FELIPE LENO DA SILVA

AUTOMATED BEE SPECIES IDENTIFICATION THROUGH
WING IMAGES

São Paulo
2015

FELIPE LENO DA SILVA

IDENTIFICAÇÃO AUTOMATIZADA DE ESPÉCIES DE
ABELHAS ATRAVÉS DE IMAGENS DE ASAS.

São Paulo
2015

FELIPE LENO DA SILVA

AUTOMATED BEE SPECIES IDENTIFICATION THROUGH WING IMAGES

Dissertation submitted to Escola Politécnica
da Universidade de São Paulo in partial
fulfillment of the requirements for the degree
of Master of Science

Area of research:
Computer Engineering

Advisor: Profa. Anna Helena Reali Costa

São Paulo
2015

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, de fevereiro de 2015.

Assinatura do autor _____

Assinatura do orientador _____

Catálogo-na-publicação

Silva, Felipe Leno da
Automated bee species identification through wing images /
F.L. da Silva. – versão corr. -- São Paulo, 2015.
82 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade
de São Paulo. Departamento de Engenharia de Computação e
Sistemas Digitais.

1.Inteligência artificial 2.Visão computacional 3.Reconheci-
mento de padrões 4.Aprendizagem de máquina I. Universidade
de São Paulo. Escola Politécnica. Departamento de Engenharia
de Computação e Sistemas Digitais II.t.

ACKNOWLEDGMENTS

Firstly, I thank my advisor, Anna Helena Reali Costa, for all the guidance and encouragement that were essential in these last years.

I would also like to express my gratitude to my family for their support, specially my mother, Solange; my father, Berto; and my sister, Paula.

I wish to thank prof. Tiago Mauricio Franco for kindly providing all the necessary data of wing images.

I also thank all members of LTI (Laboratório de Técnicas Inteligentes - USP) for all the valuable discussions regarding my research.

My special thanks for my friends Ricardo Jacomini, Heider Berlink, Juan Perafan, Walter Mayor and Allan Lima for all the support during these two years.

Finally, I gratefully acknowledge financial support from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and from the NAP Biocomp.

"The only true wisdom is knowing you know nothing."

Socrates

"If there is no struggle, there is no progress."

Frederick Douglass

ABSTRACT

Several researches focus on the study and conservation of bees, largely because of its importance for agriculture. However, the identification of bee species has hampering new studies, since it demands a very specialized knowledge and is time demanding. Although there are several methods to accomplish this task, many of them are excessively costly, restricting its applicability. For being accessible, the bee wings have been widely used for the extraction of features, since it is possible to apply morphometric techniques using just one image of the wing. As the manual measurement of various features is tedious and error prone, some systems have been developed for this purpose. However, these systems also have limitations, and there is no study concerning classification techniques that can be used for this purpose. This research aims to evaluate the feature extraction and classification techniques in order to determine the combination of more appropriate techniques for discriminating species of bees. The results of our research indicate that the use of a conjunction of Morphometric and Pixel-based features is more effective than only using Morphometric features. Our analysis also concluded that the best classification algorithms using both only Morphometric features and a conjunction of Morphometric and Pixel-based features are, respectively, Naïve Bayes and Logistic classifier. The results of this research can guide the development of new systems to identify bee species in order to assist in researches conducted by biologists.

Keywords: Artificial Intelligence, Machine Learning, Computer Vision, Supervised Learning, Bee Species Recognition, Feature Extraction, Pattern Recognition, Feature Selection.

RESUMO

Diversas pesquisas focam no estudo e conservação das abelhas, em grande parte por sua importância para a agricultura. Entretanto, a identificação de espécies de abelhas vem sendo um impedimento para a condução de novas pesquisas, já que demanda tempo e um conhecimento muito especializado. Apesar de existirem diversos métodos para realizar esta tarefa, muitos deles são excessivamente custosos, restringindo sua aplicabilidade. Por serem facilmente acessíveis, as asas das abelhas vêm sendo amplamente utilizadas para a extração de características, já que é possível aplicar técnicas morfométricas utilizando apenas uma foto da asa. Como a medição manual de diversas características é tediosa e propensa a erros, sistemas foram desenvolvidos com este propósito. Entretanto, os sistemas ainda possuem limitações e não há um estudo voltado às técnicas de classificação que podem ser utilizadas para este fim. Esta pesquisa visa avaliar as técnicas de extração de características e classificação de modo a determinar o conjunto de técnicas mais apropriado para a discriminação de espécies de abelhas. Nesta pesquisa foi demonstrado que o uso de uma conjunção de características morfométricas e fotométricas obtêm melhores resultados que o uso de somente características morfométricas. Também foram analisados os melhores algoritmos de classificação tanto usando somente características morfométricas, quanto usando uma conjunção de características morfométricas e fotométricas, os quais são, respectivamente, o Naïve Bayes e o classificador Logístico. Os Resultados desta pesquisa podem guiar o desenvolvimento de novos sistemas para identificação de espécies de abelha, objetivando auxiliar pesquisas conduzidas por biólogos.

Palavras-chave: Inteligência Artificial, Aprendizagem de Máquina, Visão Computacional, Aprendizado Supervisionado, Classificação de Abelhas, Extração de Características, Reconhecimento de Padrões, Seleção de Características.

LIST OF ABBREVIATIONS

ABIS	Automated Bee Identification System.
ANN	Artificial Neural Network.
CS	Centroid Size.
GR	Gain Ratio.
IG	Information Gain.
KNN	K Nearest Neighbors.
LDA	Linear Discriminant Analysis.
MLP	Multilayer Perceptron.
SVM	Support Vector Machine.

LIST OF SYMBOLS

α	Trade-off parameter to calculate matrix W .
β_{jc}	Regression coefficient associated to feature j and class c in a Logistic classifier.
Λ	Eigenvalues matrix.
ρ	Level of confidence to define the error margin.
θ	Rotation angle of a wing image.
σ	Standard deviation.
γ	Trade-off parameter for Fisher Separation Criterion.
∇f	Gradient of function f .
$(\bar{x}, \bar{y})$	Centroid coordinates.
(x', y')	Distances from a point to the center of mass of an image in, respectively, the x-axis and y-axis.
(x_c, y_c)	Coordinates of the center of mass of an image.
(x_{coni}, y_{coni})	Coordinates of landmark i of a consensus object.
(x_i, y_i)	Coordinates of landmark i .
$A \oplus B$	Dilation of a set A by an structuring element B .
$A \ominus B$	Erosion of a set A by an structuring element B .
$A \circledast B$	Hit-or-miss operation of the set A , where $B = B_1, B_2$. B_1 is the desired shape and B_2 is the background.
$A \oslash B$	Thinning operation of the set A by an structuring element B .
$B_{ L }^{-1}$	Bending energy matrix.
C	Correlation matrix.
$cov(x_1, x_2)$	Covariation between x_1 and x_2 .

$\mathbf{E}$	Eigenvectors matrix.
$\mathcal{E}$	Error margin.
F	Feature set.
F^d	Feature subset with d features.
f_j	Refers to feature j .
$f(x, y)$	Two-dimensional function representing an image.
$g(x, y)$	Image $f(x, y)$ after its modification by an image processing method.
$h(r_k)$	Histogram function of an image.
$\mathbf{I}$	Identity matrix.
K	number of classes
$\mathbf{L}$	Set of landmarks.
$ \mathbf{L} $	Number of landmarks.
$\mathbf{L}'$	Set of landmarks invariant to affine transformation.
m_{ij}	Second moment of an image.
n	Number of samples in a training set.
n_i	Number of samples of class i .
$\mathbf{R}'$	Relative warps scores matrix.
$\mathbf{W}$	Weight matrix of principal warps.
$\mathbf{X}$	Matrix with feature values of all samples in a training set.
x_{ij}	Value of feature j in the sample i .
$\mathbf{x}$	Vector of feature values that represents a sample.
$z_{\rho/2}$	Critical value for level of confidence ρ .
$\mathbf{A} \otimes \mathbf{B}$	Direct product of matrix $\mathbf{A}$ by matrix $\mathbf{B}$.
N	Number of experiments when performing the Wilcoxon Signed-Rank Test.

$v_{y,i}$	Accuracy value achieved by classifier y on the experiment i .
R_i	Rank of accuracy differences on experiment i when performing the Wilcoxon Signed-Rank Test.
W	Test statistic for the Wilcoxon Signed-Rank Test.

