd.22489>>.

SHELTON, J.; BRYANT, K.; ABRAMS, S.; SMALL, L.; ADAMS, J.; LEFLORE, D.; ALFORD, A.; RICANEK, K.; DOZIER, G. Genetic & evolutionary biometric security: Disposable feature extractors for mitigating biometric replay attacks. **Procedia Computer Science**, v. 8, p. 351 – 360, 2012. ISSN 1877-0509. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877050912000737>>.

SKIMAGE 0.12.2 docs. 2017. Documentação. Disponível em: <<http://scikit-image.org/docs/stable/>>.

SONNENSCHNEIN, A.; VANDERZEE, D.; PITCHERS, W. R.; CHARI, S.; DWORKIN, I. An image database of drosophila melanogaster wings for phenomic and biometric analysis. **GigaScience**, v. 4, n. 1, p. 25, 2015. ISSN 2047-217X. Disponível em: <<http://dx.doi.org/10.1186/s13742-015-0065-6>>.

SONNENSCHNEIN, A.; VANDERZEE, D.; PITCHERS, W. R.; CHARI, S.; DWORKIN, I. **Supporting material and data for "An Image Database of Drosophila melanogaster Wings for Phenomic and Biometric analysis"**. GigaScience Database, 2015. Disponível em: <<https://doi.org/10.5524/100141>>.

STARCK, J. L.; FADILI, J.; MURTAGH, F. The undecimated wavelet decomposition and its reconstruction. **IEEE Transactions on Image Processing**, v. 16, n. 2, p. 297–309, Feb 2007. ISSN 1057-7149.

U., R. A.; BHAT, P. S.; KANNATHAL, N.; RAO, A.; LIM, C. M. Analysis of cardiac health using fractal dimension and wavelet transformation. **ITBM-RBM**, v. 26, n. 2, p. 133 – 139, 2005. ISSN 1297-9562. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1297956205000227>>.

UNSER, M. Texture classification and segmentation using wavelet frames. **IEEE Transactions on Image Processing**, v. 4, n. 11, p. 1549–1560, Nov 1995. ISSN 1057-7149.

VOSS, R. F. Fractals in nature: From characterization to simulation. In: PEITGEN, H.-O.; SAUPE, D. (Ed.). **The Science of Fractal Images**. New York, NY: Springer New York, 1988. p. 21–70. ISBN 978-1-4612-3784-6. Disponível em: <https://doi.org/10.1007/978-1-4612-3784-6_1>.

WALT, S. van der; COLBERT, S. C.; VAROQUAUX, G. The numpy array: A structure for efficient numerical computation. **Computing in Science Engineering**, v. 13, n. 2, p. 22–30, March 2011. ISSN 1521-9615.

WALT, S. van der; SCHÖNBERGER, J. L.; Nunez-Iglesias, J.; BOULOGNE, F.; WARNER, J. D.; YAGER, N.; GOUILLART, E.; YU, T.; CONTRIBUTORS the scikit-image. scikit-image: image processing in Python. **PeerJ**, v. 2, p. e453, 6 2014. ISSN 2167-8359. Disponível em: <<http://dx.doi.org/10.7717/peerj.453>>.

WEBER, K. E.; DIGGINS, L. T. Increased selection response in larger populations. ii. selection for ethanol vapor resistance in *drosophila melanogaster* at two population sizes. **Genetics**, *Genetics*, v. 125, n. 3, p. 585–597, 1990. ISSN 0016-6731. Disponível em: <<http://www.genetics.org/content/125/3/585>>.

XAVIERAROCKIARAJ, S.; NITHYA, K.; DEVI, R. M. Brain tumor detection using modified histogram thresholding-quadrant approach. **Journal of Computer Applications (JCA)**, v. 5, n. 1, p. 21–25, 2012.

YAN, H.; JIANG, Y.; ZHENG, J.; PENG, C.; LI, Q. A multilayer perceptron-based medical decision support system for heart disease diagnosis. **Expert Systems with Applications**, v. 30, n. 2, p. 272 – 281, 2006. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417405001429>>.

ZELDITCH, M.; SWIDERSKI, D.; SHEETS, H. **Geometric Morphometrics for Biologists: A Primer**. 2. ed. San Diego: Elsevier Science, 2012. ISBN 9780123869043.