TABLE OF CONTENTS

1	INTRODUCTION	19
1.1	Objectives	20
1.2	Organization	20
1.3	Contributions	21
2	BACKGROUND	23
2.1	Background on Bee Identification	23
2.1.1	Bee Species Identification Process Overview	23
2.1.2	Landmark	24
2.2	Image Processing	25
2.2.1	Image Histogram	25
2.2.2	Image Gradient	26
2.2.3	Thresholding	26
2.2.4	Morphological Image Processing	27
2.2.4.1	Notation	27
2.2.4.2	Dilation	28
2.2.4.3	Erosion	29
2.2.4.4	Hit-or-Miss	29
2.2.4.5	Thinning	29
2.3	Feature Extraction	29
2.3.1	Traditional Morphometrics	30
2.3.2	Geometric Morphometrics	30
2.3.2.1	Centroid Size	30
2.3.2.2	Aligned Coordinates	32
2.3.2.3	Principal Warps	32
2.3.2.4	Relative Warps	34
2.4	Classification	34
2.4.1	Linear Discriminant Analysis	35
2.4.2	Naïve Bayes	36
2.4.3	Logistic	36
2.4.4	K Nearest Neighbors	37
2.4.5	C4.5	37

2.4.6	Multilayer Perceptron	38
2.4.7	Support Vector Machine	38
2.5	Feature Selection	39
2.5.1	Information Gain	40
2.5.2	Chi-Squared	40
2.5.3	Covariance	40
2.5.4	Fisher Separation Criterion	41
2.6	Cross-Validation	41
2.6.1	K-fold Cross-Validation	41
2.6.2	Stratified Cross-Validation	42
2.6.3	Error Margin	42
2.6.4	Wilcoxon Signed Rank Test	42
2.7	Relation Between Concepts	43
3	RELATED RESEARCHES	45
3.1	DrawWing	45
3.2	ABIS	47
3.3	tps	48
3.4	Unsolved Problems	48
4	PROPOSAL DETAILING AND EXPERIMENTAL SETUP	51
4.1	Training Set Definition	51
4.2	Feature Extraction	53
4.2.1	Morphometric Features	53
4.2.2	Pixel-based Features	56
4.2.3	Resulting Feature Vector	56
4.3	Classification Algorithms	57
4.4	Experimental Setup	59
5	RESULTS AND DISCUSSION	61
5.1	Experiment 1 - Euglossa training set using Morphometric Features . . .	61
5.2	Experiment 2 - Euglossa training set using Morphometric and Pixel-based features	64
5.3	Experiment 3 - Apis training set using Morphometric Features	67
5.4	Experiment 4 - Apis training set using Morphometric and Pixel-based features	70
5.5	Concluding Remarks	73
6	CONCLUSION AND FURTHER WORK	75

1 INTRODUCTION

Pollination is an important ecosystem service, as it is directly related to the conservation of natural environments and to food production in commercial crops (COSTANZA et al., 1997; KLEIN et al., 2007). About 87% of all crops used for human consumption are dependent on insect pollination (KLEIN et al., 2007). In the year 2005, the worldwide economic value of pollination was estimated as 9.5% of the total value of agricultural production, thus accounting for €153 billion (GALLAI et al., 2009).

The main pollinator globally available is the honey bee *Apis mellifera*. This bee has been facing the so-called Colony Collapse Disorder in the USA and similar phenomena in Europe. The disorder caused, in the last years, the death of hundreds of thousands of colonies (VANENGELSDORP et al., 2008). The global stocks of managed *Apis mellifera* bees have grown in the last decades, although in a rate much slower than the demand for commercial pollinators, generating several concerns about this problem (POTTS et al., 2010). It is urgent to preserve this important pollinator; the accurate identification of the different subspecies is a measure that can help in that goal.

Due to their great importance to agriculture, various researches have been conducted aiming at the study and conservation of bee species. However, some bee species are really difficult to distinguish, thus, bee species identification has been hampering the progress of researches in this area, since the correct identification of some species requires time and specialized knowledge.

Some techniques can be used for species identification, e.g., isoenzymes or DNA analysis, but molecular and biochemical methods are expensive (FRANCOY et al., 2008). Aiming a cheaper procedure, the identification through manually measured morphometric features from wings, sternites and legs was developed (DALY; BALLING, 1978; RINDERER et al., 1986). Nevertheless, the manual measurement of several physical features requires time and expertise (FRANCOY et al., 2008). More recent studies have shown that features extracted from patterns of wing venation are good discriminatory elements to differentiate among insect species (WEEKS et al., 1997; NIELSEN et al., 1999; FRANCOY et al., 2008, 2009; SCHRÖDER et al., 2002). Since the bee wing is easily accessible, the development of new computer-aided methods to species identification was made possible.

Due to its low costs and high efficacy, the morphometry (ROHLF; BOOKSTEIN, 1990) became one of the most used methods for this purpose. Although some systems have been developed to facilitate this task (see Chapter 3), there are still improve-

ments to be done. Most of the researches on this area uses a linear discrimination technique as classifier (FRANCOY et al., 2008; KOCA; KANDEMIR, 2013; FRANCOY; GONÇALVES; JONG, 2012; MEULEMEESTER et al., 2012; MICHEZ et al., 2009), and only a few articles tried to use any other classification method (ROTH et al., 1999), thus, none systematic study assessed what is the best classification method to bee species identification through a comparison between various state-of-art techniques.

Usually the feature extraction process is performed with one of two morphometric techniques, called *Geometric Morphometrics* and *Traditional Morphometrics* (see Sections 2.3.2 and 2.3.1). *Geometric Morphometrics* was reported to get better results in several researches (TOFILSKI, 2008b; KOCA; KANDEMIR, 2013).

There are also some works focused on processing a wing image and extracting features from pixels (DRAUSCHKE et al., 2007; ROTH et al., 1999), where different measures are extracted from images and used for classification, without use of morphometric features.

Researches have achieved good performance both using morphometric data (TOFILSKI, 2008b; FRANCOY et al., 2008; FRANCOY; FRANCO; ROUBIK, 2012) and pixel-based features (ROTH et al., 1999) separately, but there is no work evaluating if those two types of information can be simultaneously used for improving classification performance.

1.1 Objectives

In this work we have two main objectives:

1. Evaluate the feature extraction methods applied to bee species identification. We aim to evaluate if the simultaneous use of morphometric and pixel-based features can improve the classification performance.
2. Define the best classification techniques applied to bee species identification. For this evaluation, both morphometrical and pixel-based features will be used.

1.2 Organization

This text was structured as follows:

- **Chapter 2 – Background:** covers basic concepts needed to fully understand this text.
- **Chapter 3 – Related Researches:** provides a literature review about bee species identification, outlining pros and cons of each approach.

- **Chapter 4 – Proposal Detailing and Experimental Setup:** describes the experiments performed to fulfill our objectives and all decisions regarding the experiments, such as the training set definition, feature extraction and classification algorithms.
- **Chapter 5 – Results and Discussion:** presents the results of our experiments and its discussion.
- **Chapter 6 – Conclusion and Further Work:** concludes this document and points toward possible further researches for this domain.

1.3 Contributions

The most significant contributions of this dissertation will be outlined in this Section.

- **Classification evaluation for Bee Species Identification** - We have chosen seven classifiers among the state of the art methods and defined the most appropriate method to bee species identification through experiments with two training sets. The best classification algorithm was defined both using only Morphometric features and using a conjunction of Morphometric and Pixel-based features.
- **Feature evaluation for Bee Species Identification** - In this work, we empirically showed that the addition of Pixel-based features can improve the classification performance, comparing with classification using only Morphometric features.
- **Feature Selection evaluation for Bee Species Identification** - For each one of the seven evaluated classifiers, we defined a suitable Feature Selection method and evaluated if Feature Selection was able to improve the classification performance.

To the best of our knowledge, none published work achieved any of these contributions prior to this research. Some of the material contained in this dissertation has appeared in the following articles (SANTANA et al., 2014; SILVA; JACOMINI; COSTA, 2014; SILVA et al., 2015 (Under revision)):

1. SANTANA, F. S.; COSTA, A. H. R.; TRUZZI, F. S.; SILVA, F. L.; SANTOS, S. L.; FRANCOY, T. M.; SARAIVA, A. M. A reference process for automating bee species identification based on wing images and digital image processing. *Ecological Informatics*, v. 24, p. 248–260, 2014.

In this article, the bee identification process was fully described, and we carried

some preliminary tests to define if Pixel-based features could improve classification performance.

2. SILVA, F. L.; JACOMINI, R. S.; COSTA, A. H. R. Ensemble learning applied to bee species identification using wing images. In: *II Symposium on Knowledge Discovery, Mining and Learning*, 2014. p. 1–8.

Some preliminary tests were carried to define if Ensemble-based classifiers could perform better than ordinary classifiers to bee species identification.

3. SILVA, F. L.; SELLA, M. L. G.; FRANCOY, T. M.; COSTA, A. H. R. Evaluating classification and feature selection techniques for honeybee subspecies identification by wing images. *Computers and Electronics in Agriculture*, 2015 (Under revision).

We have performed experiments similar to Experiment 3 in this document (Section 5.3), defining the best classification method when using only Morphometric features. We have also evaluated Feature Selection techniques in this article.

2 BACKGROUND

This chapter will describe all concepts needed to fully understand this text. None of this content was developed in this research, nevertheless each one of them were used or referred in some way. In Section 2.1.1 we describe the Bee Species Identification process. In Section 2.2 we explain some image processing concepts that are needed to understand our Feature Extraction process, and some methods used in related works that are described in Chapter 3. We then explain how the Feature Extraction process is usually performed to extract Morphometric features in Section 2.3. The classification Algorithms are described in Section 2.4, whereas we describe Feature Selection methods in Section 2.5. We explain the performance evaluation methods used in this work in Section 2.6 and, finally, Section 2.7 concludes this chapter with the relation between these concepts.

2.1 Background on Bee Identification

"One well-worn, and probably accurate, estimate says that one-third of the human diet can be traced directly, or indirectly, to bee pollination." (DE-LAPLANE; MAYER; MAYER, 2000)

As explained in the Introduction (Chapter 1), the discrimination between bee species requires expertise and time, what is not always available. After the development of softwares that identify some bee species by extracting features from bee wings (Chapter 3), the species identification became faster and efficient. We describe in this section the process to Bee Species Identification using wing images.

2.1.1 Bee Species Identification Process Overview

We will describe here a simplification of the full Bee Species Identification reference process, that is fully described in (SANTANA et al., 2014). Figure 2.1 shows how bee species are identified from a wing image. The first step is to extract features from a training set composed of wing images and their correspondent label (indicating the species of each wing). All images are labeled by an expert, that also marks all *landmarks* when Morphometric features are used. After all features are extracted, an optional Feature Selection method can be executed, where the most irrelevant features will be removed.

All remaining feature values and labels for each image are given to a classifier on the Classifier Training step. After that, a trained classifier is stored to be posteriorly used. When the species identification is required to an unlabeled image, the Feature Extraction step must be performed on the new image, and the feature values are given to the trained classifier. Finally, the classifier will output the predicted species, concluding the Bee Species Identification Process.

Figure 2.1 – Features that represent the wing are extracted on the Feature Extraction step from a training set composed of wing images and their correspondent label (given by an expert). The Feature Selection step is optional, where relevant features will be chosen to be used on the next steps, while irrelevant ones will be discarded. The training set must be given to a classifier on the Classifier Training step. Finally, a trained classifier will be able to identify the species of a new wing image and label the observation.

2.1.2 Landmark

Landmark is a central concept to morphometry. A landmark is a specific point on a biological form located according to some rule (BOOKSTEIN, 1991). Specifically to bee identification, landmarks are vein junctions in the wing (see Figure 2.2). Each landmark is labeled with a number, and, for all bee wing images, landmarks with the same label will correspond to the same location on the wing, disregarding the wing orientation or position on the image.

Figure 2.2 – *Euglossa imperialis* forewing with landmarks (marks in the wing).

2.2 Image Processing

Generally speaking, Image Processing can be defined as any method that receives as input an image and outputs a modified image. Methods for this purpose can emphasize important characteristics on images, that can be further used to find patterns and features on images. An image can be defined as a two-dimensional function $f(x, y)$, where x and y are spatial coordinates, and the amplitude of f on any pair of coordinates is the intensity of the image on that given point (GONZALEZ; WOODS; EDDINS, 2003). Usually grayscale images are represented by functions ranging from 0 (black) to 255 (white), and RGB images are represented by three channels, each one having its own $f(x, y)$ function. This chapter will describe image processing methods that are relevant to this research or to related works.

2.2.1 Image Histogram

Histograms are the basis for numerous image processing techniques, being used for image enhancement, segmentation, etc (GONZALEZ; WOODS, 2002).

A normalized histogram is extracted from an image as:

$$h(r_k) = n_k / n_{image} \quad (2.1)$$

where r_k is the k th gray level, n_k is the number of pixels in the image having intensity r_k , and n_{image} is the total number of pixels in the image.

Histograms are usually visualized as a bar graph, where all the possible values (usually 0 to 255) are represented in the x-axis, and the frequency of pixels that have a given value is represented in the y-axis. Figure 2.3a is a grayscale image and Figure 2.3b is its histogram.

Figure 2.3 – A grayscale image (a) and its histogram (b).

2.2.2 Image Gradient

The magnitude of the gradient of an image is used to detect abrupt changes of intensity on images, being applied to edge identification (GONZALEZ; WOODS, 2002). The gradient for a function $f(x, y)$ can be computed as:

$$\nabla \mathbf{f} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}. \quad (2.2)$$

Usually we use the magnitude of this gradient, expressed as:

$$\nabla f = \text{mag}(\nabla \mathbf{f}) = \sqrt{G_x^2 + G_y^2} \quad (2.3)$$

Note that we need to compute the first derivative to calculate the gradient on Equation (2.2). This derivative can be approximated by convoluting a 3×3 mask (e. g., the *Sobel Operator*) (GONZALEZ; WOODS, 2002).

2.2.3 Thresholding

This method aims to divide the image in 2 labels (usually referred as *object* and *background*). This task is performed by comparing each pixel of the image with a threshold

k (parameter informed by the user). The thresholded image $g(x, y)$ is defined as (GONZALEZ; WOODS, 2002):

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) > k \\ 0 & \text{if } f(x, y) \leq k \end{cases} \quad (2.4)$$

where $f(x, y)$ is the original image and k is the threshold informed by the user.

This method is particularly effective to images with bimodal thresholds, where the desired object has intensity values roughly constant and different from the background. Therefore, this method is ineffective to noisy images and when the image does not have a bimodal histogram.

2.2.4 Morphological Image Processing

Morphological Image Processing is the name assigned to a branch of image processing techniques for extracting image segments that are useful in some way through a set theory language. This group of techniques has this name because it relies on the analysis of shape and form of objects (SOILLE, 2003). None of these techniques were directly employed in our research. However, we will describe operations that were used in related researches (Chapter 3).

We will present here the formulation of methods for binary images, however it is also possible to formulate all of them to grayscale images (GONZALEZ; WOODS, 2002).

A simple way to transform an grayscale image in a binary image is by the thresholding operation (Equation (2.4)).

Only a small portion of *Morphological Image Processing* methods are described in this work, since there is a huge amount of research in this area and only a small number of these techniques have been cited in this work.

2.2.4.1 Notation

For Morphological Image Processing we use the same notation as in set theory. Let A be a set in Z^2 (Here, A represents a binary image), if $a = (a_1, a_2)$ is an element of A , we write:

$$a \in A. \quad (2.5)$$

If a is not an element of A :

$$a \notin A. \quad (2.6)$$

By this notation we mean that, if the pixel on the coordinate (a_1, a_2) has intensity 1, then $a \in A$. However, if this pixel has intensity 0, then $a \notin A$. If every element in A is also present in set B , then we say that A is a subset of B , denoted as:

$$A \subseteq B. \quad (2.7)$$

If we perform the *union* of two sets A and B , the resulting set will have all elements belonging to A or B (or both), denoted as:

$$C = A \cup B. \quad (2.8)$$

The *intersection* is the set of all elements belonging to both A and B :

$$D = A \cap B \quad (2.9)$$

The *complement* is the set of all elements not contained in a given set. For images, this is processed by inverting all intensity values (i.e. 1 becomes 0 and 0 becomes 1).

$$A^c = \{w | w \notin A\}. \quad (2.10)$$

The *difference* of two sets is defined as (GONZALEZ; WOODS, 2002):

$$A - B = A \cap B^c. \quad (2.11)$$

The *reflection* consists in assigning to each coordinate the value of its opposite coordinate. Defined as:

$$\hat{B} = \{w | w = -b, \text{ for } b \in B\}. \quad (2.12)$$

Finally, the translation of a given set by a point $z = (z_1, z_2)$ is defined as:

$$(A)_z = \{c | c = a + z, \text{ for } a \in A\}. \quad (2.13)$$

Some operations rely on the utilization of a set, usually called as *structuring element*. These operations process the image in small subdivisions, where the size and form of these subdivisions are defined by the structuring element.