ZHOU, J.; LAMICHHANE, S.; STERNE, G.; YE, B.; PENG, H. Biocat: a pattern recognition platform for customizable biological image classification and annotation. **BMC Bioinformatics**, BioMed Central, v. 14, 2013. ISSN 1471-2105. Disponível em: <<http://dx.doi.org/10.1186/1471-2105-14-291>>.

ÖZTÜRK Şaban; AKDEMİR, B. Comparison of edge detection algorithms for texture analysis on glass production. **Procedia - Social and Behavioral Sciences**, v. 195, p. 2675 – 2682, 2015. ISSN 1877-0428. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877042815039567>>.

APÊNDICE A – RESULTADOS ADICIONAIS

Nesta Seção são mostrados os resultados adicionais obtidos. As imagens utilizadas foram adquiridas com microscópios Olympus e Leica, o primeiro com ampliação de 20X e o último com ampliação de 20X e 40X (SONNENSCHNEIN *et al.*, 2015a; SONNENSCHNEIN *et al.*, 2015b), mostradas na Seções seguintes. Mais informações sobre as imagens são expostas na Seção 2.3. Foram classificados o genótipo, com ambos os gêneros e apenas amostras de fêmeas, e gênero, com todos os genótipos e apenas amostras de SAM.

A.1 Resultados para as imagens do microscópio Olympus 20X.

A.1.1 Genótipo

Tabela 7 – Acerto obtido na classificação de genótipo com ambos os gêneros e imagens do microscópio Olympus e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 51,00 ± 5,63 | 60,00 | B | 55,78 ± 7,97 | 73,33 |
| <i>SVM</i> | | 51,56 ± 5,27 | 63,33 | | 55,56 ± 4,35 | 64,44 |
| <i>MLP</i> | | 47,00 ± 7,54 | 60,00 | | 51,33 ± 5,80 | 60,00 |
| <i>LDA</i> | | 50,56 ± 5,47 | 58,89 | | 54,00 ± 3,96 | 60,00 |
| <i>QDA</i> | | 53,56 ± 6,39 | 67,78 | | 54,78 ± 7,16 | 66,67 |
| <i>KNN</i> | | 47,67 ± 5,06 | 56,67 | | 58,22 ± 4,39 | 66,67 |
| <i>RF</i> | C | 59,81 ± 1,81 | 63,20 | D | 59,08 ± 1,85 | 61,80 |
| <i>SVM</i> | | 60,08 ± 2,07 | 63,20 | | 59,32 ± 1,33 | 61,00 |
| <i>MLP</i> | | 53,20 ± 2,43 | 56,80 | | 52,80 ± 3,44 | 56,60 |
| <i>LDA</i> | | 57,23 ± 2,84 | 61,07 | | 56,36 ± 2,97 | 60,00 |
| <i>QDA</i> | | 57,31 ± 2,17 | 60,53 | | 58,20 ± 1,57 | 59,60 |
| <i>KNN</i> | | 55,17 ± 1,81 | 58,13 | | 57,80 ± 1,59 | 59,80 |

Fonte – O próprio autor.

Tabela 8 – Acerto obtido na classificação de genótipo com apenas fêmeas e imagens do microscópio Olympus e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 66,89 ± 6,39 | 76,67 | B | 67,22 ± 4,03 | 72,22 |
| <i>SVM</i> | | 64,89 ± 4,00 | 71,11 | | 65,11 ± 6,48 | 75,56 |
| <i>MLP</i> | | 56,00 ± 4,81 | 64,44 | | 57,00 ± 5,08 | 65,56 |
| <i>LDA</i> | | 60,78 ± 3,36 | 67,78 | | 61,89 ± 5,19 | 67,78 |
| <i>QDA</i> | | 64,89 ± 4,60 | 71,11 | | 65,22 ± 4,26 | 72,22 |
| <i>KNN</i> | | 62,33 ± 3,68 | 66,67 | | 62,44 ± 4,18 | 68,89 |
| <i>RF</i> | C | 68,45 ± 1,58 | 71,20 | D | 69,74 ± 1,82 | 72,40 |
| <i>SVM</i> | | 65,01 ± 2,31 | 68,53 | | 66,74 ± 1,31 | 68,20 |
| <i>MLP</i> | | 55,95 ± 3,38 | 61,87 | | 56,62 ± 2,23 | 59,80 |
| <i>LDA</i> | | 62,43 ± 1,95 | 66,40 | | 63,08 ± 1,84 | 65,80 |
| <i>QDA</i> | | 65,68 ± 1,08 | 67,73 | | 66,24 ± 1,55 | 68,20 |
| <i>KNN</i> | | 62,45 ± 3,03 | 66,67 | | 65,72 ± 1,93 | 70,40 |

Fonte – O próprio autor.