2.2.4.2 Dilation

The *dilation* of a set A by an structuring element B is defined as:

$$A \oplus B = \{z | [(\hat{B})_z \cap A] \subseteq A\}. \quad (2.14)$$

This operation expands the objects in the image. Intuitively, every pixel with value 0 that is touching a pixel with value 1 will have value 1 after the dilation (SMITH, 1997).

2.2.4.3 Erosion

The erosion is the dual operation of dilation, and is defined by:

$$A \ominus B = \{z | (B)_z \subseteq A\}. \quad (2.15)$$

Erosion can be used to remove irrelevant detail (noise) in an image. Every pixel with value 1 that is touching a pixel with value 0 will have value 0 after the erosion (SMITH, 1997). The Erosion and Dilation are the base to most of complex operations.

2.2.4.4 Hit-or-Miss

This operation is used for shape detection (GONZALEZ; WOODS, 2002). The result of this operation will be an image where each pixel with intensity equals to 1 corresponds to a location where the desired shape was located. Let $B = (B_1, B_2)$, where B_1 is the desired shape and B_2 is the background ($B_2 = B_1^c$), this operation can be processed as:

$$A \circledast B = (A \ominus B_1) \cap (A^c \ominus B_2). \quad (2.16)$$

2.2.4.5 Thinning

This morphological operation is used to reduce the thickness of the selected object on a binary image. The thinning of A by a structuring element B , denoted $A \oslash B$ can be defined as (GONZALEZ; WOODS, 2002):

$$\begin{aligned} A \oslash B &= A - A \circledast B \\ &= A \cap (A \circledast B)^c. \end{aligned} \quad (2.17)$$

Where $A \circledast B$ is the hit-or-miss operation.

2.3 Feature Extraction

As explained in Section 2.4, the classification of unlabeled samples is done by processing feature values. These features, in the bee identification domain, must be extracted from wing images in some way. This Section describes the most used feature extraction methods for bee identification.

2.3.1 Traditional Morphometrics

The first method to extract features from bee forewings, now known as *Traditional Morphometrics*, compares shape variations on wings through angles and distances (KOCA; KANDEMIR, 2013). These angles and lengths are listed in Table 2.1 and Figure 2.4 shows how these features are measured in the wing.

These features could be extracted manually by measuring the wing or with assistance of computer softwares (e.g., tpsDig (ROHLF, 2010a) helps to measure angles).

Table 2.1 – *Traditional Morphometrics* features, extracted from (AYŞEGÜL GÖNÜLŞEN, 2004).

Angle Parameters	Length Parameters
Angle EAB (A4)	Forewing Length, $FL = L1+L2$
Angle EBA (B4)	Forewing Width, FW
Angle BDG (D7)	Length 1, L1
Angle FGD (G18)	Length 2, L2
Angle OKF (K19)	Cubital index a
Angle ROQ (O26)	Cubital index b
Angle HEI (E9)	Distance c
Angle PNJ (N23)	Distance d
Angle NJM (J16)	
Angle IJH (J10)	
Angle ILE (L13)	

More recently, *Traditional Morphometrics* have been reported to get worst results than *Geometric Morphometrics* (TOFILSKI, 2008b; MIGUEL et al., 2011; KOCA; KANDEMIR, 2013), becoming a rarely-used method.

2.3.2 Geometric Morphometrics

Geometric Morphometrics is a collection of approaches to compare shapes of sets of cartesian coordinate data. Usually, landmarks are used to define shapes of bee wings. *Geometric Morphometrics* have been reported to get the best result for bee identification (TOFILSKI, 2008b; KOCA; KANDEMIR, 2013), and is today the main method for feature extraction. The next subsections describe features from *Geometric Morphometrics* that can be extracted from landmarks. All these features can be easily obtained in the tpsRelw software (ROHLF, 2010b).

2.3.2.1 Centroid Size

The Centroid Size (CS) is a widely used feature from Geometric Morphometrics (FRANCOY; GONÇALVES; JONG, 2012; MEULEMEESTER et al., 2012; TOFILSKI, 2008b), and can be calculated as follows (BOOKSTEIN, 1991):

Figure 2.4 – (a): Length Parameters. (b): Angle Parameters (Extracted from (AYŞEGÜL GÖNÜLŞEN, 2004)).

$$CS = \frac{\sqrt{\sum_{i=1}^{|L|} (x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{|L|}, \quad (2.18)$$

where the set of landmarks L is composed of $|L|$ landmark coordinates (x_i, y_i) , $i = 1, 2, \dots, |L|$, and $(\bar{x}, \bar{y})$ is the centroid of L . Usually the centroid size is calculated without the last division by $|L|$ (aiming to save computational time, since it is a constant

value for all wings), however, we have calculated the centroid size with this final step to reduce any bias regarding the dimensional difference between features (in the training phase).

The centroid can be computed as follows:

$$\bar{x} = \frac{1}{|L|} \sum_{i=1}^{|L|} x_i \text{ and } \bar{y} = \frac{1}{|L|} \sum_{i=1}^{|L|} y_i. \quad (2.19)$$

where (x_i, y_i) are the coordinates of the landmark i .

2.3.2.2 Aligned Coordinates

These features are used in all published works of our knowledge (KOCA; KANDEMIR, 2013; FRANCOY et al., 2011, 2008; FRANCOY; GONÇALVES; JONG, 2012; KANDEMIR; OZKAN; FUCHS, 2011; MEULEMEESTER et al., 2012; TOFILSKI, 2008b; FRANCOY; FRANCO; ROUBIK, 2012) and consist in performing rotation, scale and translations transformations on the configurations of landmarks, aiming to obtain configurations of coordinates invariant to these operations. This procedure can be performed in a number of ways, e.g., calculating the Bookstein coordinates (BOOKSTEIN, 1991) or performing an Orthogonal Procrustes Analysis (ROHLF; SLICE, 1990). In this work, we provided the rotation invariance with a procedure based on second moments, whereas the scale and translation invariance were achieved with an Orthogonal Procrustes Analysis, as described in Section 4.2.1.

2.3.2.3 Principal Warps

The Principal Warps (BOOKSTEIN, 1989) is a tool to shape variation analysis, and have been used to extract features to bee species recognition successfully (MEULEMEESTER et al., 2012; FRANCOY; FRANCO; ROUBIK, 2012). The first step is to define a reference object (called consensus object) to be used in the following computations. In this work, the consensus object was defined calculating, to each landmark, the mean value of the (x, y) coordinates of all samples of our entire training set, resulting in $2|L|$ coordinates (for $|L|$ landmark points). Note that the coordinates must be aligned (Section 2.3.2.2) prior to this computation.

The next step is to compute the bending energy matrix ($B_{|L|}^{-1}$) for the consensus object. This can be done assembling the partitioned matrix (ROHLF, 1993):

$$B = \begin{bmatrix} P & Q \\ Q^t & O \end{bmatrix} \quad (2.20)$$

where

$$\mathbf{P} = \begin{bmatrix} 0 & U(r_{12}) & U(r_{13}) & \dots & U(r_{1|L|}) \\ U(r_{21}) & 0 & U(r_{23}) & \dots & U(r_{2|L|}) \\ U(r_{31}) & U(r_{32}) & 0 & \dots & U(r_{3|L|}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U(r_{|L|1}) & U(r_{|L|2}) & U(r_{|L|3}) & \dots & 0 \end{bmatrix} \quad (2.21)$$

where $|L|$ is the number of landmarks and U stands for the function:

$$U(r_{ij}) = r_{ij}^2 \ln r_{ij}^2 \quad (2.22)$$

and r_{ij} is the distance between the landmarks i and j of the consensus object, and:

$$\mathbf{Q} = \begin{bmatrix} 1 & x_{con1} & y_{con1} \\ 1 & x_{con2} & y_{con2} \\ \vdots & \vdots & \vdots \\ 1 & x_{con|L|} & y_{con|L|} \end{bmatrix} \quad (2.23)$$

where (x_{coni}, y_{coni}) are the coordinates of the landmark i of the consensus object. The matrix $\mathbf{O}$ is an 3×3 matrix all zeros.

With $\mathbf{B}$ fully built, we then compute its inverse and extract the bending energy matrix $\mathbf{B}_{|L|}^{-1}$, i.e., the upper-left $|L| \times |L|$ block of the inverse of $\mathbf{B}$. An eigen decomposition (PARLETT, 1998) of the bending energy matrix is performed:

$$\mathbf{B}_{|L|}^{-1} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^t \quad (2.24)$$

where $\mathbf{\Lambda}$ and $\mathbf{E}$ are $|L| \times |L|$ matrices of, respectively, eigenvalues and eigenvectors. Three eigenvalues will be equal to zero, corresponding to the affine components (translation, scale and rotation) (ROHLF, 1993). These eigenvalues are removed from $\mathbf{\Lambda}$ and their respective eigenvectors from $\mathbf{E}$, reducing $\mathbf{\Lambda}$ to a $|L| - 3 \times |L| - 3$ matrix and $\mathbf{E}$ to a $|L| \times |L| - 3$ matrix.

Finally, we can compute the weight matrix $\mathbf{W}$ to be used as feature:

$$\mathbf{W} = [\mathbf{W}_x \mid \mathbf{W}_y], \quad (2.25)$$

where

$$\mathbf{W} = \frac{1}{\sqrt{n}} \mathbf{V} (\mathbf{I}_2 \otimes \mathbf{E} \mathbf{\Lambda}^{-\alpha/2}), \quad (2.26)$$

$$\mathbf{V} = [\mathbf{V}_x \mid \mathbf{V}_y] \quad (2.27)$$

and

$$\mathbf{V}_x = \mathbf{X}_x - \mathbf{1}_n \otimes [1 \mid 0] \mathbf{X}_c, \quad (2.28)$$

$$\mathbf{V}_y = \mathbf{X}_y - \mathbf{1}_n \otimes [0 \mid 1] \mathbf{X}_c. \quad (2.29)$$

The symbol $\otimes$ refers to a direct product of two matrices, n is the number of specimens, $\mathbf{X}_x$ is a matrix with x coordinates of each specimen (each row refers to a specimen, while each column refers to a landmark) and $\mathbf{X}_y$ refers to y coordinates. $\mathbf{1}_n$ is a column of n ones and $\mathbf{X}_c$ is a matrix with the consensus object coordinates.

The parameter α was introduced due to the suggestion of (BOOKSTEIN, 1991). If $\alpha = 1$ large-scale variations have more weight than small-scale variations, and if $\alpha = 0$ all variations have the same weight.

2.3.2.4 Relative Warps

This method has already been used successfully in this domain (FRANCOY; GONÇALVES; JONG, 2012; MEULEMEESTER et al., 2012). A Singular Value Decomposition (GOLUB; KAHAN, 1965) is performed on the weight matrix of equation (2.25):

$$\mathbf{W} = \mathbf{S} \mathbf{D} \mathbf{R}^t \quad (2.30)$$

where $\mathbf{S}$ is an $n \times n$ unitary matrix, $\mathbf{D}$ is an $n \times 2|L|$ diagonal matrix and $\mathbf{R}^t$ is a $2|L| \times 2|L|$ unitary matrix.

We then express the matrix of relative warps in terms of the original x, y -coordinate system (instead of expressing in terms of principal warps) (BOOKSTEIN, 1991):

$$\mathbf{R}' = (\mathbf{I}_2 \otimes \mathbf{E} \Lambda^{-\alpha/2}) \mathbf{R}. \quad (2.31)$$

2.4 Classification

In the context of Machine Learning, classification is the assignment of a label to an observation, i.e., it is a technique used to predict class membership for an unknown observation (RUSSELL; NORVIG, 2003).

In the general case of supervised learning, we want to classify a collection of observations $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_w$ into one of K predefined classes, where w is the number of observations to be classified. The data needed to perform the classification can be organized in an $n \times p$ matrix $\mathbf{X} = (x_{ij})$, where x_{ij} represents the measured value of the feature j in the sample $\mathbf{x}_i$ and n is the number of samples in the training set. Thus, each row of the matrix $\mathbf{X}$ (i.e., each sample of the collection) is associated with a class

label c_i (TARCA et al., 2007). The X matrix and its associated labels form the training set that will be used to train classifiers.

The utilization of this type of learning implies in a classification performed in two phases: The *Training* phase and the *Classification* phase. On the first phase, the classifier will learn how it is possible to classify new observation through analysis of already labeled samples in a training set. On the second phase, the model learnt on the first phase will be used to label new observations.

Specifically to bee species classification, each class label represents a species (or subspecies, depending on the application), and each feature is a value measured somehow from collected specimens (e.g., features shown in Chapter 4.2). Each classifier has an unique procedure to learn how to classify new observations from a training set. The classifiers used in this research will be described on the next subsections.

2.4.1 Linear Discriminant Analysis

Also known as Fisher's Discriminant Analysis, the Linear Discriminant Analysis (LDA) finds a linear combination of features that allows the discrimination of instances belonging to different classes (FISHER, 1936).

This can be done by finding the projective vector W that maximizes the following Fisher Separation Criterion (LEI; LIAO; LI, 2012):

$$J = \frac{|W^T S_b W|}{|W^T S_w W|}. \quad (2.32)$$

The solution of W can be found solving the eigen problem $S_b W = \lambda S_w W$, and:

$$S_b = \frac{1}{n} \sum_{i=1}^K n_i (M^i - M)(M^i - M)^T, \quad (2.33)$$

$$S_w = \frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i} (X_j^i - M^i)(X_j^i - M^i)^T \quad (2.34)$$

where K is the number of classes, n is the total number of samples in the training set, n_i is the number of samples of class i , M^i is the mean vector of class i , M is the mean vector of the whole training set and X_j^i is the vector of feature values of specimen j belonging to class i . These calculations are done in the training phase. In the classification phase, LDA compares the linear combination of measured feature values from an unlabeled observation with the training set, finding the most suitable class and labeling the observation.

Linear separation techniques were used to compare the performance of feature extraction methods from bee wing images (TOFILSKI, 2008b; MIGUEL et al., 2011; KOCA; KANDEMIR, 2013; ROTH et al., 1999).

2.4.2 Naïve Bayes

The Naïve Bayes classifier estimates the probability of new observations be of each class, labeling them according to the class that maximizes this value. This probability is obtained for each class in the classification phase based on the Bayes' theorem, assuming that all features are independent (JOHN; LANGLEY, 1995), resulting in:

$$P(C = c|\mathbf{x}) = P(C = c) \prod_{j=1}^p P(F_j = x_j|C = c), \quad (2.35)$$

where $P(C = c)$ is the probability of the observation belonging to class c , $\mathbf{x}$ stands for a vector of feature values, p is the number of features, and $P(F_j = x_j|C = c)$ is the probability of feature F_j has value x_j given class c .

The probability $P(F_j = x_j|C = c)$ usually is estimated assuming that all features values for a given class have a normal distribution, these distributions are defined for each class in the training phase.

This classifier assumes the independency between features to act in a relaxed problem and, consequently, reduce the computational effort. In practice, Naïve Bayes has already been used in various applications and can compete against more sophisticated classifiers, being successfully applied in several cases, e.g., text classification and medical diagnostics (RISH, 2001).

2.4.3 Logistic

The Logistic classifier builds a ranking of probabilities of an observation being of each possible class, given a feature vector, labeling the observation with the most probable class. Unlike the Naïve Bayes, the Logistic classifier does not assume statistical independency of features.

The probability of an observation $\mathbf{x}$ being of class c is determined in the classification phase by:

$$P(C = c|\mathbf{x}) = \frac{1}{1 + e^{-func(c,\mathbf{x})}}, \quad (2.36)$$

where

$$func(c, \mathbf{x}) = \beta_{0,c} + \sum_{j=1}^p \beta_{jc} x_j, \quad (2.37)$$

where $\beta_{j,c}$ is a regression coefficient associated to feature j and to class c , and x_j is the value of the j -th feature (WITTEN; FRANK; HALL, 2011). Regression coefficients are calculated in the training phase.

2.4.4 K Nearest Neighbors

The K Nearest Neighbors (KNN) classifier is an instance-based learning algorithm, i.e., KNN stores all the training set rather than building a prediction model. Each sample is taken as a point in a cartesian space with p dimensions and the classification phase is performed in three steps:

1. Define the point corresponding to the observation in the same cartesian space of the samples.
2. Define the k nearest neighbors (samples) to the observation.
3. Label the observation with the most frequent class among the defined k neighbors.

In order to perform classification, a metric is needed to calculate the distance between two points, e.g., the Euclidian distance (RUSSELL; NORVIG, 2003). During the training phase, KNN only stores the provided samples and does not perform any calculation. This classifier has been used in various applications, obtaining, for example, good results in parasitic species recognition (SHINN; KAY; SOMMERVILLE, 2000).

2.4.5 C4.5

This classifier builds a decision tree in a training phase for posterior use in the classification of new observations. An one-feature test is chosen to build the decision tree from a training set T , so that y mutually exclusive outputs are defined and T is splitted in y subsets, where T_i has all the instances of the output i . After defining this test, the decision tree will have a vertex identifying the selected test and an edge to each possible output.