A.1.2 Gênero

Tabela 9 – Acerto obtido na classificação de gênero com todos os genótipos] e imagens do microscópio Olympus e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | E | 77,22 ± 3,01 | 82,22 | F | 75,86 ± 1,31 | 77,40 |
| <i>SVM</i> | | 77,33 ± 3,25 | 78,33 | | 72,78 ± 1,44 | 75,40 |
| <i>MLP</i> | | 71,17 ± 4,29 | 78,33 | | 72,02 ± 2,95 | 77,20 |
| <i>LDA</i> | | 72,28 ± 3,79 | 78,89 | | 71,68 ± 1,69 | 74,40 |
| <i>QDA</i> | | 74,33 ± 3,16 | 78,89 | | 73,00 ± 1,13 | 74,60 |
| <i>KNN</i> | | 76,83 ± 1,70 | 78,89 | | 73,90 ± 2,05 | 76,80 |
| <i>RF</i> | G | 75,09 ± 1,22 | 76,53 | H | 75,95 ± 0,83 | 77,30 |
| <i>SVM</i> | | 71,68 ± 0,95 | 73,07 | | 71,62 ± 1,46 | 73,30 |
| <i>MLP</i> | | 71,25 ± 1,00 | 72,80 | | 70,74 ± 1,51 | 72,40 |
| <i>LDA</i> | | 71,43 ± 0,90 | 72,80 | | 70,87 ± 1,59 | 72,90 |
| <i>QDA</i> | | 72,89 ± 1,30 | 74,80 | | 72,42 ± 1,48 | 74,10 |
| <i>KNN</i> | | 74,36 ± 0,96 | 75,47 | | 74,95 ± 1,44 | 76,60 |

Fonte – O próprio autor.

Tabela 10 – Acerto obtido na classificação de gênero com amostras de SAM e imagens do microscópio Olympus e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | E | 78,61 ± 7,75 | 91,67 | F | 77,00 ± 4,88 | 82,00 |
| <i>SVM</i> | | 78,89 ± 8,20 | 91,67 | | 78,80 ± 3,85 | 85,00 |
| <i>MLP</i> | | 70,56 ± 7,43 | 77,78 | | 76,10 ± 4,46 | 81,00 |
| <i>LDA</i> | | 79,44 ± 7,31 | 91,67 | | 77,70 ± 3,97 | 84,00 |
| <i>QDA</i> | | 78,61 ± 3,48 | 86,11 | | 79,70 ± 2,87 | 84,00 |
| <i>KNN</i> | | 72,22 ± 6,93 | 83,33 | | 72,30 ± 3,06 | 77,00 |
| <i>RF</i> | G | 78,80 ± 1,50 | 81,33 | H | 80,75 ± 3,80 | 85,00 |
| <i>SVM</i> | | 78,87 ± 1,96 | 83,33 | | 78,75 ± 1,99 | 82,00 |
| <i>MLP</i> | | 76,00 ± 3,47 | 84,00 | | 77,25 ± 3,01 | 80,50 |
| <i>LDA</i> | | 78,47 ± 2,25 | 81,33 | | 80,35 ± 2,68 | 82,50 |
| <i>QDA</i> | | 77,93 ± 2,25 | 80,67 | | 81,90 ± 2,34 | 87,00 |
| <i>KNN</i> | | 75,07 ± 3,28 | 80,67 | | 75,60 ± 2,57 | 80,50 |

Fonte – O próprio autor.

A.2 Resultados para as imagens do microscópio Leica 20X.