After iterating this procedure through all the features, the final decision tree is obtained. The feature order to define the tests is given by the Gain Ratio (GR), calculated by:

$$GR(T, f) = \frac{IG(T, f)}{SI(T, f)}, \quad (2.38)$$

where

$$IG(T, f) = H(T) - \sum_{j=1}^y \left(\frac{|T_j|}{|T|} H(T_j) \right) \text{ and} \quad (2.39)$$

$$H(T) = - \sum_{c=1}^K (freq(c|T) \log_2 freq(c|T)) , \quad (2.40)$$

where

$$SI(T, f) = - \sum_{j=1}^y \left(\frac{|T_j|}{|T|} \log_2 \left(\frac{|T_j|}{|T|} \right) \right) , \quad (2.41)$$

where T is the training set, f is a feature of T , K is the number of classes and IG is the information gain. To calculate the gain, T is partitioned in y subsets (T_j) according to the values of the feature f in T , $freq(c|T)$ stands for the relative frequency of the class c in the set T , and H is the entropy (QUINLAN, 1993). This classifier was successfully applied to assist in the decision-making of pig farming (KIRCHNER; TOLLE; KRIETER, 2004).

2.4.6 Multilayer Perceptron

An Artificial Neural Network (ANN) is a computational model inspired in biological processes of the brain. ANNs are composed by neurons, which are units connected by directed bonds that perform the input processing. ANN can be used in classification problems.

Multilayer Perceptron (MLP) is the most popular type of ANN, where multiple layers of neurons are trained with the backpropagation method (BASHEER; HAJMEER, 2000). For this type of ANN, the signals are propagated from the input to the output layer through a hidden layer (see Figure 2.5), and each neuron in the hidden layer associates an weight to each input. After processing the inputs using a nonlinear function (usually the logistic function, Equation (2.36)) the neuron will pass the processed value to the next layer. The output layer, similarly as the hidden layer, processes the inputs of the hidden layer neurons and outputs the estimated class label. The weights of each neuron is learned during the training phase, and are estimated minimizing some loss function (TARCA et al., 2007).

MLPs are used for classification due to its ability to distinguish non-linear data (BASHEER; HAJMEER, 2000). This classifier has been applied in various situations, showing the best performance to identification of species of the genus *Euglossa* with pixel-based features (SANTANA et al., 2014).

2.4.7 Support Vector Machine

The Support Vector Machine (SVM) finds an optimal hyperplane separating the samples, i.e., the hyperplane that defines the bigger separation margin between different classes. The points defined by the training set that are in the margin (closer to the hyperplane) are called *Support Vectors*.

Figure 2.5 – Representation of a Multilayer Perceptron Neural Network (extracted from (TARCA et al., 2007)). x_1 and x_2 are values of features measured from an unlabeled observation, W_{ij} and α_{ij} are the weights associated with, respectively, bonds of input layer and hidden layer, and bonds of hidden layer and output layer. Each neuron of the input layer will be fed with the value of its respective feature and each neuron of output layer corresponds to a possible class label g . The observation will be labeled with the label corresponding to the activated neuron of output layer.

However, it can be impossible to perform a linear separation between classes; in this case, the feature vector x of p dimensions is transformed into a vector of N dimensions, so that a separation can be found in a higher dimension. A function $\phi : \mathbb{R}^p \rightarrow \mathbb{R}^N$ is chosen to transform the feature vector, enabling to define a hyperplane for non-linear data (CORTES; VAPNIK, 1995). The dot product of two vectors u and v in the space $\mathbb{R}^N$ is called kernel function, which is used to find the relative position of a new observation to the hyperplane, expressed by:

$$\phi(u) \cdot \phi(v) = \text{Kernel}(u, v) \quad (2.42)$$

The hyperplane is defined in the training phase, while the classification phase consists in finding the most suitable class for the new observation according to its relative position. These procedures are performed to binary discrimination; solutions for multiclass classifications are analyzed in (DUAN; KEERTHI, 2005). This classifier was successfully used for bee species identification in (ROTH et al., 1999).

2.5 Feature Selection

Classifiers can be susceptible to irrelevant, redundant or noisy information, hampering its performance. To improve the classification precision and the execution time of the classifiers, it is possible to select features in order to identify the more relevant ones to classification, removing inappropriate features. However, there is no method that performs better in all possible domains (HALL; HOLMES, 2003); therefore, we will describe some methods that have been used in our experiments.

2.5.1 Information Gain

This feature selection method chooses the most relevant features by their Information Gain, that is based on the entropy and can be calculated by the Equation (2.39), as explained in Section 2.4.5.

2.5.2 Chi-Squared

The Chi-Squared tests if the measured features are independent to the output class (MANNING; RAGHAVAN; SCHÜTZE, 2008). This method compares the expected chi-squared distribution, supposing that the feature is independent to the class, with the observed distribution in the training set. The greater the distance between the expected and the obtained distributions, the feature will be more relevant to classification.

2.5.3 Covariance

Based on good results achieved in (ZHANG; WANG; QU, 2008), we slightly modified its feature selection method, where the covariance between features is the criterium to choose the most relevant features. The first step is to determine the matrix of the absolute value of correlation between features for the training set, as follows (ZHANG; WANG; QU, 2008): Let $\mathbf{X}$ be the matrix of calculated values of each features for all the available specimens, and $\mathbf{f}_j = (x_{1j}, x_{2j}, \dots, x_{nj})$ is the data corresponding to feature j . $\mathbf{X}$ will be of size $n \times p$, where n is the number of samples in the training set and p is the number of extracted features. Then,

$$\mathbf{C} = (|\text{corr}_{ij}|)_{p \times p}, \quad (2.43)$$

where $\mathbf{C}$ is the desired correlation matrix and

$$\text{corr}_{ij} = \frac{\text{cov}(\mathbf{f}_i, \mathbf{f}_j)}{\sigma_{f_i} \sigma_{f_j}}, \quad (2.44)$$

$$\text{cov}(\mathbf{f}_i, \mathbf{f}_j) = \frac{1}{n} \sum_{l=1}^n (x_{li} - \bar{\mathbf{f}}_i)(x_{lj} - \bar{\mathbf{f}}_j). \quad (2.45)$$

σ_f refers to the standard deviation of the vector $\mathbf{f}$ and $\text{cov}(\mathbf{f}_i, \mathbf{f}_j)$ to the covariance between the two vectors. Instead of defining a threshold to select a feature subset (as in (ZHANG; WANG; QU, 2008)), we select the maximum value in $\mathbf{C}$ (excluding the elements on the diagonal, that have always the value 1.0) and remove the feature that corresponds to this value. This procedure is iterated until we reach a desired number of features.

2.5.4 Fisher Separation Criterion

The LDA classifier (see Section 2.4.1) needs to find the coefficients that gather samples of the same class while maximizing the margin between different classes; this is done maximizing the Fisher criterion (Equation 2.32). While used to build a classifier, the Fisher Separation Criterion can also be used to perform feature selection, being formulated as (LEI; LIAO; LI, 2012):

$$F^d = \arg \max_F J(F^d) \quad (2.46)$$

where F^d is the feature subset containing d features, F is the complete feature set, and J is the same from Equation (2.32). Directly selecting d features is computationally expensive, but the feature subset can be defined iteratively with the equation (LEI; LIAO; LI, 2012):

$$f_{k+1} = \arg \max_{f \in F/F^k} \{ \min_{f' \in F^k} \Delta J(f|f') - \gamma \max_{f' \in F^k} \text{corr}(f, f') \} \quad (2.47)$$

where F^k is the set of features already selected, F is the complete feature set, $\text{corr}(f, f')$ is the correlation between features f and f' (Equation 2.44), and γ is a trade-off parameter. $\Delta J(f|f')$ is called Fisher Separation Improvement, and is defined as follows:

$$\Delta J(f|f') = J(f, f') - J(f'). \quad (2.48)$$

2.6 Cross-Validation

As explained in Section 2.4, all classifier learns how to perform the classification through a labeled training set. This approach leads to what we call as *overfitting*. That means the classifier has excellent performance to classify the training set, however does not keep the same performance classifying new, previously unknown, observations (RUSSELL; NORVIG, 2003). The Cross-Validation is a method to obtain an estimate of the accuracy rate of a classifier trying to avoid the bias induced by overfitting.

2.6.1 K-fold Cross-Validation

The k-fold Cross-Validation randomly splits the training set in k mutually exclusive subsets (folds), with approximately $\frac{n}{k}$ samples in each fold, where n is the number of samples in the training set. After the training set division, k experiments are performed, where one fold will be used as the test set and the other $k - 1$ will be used as training

set. After the evaluation of all folds, the mean of the experiments is calculated and will be taken as the performance of the classifier on the Cross-Validation.