A.2.1 Genótipo

Tabela 11 – Acerto obtido na classificação de genótipo com ambos os gêneros e imagens do microscópio Leica e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 56,78 ± 4,55 | 65,56 | B | 54,67 ± 3,35 | 58,89 |
| <i>SVM</i> | | 55,33 ± 5,68 | 63,33 | | 54,67 ± 4,56 | 61,11 |
| <i>MLP</i> | | 50,78 ± 5,52 | 61,11 | | 51,00 ± 3,94 | 56,67 |
| <i>LDA</i> | | 55,67 ± 4,93 | 62,22 | | 55,44 ± 5,20 | 63,33 |
| <i>QDA</i> | | 59,00 ± 4,52 | 65,56 | | 58,00 ± 5,31 | 65,56 |
| <i>KNN</i> | | 54,78 ± 5,67 | 63,33 | | 53,22 ± 4,33 | 56,67 |
| <i>RF</i> | C | 55,73 ± 2,18 | 58,67 | D | 57,80 ± 1,76 | 60,80 |
| <i>SVM</i> | | 56,16 ± 0,98 | 57,87 | | 55,78 ± 2,01 | 58,60 |
| <i>MLP</i> | | 52,05 ± 2,72 | 55,73 | | 50,76 ± 1,93 | 54,60 |
| <i>LDA</i> | | 57,87 ± 1,77 | 60,53 | | 57,20 ± 1,23 | 59,80 |
| <i>QDA</i> | | 61,17 ± 1,92 | 64,80 | | 60,24 ± 1,59 | 63,00 |
| <i>KNN</i> | | 55,41 ± 2,14 | 58,93 | | 56,20 ± 1,37 | 58,60 |

Fonte – O próprio autor.

Tabela 12 – Acerto obtido na classificação de genótipo com apenas fêmeas e imagens do microscópio Leica e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx.(%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|-------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 61,00 ± 6,38 | 70,00 | B | 64,22 ± 5,31 | 71,11 |
| <i>SVM</i> | | 62,78 ± 3,11 | 66,67 | | 62,89 ± 4,06 | 67,78 |
| <i>MLP</i> | | 57,89 ± 3,76 | 62,22 | | 62,22 ± 3,78 | 66,67 |
| <i>LDA</i> | | 59,33 ± 4,78 | 65,56 | | 58,56 ± 3,36 | 63,33 |
| <i>QDA</i> | | 65,67 ± 2,84 | 70,00 | | 64,67 ± 3,77 | 71,11 |
| <i>KNN</i> | | 59,11 ± 4,05 | 66,67 | | 64,22 ± 4,25 | 71,11 |
| <i>RF</i> | C | 63,73 ± 1,61 | 66,13 | D | 64,02 ± 2,49 | 67,40 |
| <i>SVM</i> | | 66,29 ± 1,08 | 68,00 | | 66,08 ± 2,12 | 71,20 |
| <i>MLP</i> | | 61,36 ± 2,68 | 66,40 | | 62,00 ± 3,11 | 65,40 |
| <i>LDA</i> | | 62,75 ± 1,93 | 65,87 | | 61,42 ± 2,86 | 66,40 |
| <i>QDA</i> | | 68,19 ± 1,96 | 72,00 | | 66,70 ± 1,90 | 70,40 |
| <i>KNN</i> | | 62,59 ± 1,53 | 65,60 | | 62,06 ± 1,54 | 64,60 |

Fonte – O próprio autor.

A.2.2 Gênero

Tabela 13 – Acerto obtido na classificação de gênero com todos os genótipos e imagens do microscópio Leica e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | E | 66,50 ± 2,50 | 71,67 | F | 66,46 ± 2,29 | 70,80 |
| <i>SVM</i> | | 67,50 ± 2,66 | 72,22 | | 65,12 ± 1,43 | 68,20 |
| <i>MLP</i> | | 68,28 ± 3,16 | 73,89 | | 67,18 ± 2,00 | 71,20 |
| <i>LDA</i> | | 68,11 ± 3,65 | 72,78 | | 66,14 ± 1,21 | 68,80 |
| <i>QDA</i> | | 65,17 ± 3,70 | 72,22 | | 65,38 ± 2,57 | 70,40 |
| <i>KNN</i> | | 65,06 ± 3,84 | 69,44 | | 64,10 ± 1,42 | 66,00 |
| <i>RF</i> | G | 67,39 ± 1,67 | 70,40 | H | 68,04 ± 1,49 | 70,30 |
| <i>SVM</i> | | 67,95 ± 1,18 | 69,60 | | 67,62 ± 1,39 | 69,40 |
| <i>MLP</i> | | 69,01 ± 1,71 | 70,93 | | 68,16 ± 1,60 | 70,10 |
| <i>LDA</i> | | 67,71 ± 1,38 | 70,00 | | 67,41 ± 1,45 | 69,70 |
| <i>QDA</i> | | 64,85 ± 2,45 | 69,47 | | 65,19 ± 2,12 | 67,60 |
| <i>KNN</i> | | 65,97 ± 0,95 | 67,47 | | 66,07 ± 1,81 | 68,60 |

Fonte – O próprio autor.

Tabela 14 – Acerto obtido na classificação de gênero com amostras de SAM e imagens do microscópio Leica e ampliação 20X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | E | 74,72 ± 5,31 | 86,11 | F | 70,00 ± 3,94 | 75,00 |
| <i>SVM</i> | | 73,06 ± 6,29 | 80,56 | | 71,90 ± 4,58 | 77,00 |
| <i>MLP</i> | | 70,00 ± 5,68 | 77,78 | | 71,40 ± 3,86 | 76,00 |
| <i>LDA</i> | | 70,56 ± 6,17 | 77,78 | | 70,60 ± 4,33 | 78,00 |
| <i>QDA</i> | | 61,94 ± 8,39 | 77,78 | | 62,10 ± 4,51 | 73,00 |
| <i>KNN</i> | | 72,22 ± 4,14 | 77,78 | | 71,60 ± 3,86 | 77,00 |
| <i>RF</i> | G | 74,73 ± 3,42 | 80,00 | H | 72,10 ± 2,87 | 76,50 |
| <i>SVM</i> | | 75,53 ± 3,28 | 80,67 | | 76,65 ± 2,25 | 80,50 |
| <i>MLP</i> | | 72,67 ± 5,04 | 78,00 | | 74,70 ± 2,63 | 78,00 |
| <i>LDA</i> | | 73,60 ± 2,69 | 78,67 | | 74,10 ± 2,33 | 78,50 |
| <i>QDA</i> | | 64,80 ± 3,90 | 71,33 | | 65,70 ± 2,37 | 71,50 |
| <i>KNN</i> | | 73,93 ± 3,85 | 80,67 | | 72,25 ± 2,57 | 75,50 |

Fonte – O próprio autor.

A.3 Resultados para as imagens do microscópio Leica 40X.

A.3.1 Genótipo

Tabela 15 – Acerto obtido na classificação de genótipo com ambos os gêneros e imagens do microscópio Leica e ampliação 40X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 52,78 ± 3,86 | 57,78 | B | 57,78 ± 5,93 | 67,78 |
| <i>SVM</i> | | 57,89 ± 6,12 | 68,89 | | 58,44 ± 4,60 | 64,44 |
| <i>MLP</i> | | 48,44 ± 6,35 | 62,22 | | 48,11 ± 6,93 | 56,67 |
| <i>LDA</i> | | 57,33 ± 5,98 | 66,67 | | 59,33 ± 4,23 | 65,56 |
| <i>QDA</i> | | 56,56 ± 5,96 | 65,56 | | 56,78 ± 4,84 | 63,33 |
| <i>KNN</i> | | 53,56 ± 4,44 | 63,33 | | 55,22 ± 3,63 | 61,11 |
| <i>RF</i> | C | 58,32 ± 1,68 | 60,53 | D | 58,92 ± 2,35 | 63,20 |
| <i>SVM</i> | | 63,01 ± 1,72 | 65,60 | | 61,44 ± 1,48 | 63,60 |
| <i>MLP</i> | | 52,85 ± 2,08 | 56,00 | | 52,22 ± 4,12 | 58,40 |
| <i>LDA</i> | | 63,60 ± 1,02 | 64,80 | | 61,60 ± 1,51 | 64,00 |
| <i>QDA</i> | | 61,17 ± 1,64 | 63,73 | | 58,88 ± 0,90 | 60,40 |
| <i>KNN</i> | | 57,20 ± 1,80 | 59,20 | | 57,02 ± 2,51 | 60,60 |

Fonte – O próprio autor.

Tabela 16 – Acerto obtido na classificação de genótipo com apenas fêmeas e imagens do microscópio Leica e ampliação 40X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx.(%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|-------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | A | 78,33 ± 5,37 | 86,67 | B | 73,22 ± 3,79 | 76,67 |
| <i>SVM</i> | | 78,87 ± 4,65 | 85,56 | | 75,00 ± 3,93 | 80,00 |
| <i>MLP</i> | | 73,78 ± 6,55 | 83,33 | | 69,78 ± 3,91 | 74,44 |
| <i>LDA</i> | | 77,11 ± 3,02 | 81,11 | | 77,67 ± 3,41 | 82,22 |
| <i>QDA</i> | | 78,00 ± 3,18 | 83,33 | | 76,56 ± 3,94 | 83,33 |
| <i>KNN</i> | | 76,44 ± 4,88 | 86,67 | | 71,44 ± 5,37 | 77,78 |
| <i>RF</i> | C | 76,77 ± 1,78 | 78,67 | D | 77,34 ± 2,17 | 81,00 |
| <i>SVM</i> | | 76,96 ± 1,32 | 78,93 | | 77,06 ± 1,41 | 79,80 |
| <i>MLP</i> | | 70,00 ± 2,86 | 75,20 | | 69,18 ± 3,01 | 73,40 |
| <i>LDA</i> | | 77,52 ± 1,48 | 80,80 | | 77,08 ± 1,23 | 79,40 |
| <i>QDA</i> | | 76,51 ± 1,18 | 78,40 | | 77,74 ± 1,61 | 80,00 |
| <i>KNN</i> | | 75,33 ± 1,93 | 77,60 | | 75,38 ± 1,74 | 77,60 |