2.6.2 Stratified Cross-Validation

The Stratified Cross-Validation is a variation of the k-fold Cross-Validation, where the training set is splitted so that the proportion of labels in each of the folds is roughly the same as in the training set (KOHAVI, 1995).

2.6.3 Error Margin

The evaluation of a classifier through a Cross-Validation method has an experimental nature. Thus, any metric extracted by this procedure will be the result of a sampling of the problem (in this case, the training set is a sample of all possible bee wings that exist in the world). In these cases, it is desired to define an error margin to verify if the difference between two classifiers is statistically relevant. Supposing that the accuracy percentage of each classifier is normally distributed, an error margin can be defined as:

$$\mathcal{E} = z_{\rho/2} \frac{\sigma}{\sqrt{s}}, \quad (2.49)$$

where $z_{\rho/2}$ is the critical value for the level of confidence ρ . As we adopted the level of confidence of 95%, $\rho = 0.05$ and $z_{\rho/2} = 1.96$. s is the sample size and σ is the standard deviation of the accuracy rate. As we do not know the standard deviation, it can be defined as follows:

$$\sigma = \sqrt{\frac{1}{s-1} \sum_{i=1}^s (a_i - \bar{a})^2}, \quad (2.50)$$

where a_i is the observed accuracy in an experiment i and $\bar{a}$ is the mean value of all accuracy values extracted from all experiments.

2.6.4 Wilcoxon Signed Rank Test

This method is a non-parametric statistical hypothesis test (WILCOXON, 1945). We have used it in this work to verify if the difference in performance of two classifiers is statistically significant. Given two vectors of values (in our case, accuracies extracted from cross-validation tests), this test will assert if the two vectors are different enough to reject the null hypothesis, i. e., the difference between the two performances is statistically significant. This evaluation is done as follows (the evaluation is done for pairs of classifiers):

Let N be the number of cross-validation tests (experiments), thus, there are a total of $2N$ accuracy values. $v_{1,i}$ and $v_{2,i}$ denote the accuracy values from, respectively, the

first and the second classifier. For $i = 1, \dots, N$, calculate $|v_{2,i} - v_{1,i}|$ and $\text{sgn}(v_{2,i} - v_{1,i})$, where sgn is the sign function. After that procedure, exclude all pairs where $|v_{2,i} - v_{1,i}| = 0$ and order the remaining pairs by the calculated difference in crescent order. Define a Rank R_i to each pair, where the smallest receives $R_1 = 1$ and ties receive the average of the ranks they span. Then, calculate the test statistic W as in Equation (2.51).

$$W = \left| \sum_{i=1}^N \text{sgn}(v_{2,i} - v_{1,i}) R_i \right| \quad (2.51)$$

Finally, when $N \geq 10$, a z-score can be calculated as:

$$z = \frac{W - 0.5}{\sigma_W} \quad (2.52)$$

where:

$$\sigma_W = \sqrt{\frac{N(N+1)(2N+1)}{6}}. \quad (2.53)$$

The difference in the accuracies can be considered statistically significant when $z > z_{\rho/2}$. We chose a confidence level of 95% for this test, thus, $z_{\rho/2} = 1.96$.

2.7 Relation Between Concepts

All the concepts described in this chapter form the background needed to understand the state of art of bee identification and our proposal. Methods for each one of the steps of the Bee Species Identification Process were described in this Chapter. The next chapter will describe the major researches on bee identification, and what can be further improved.

3 RELATED RESEARCHES

After the introduction of wings as source of features for bee identification (WEEKS et al., 1997; NIELSEN et al., 1999; FRANCOY et al., 2008, 2009; SCHRÖDER et al., 2002), much effort have been devoted to perform this task efficiently and precisely. The reason to rely on these procedures is that wings are easily accessible, what even allows to take a photo of a living specimen without killing it. Additionally, the venation of the wing can be easily seen in two-dimensions images, facilitating visual inspection and even image processing of these images. The adoption of wings allowed the development of some automated and semi-automated methods to perform bee identification. With these computer-assisted methods, a non-expert biologist could identify the species of a bee easily and readily. The researches that lead to a major level of automation of this task will be described in this chapter.

3.1 DrawWing

The DrawWing (TOFILSKI, 2004) system achieved a high level of automation on the venation identification in wing images. Using a flatbed scanner it is possible to easily obtain a diagram of wings from bees or others insects as shown in Figure 3.1.

Figure 3.1 – Diagram of a dragonfly (*Aeshna juncea*) forewing generated by *DrawWing*. Adapted from (TOFILSKI, 2004).

The system relies on two threshold values (Section 2.2.3) to build a model of the wing from images, discriminating the wing outline and venation (Although this values should be set manually, DrawWing estimates thresholds that work in most cases). The first step is to determine the wing location, which is done finding the object with maximum outline in the image after applying the threshold. After that, the wing venation is extracted applying the second threshold. The venation outline is transformed in a skeleton through a thinning algorithm (Section 2.2.4.5) and the short-lengthed veins

are removed. These steps will only be successful on images with darker wings presented in an uniform light background (TOFILSKI, 2008a), images that can be easily extracted with a scanner, but harder to get from a microscope.

After detecting the wing's position, three characteristic points are detected: anterior, posterior and apex (as shown in Figure 3.2). The tangents to the wing outline at the anterior and posterior points are parallel to each other and perpendicular to the line crossing them. The anterior and posterior points are automatically detected by examining outline points in pairs until they meet the criteria and the apex point is the extremum at which the outline curvature is smaller. Detecting these points it is possible to crop the wing from the image and perform a rotation to normalize the orientation. Resulting in an image as Figure 3.3.

Figure 3.2 – The wing outline with three characteristic points marked: A: anterior point, B: posterior point and C: apex point. Tangents to the outline at the anterior and posterior points are perpendicular to the line crossing the points. The anterior and posterior points demarcate the wing width, which can be used for wing positioning and scaling. There are two extrema of the outline on opposite sides of the wing width. The apex point is the extremum at which the outline curvature is smaller. Extracted from (TOFILSKI, 2008a).

Figure 3.3 – A standard image produced by *DrawWing*. Extracted from (TOFILSKI, 2008a).

To identify vein junctions (*landmarks*), a gradient of the image (Section 2.2.2) is calculated and the median of gradient values is taken as a threshold. Pixels darker than this threshold is taken as veins and their outline is reduced to a skeleton. Pixels with three or 4 neighbours in the skeleton are considered as vein junctions candidates. These candidates are compared with the expected junctions from a typical honeybee wing, in order to find the real landmarks. This step imposes that the author informs where is the expected position of landmarks for every species to be discriminated.

This is the major drawback of this software, since is impossible to classify species that have not been analyzed by the author. Actually, *DrawWing* performs identification of three species and landmark identification of two genus.

This system also does not work well with microscope images, demanding scanner-generated wing images.

Despite its limitations, *DrawWing* contributed to speed up the identification of bee species, performing this task much faster than marking the landmarks manually and maintaining good results (TOFILSKI, 2008b).

3.2 ABIS

The *Automated Bee Identification System* (ABIS) (ROTH et al., 1999) was the result of an interdisciplinary research project involving zoologists and computer scientists. After marking landmarks on samples in the training set, ABIS will automatically detect landmarks on new observations. One of the first researches to use digital image processing to extract features and automatically mark landmarks (STEINHAGE et al., 2006), this system identify bees with accurate precision and with less effort than manually marking landmarks (to identify new observations)(DRAUSCHKE et al., 2007).

Initially, three cells are extracted (cells B,D1 and D2 of Image 3.4) and, comparing the cells location of the observation with others on the training set, the remaining cells are extracted (ROTH et al., 1999).

Figure 3.4 – Digitalized image of a bee wing with labeled cells (Extracted from (ROTH et al., 1999)).

A number of features including cells area and circumference, veins length of curvature, and distances and angles between landmarks are used to bee identification (the article does not specify all used features). Pixel values from areas of the image that represents the wing are also used. The *ABIS* was one of the first researches to use pixel-based features for bee classification, and the first one to use the SVM classifier (ROTH et al., 1999). The major drawback of this system is that its feature extraction method prevent the identification of stingless bees (FRANCOY; FONSECA, 2010). Having to manually mark the landmarks in the training phase is also a limitation.

3.3 tps

The tps softwares (ROHLF, 2014) are a group of software developed by Rohlf to analyze shape variation, landmark superimposition, etc.