Fonte – O próprio autor.

A.3.2 Gênero

Tabela 17 – Acerto obtido na classificação de gênero com todos os genótipos e imagens do microscópio Leica e ampliação 40X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | E | 80,17 ± 2,65 | 84,44 | F | 79,34 ± 2,29 | 82,20 |
| <i>SVM</i> | | 78,44 ± 2,08 | 82,22 | | 77,64 ± 1,48 | 79,60 |
| <i>MLP</i> | | 77,11 ± 1,94 | 79,44 | | 78,14 ± 1,66 | 79,80 |
| <i>LDA</i> | | 75,94 ± 2,33 | 79,44 | | 75,42 ± 1,52 | 78,00 |
| <i>QDA</i> | | 75,22 ± 1,91 | 77,78 | | 74,46 ± 2,61 | 79,00 |
| <i>KNN</i> | | 79,06 ± 2,92 | 84,44 | | 78,60 ± 1,69 | 81,40 |
| <i>RF</i> | G | 79,75 ± 1,58 | 82,93 | H | 81,01 ± 0,67 | 82,50 |
| <i>SVM</i> | | 77,41 ± 1,09 | 79,20 | | 78,44 ± 0,67 | 79,30 |
| <i>MLP</i> | | 77,20 ± 1,53 | 79,60 | | 78,07 ± 1,27 | 79,70 |
| <i>LDA</i> | | 74,77 ± 1,41 | 77,47 | | 75,29 ± 1,18 | 78,10 |
| <i>QDA</i> | | 74,20 ± 2,10 | 79,33 | | 76,19 ± 1,95 | 79,90 |
| <i>KNN</i> | | 79,36 ± 1,49 | 82,40 | | 80,25 ± 0,89 | 82,30 |

Fonte – O próprio autor.

Tabela 18 – Acerto obtido na classificação de gênero com amostras de SAM e imagens do microscópio Leica e ampliação 40X.

| Class. | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) | Grupo | Acerto médio ± des. pad. (%) | Acerto máx. (%) |
|------------|-------|---------------------------------|--------------------|-------|---------------------------------|--------------------|
| <i>RF</i> | | 86,39 ± 7,34 | 97,22 | | 85,00 ± 3,83 | 90,00 |
| <i>SVM</i> | | 88,06 ± 5,56 | 94,44 | | 86,40 ± 3,57 | 92,00 |
| <i>MLP</i> | E | 85,28 ± 6,29 | 91,67 | F | 85,70 ± 3,13 | 90,00 |
| <i>LDA</i> | | 85,83 ± 5,15 | 91,67 | | 84,70 ± 3,06 | 89,00 |
| <i>QDA</i> | | 84,72 ± 5,75 | 91,67 | | 85,20 ± 3,22 | 89,00 |
| <i>KNN</i> | | 75,28 ± 8,33 | 88,89 | | 76,50 ± 2,01 | 80,00 |
| <i>RF</i> | | 82,60 ± 2,38 | 85,33 | | 82,10 ± 2,56 | 87,00 |
| <i>SVM</i> | G | 83,93 ± 2,21 | 86,67 | H | 83,15 ± 1,86 | 86,00 |
| <i>MLP</i> | | 83,33 ± 1,81 | 86,67 | | 83,10 ± 1,68 | 85,50 |
| <i>LDA</i> | | 83,47 ± 1,36 | 84,67 | | 83,20 ± 2,04 | 86,00 |
| <i>QDA</i> | | 82,13 ± 2,03 | 86,00 | | 81,25 ± 3,06 | 87,00 |
| <i>KNN</i> | | 76,87 ± 3,06 | 80,00 | | 79,20 ± 2,29 | 84,00 |

Fonte – O próprio autor.