Specifically to bee identification, tpsDig (ROHLF, 2010a) allows to mark landmarks, and tpsRelw (ROHLF, 2010b) processes Aligned Coordinates, Principal Warps and Relative Warps (Section 2.3.2).

These two software are the current standard to extract Geometric Morphometrics features for bee identification, and have been use in various researches to mark landmarks or process Principal Warps and Relative Warps (KOCA; KANDEMIR, 2013; FRANCOY et al., 2011, 2008).

However, since these softwares were not developed exclusively to bee identification, landmarks must be marked manually, which takes time of the specialist to identify even a small amount of wings.

3.4 Unsolved Problems

As described in this Chapter, several softwares have been developed to optimize the bee species identification process. *DrawWing* automatically identifies landmarks on wings. However, its method depends on a definition of the approximate location of the landmarks. This imposition prevents an identification of landmarks that works for all species of bees, and not only to species previously defined by its author. Even though *ABIS* provided a novel level of automation, it is still needed the manual definition of landmarks on samples of the training set and it is impossible to classify stingless bee species.

This lack of flexibility resulted in the utilization of the *tps* software package, that is not specialized to bee identification. This software allows the calculation of morphometric features for any species, however, every landmark must be marked manually and this software does not extract any feature from images. Thus, it is still undefined if the use of pixel-based features along with morphometric features would lead to better results.

Furthermore, almost all literature concerning bee identification uses a linear discrimination method as classifier (In fact, for the best of our knowledge, the SVM classifier used by *ABIS* was the first attempt of using a different classification algorithm). It is still unexplored the effectiveness of various classification methods for bee identification.

In this work we focus on evaluating if new Feature Extraction and Classification techniques can improve the classification performance. While *DrawWing* relies on Morphometric features, the *ABIS* system extracts features based on pixel values. Both

systems achieved good results for Bee Species Identification, now, our work combines features based in Geometric Morphometrics and Pixel values, in order to evaluate if the classification performance is increased by this combination. We also evaluate Feature Selection and Classification techniques that were unexplored in this domain.

The results of our research will unveil the best combination of Feature Extraction, Feature Selection and Classification techniques among all evaluated methods. This novel contribution can help on the development of new softwares for Bee Species Identification in order to deal with gaps until now unsolved by automated systems.

4 PROPOSAL DETAILING AND EXPERIMENTAL SETUP

In this work we aim to answer to the following questions:

1. Would a conjunction of Morphometric and Pixel-based features achieve better performance than only Morphometric features?
2. Could Feature Selection improve classification performance for Bee Species Identification?
3. What are the best classifiers (both when using only Morphometric features and when using Morphometric and Pixel-based features)?

After answering to these question we will have fulfilled the objectives from Section 1.1. Thus, the experiments will focus in answering those questions. All aspects of the experiments will be discussed in this chapter. The main characteristics of our training sets will be discussed in Section 4.1. The methods for Feature Extraction will be described in Section 4.2, while the classification algorithms to be evaluated are described on Section 4.3. Finally, the experiment setup is explained in Section 4.4.

4.1 Training Set Definition

We used two training sets in our experiments, where each sample is a real bee wing properly labeled by an expert. The first one, fully specified in Table 4.1, has 138 samples of 5 species of the *Euglossa* genus. The number of samples per species is roughly the same. These species are similar and difficult to distinguish to non-experts, however, this training set represents a relatively easy identification problem for classifiers, where there are only 5 possible classes.

The second training set (specified in Table 4.2) has 1785 samples of 26 subspecies of *Apis mellifera* bees. Since this training set has more classes of species that are more similar than in the previously training set, this one imposes a much harder classification problem.

These two training sets were built aiming to simulate two classification problems that are possible to occur in a real life situation, with different difficulty level. Both training sets will be used in our experiments.

The expert manually marked landmarks for all samples in both training sets using the tpsDig (ROHLF, 2010a) software. In the first one 18 landmarks were marked

(Figure (2.2)), whereas in the second one 19 landmarks were identified in each image (Figure (4.1)).

Figure 4.1 – Example of *Apis mellifera adamii* forewing with 19 landmarks (squares on the vein junctions). This is a sample from training set 2.

Table 4.1 – Number of samples per species on the first training set, totalizing 138 images.

Species	Images
<i>Euglossa flammea</i>	29
<i>Euglossa ignita</i>	26
<i>Euglossa imperialis</i>	29
<i>Euglossa orellana</i>	26
<i>Euglossa chalybeata</i>	28

Table 4.2 – Number of samples per subspecies of *Apis mellifera* on the second training set, totalizing 1785 images.

Subspecies	Images
<i>adamii</i>	44
<i>adansonii</i>	113
<i>anatoliaca</i>	50
<i>armeniaca</i>	56
<i>capensis</i>	30
<i>carnica</i>	150
<i>caucasica</i>	120
<i>cecropia</i>	88
<i>cypria</i>	40
<i>iberica</i>	20
<i>intermissa</i>	59
<i>jemenitica</i>	123
<i>lamarckii</i>	70
<i>ligustica</i>	109
<i>litorea</i>	54
<i>macedonica</i>	20
<i>major</i>	10
<i>meda</i>	78
<i>mellifera</i>	132
<i>monticola</i>	72
<i>ruttneri</i>	49
<i>sahariensis</i>	20
<i>scutellata</i>	118
<i>sicula</i>	10
<i>syriaca</i>	81
<i>unicolor</i>	69

4.2 Feature Extraction

In order to fulfill the objectives of this research (Section 1.1), we need to extract two types of features from the data described on Section 4.1: (1) Morphometric Features; and (2) Pixel-based features. The following sections describe our procedure to extract those features.

4.2.1 Morphometric Features

Although Geometric Morphometrics have been reported to achieve better results than Traditional Morphometrics methods (TOFILSKI, 2008b; MIGUEL et al., 2011; KOCA; KANDEMIR, 2013), there is no clear consensus about which and how many features

should be used to maximize the classification rate, as the number and type of features vary in different articles.

We extracted all landmark-based features from Geometric Morphometrics that had already been used successfully for bee species identification in the literature (FRANCOY et al., 2008; TOFILSKI, 2008b; SANTANA et al., 2014; KOCA; KANDEMIR, 2013; FRANCOY; GONÇALVES; JONG, 2012; MEULEMEESTER et al., 2012; FRANCOY et al., 2011; KANDEMIR; OZKAN; FUCHS, 2011; FRANCOY; FRANCO; ROUBIK, 2012).

The first feature to be calculated is the Centroid Size (see Section 2.3.2.1), which is a widely used feature from Geometric Morphometrics (FRANCOY; GONÇALVES; JONG, 2012; MEULEMEESTER et al., 2012; TOFILSKI, 2008b). Note that L corresponds to the landmark positions that were marked by the expert in our training sets (Section 4.1).

After Centroid Size computation, the next extracted features are the Aligned Coordinates (Section 2.3.2.2). To provide rotation invariance, we took the landmarks shown in Figure 4.2 as vertexes of a regular polygon, from which an image was generated and any position inside the polygon was set to white and all the others to black (Figure 4.3a).

Figure 4.2 – Landmarks chosen to generate a regular polygon (numbered). The numbers correspond to the order that the landmarks were taken to build the polygon.

Then, we computed the second moments of the polygon image as follows (GRIMSON, 1990):

$$\begin{aligned} m_{2,0} &= \sum_x \sum_y x'^2 b(x, y), \\ m_{1,1} &= \sum_x \sum_y x' y' b(x, y), \\ m_{0,2} &= \sum_x \sum_y y'^2 b(x, y), \end{aligned} \tag{4.1}$$

where $b(x, y)$ is the intensity of pixel (x, y) (white color corresponds to intensity 1, while black color corresponds to intensity 0), and x' and y' are:

$$\begin{aligned} x' &= x - x_c, \\ y' &= y - y_c, \end{aligned} \tag{4.2}$$

where (x_c, y_c) is the position of the center of mass of the image:

$$x_c = \frac{\sum_x \sum_y x b(x, y)}{\sum_x \sum_y b(x, y)} \quad \text{and} \quad y_c = \frac{\sum_x \sum_y y b(x, y)}{\sum_x \sum_y b(x, y)}. \quad (4.3)$$

We can use the moments in Equation (4.1) to determine the angle of rotation θ of the wing (in radians) (GRIMSON, 1990):

$$2\theta = \arcsin \left(\frac{2m_{1,1}}{\sqrt{4m_{1,1}^2 + (m_{2,0} - m_{0,2})^2}} \right). \quad (4.4)$$

After rotating all the configurations of landmarks by their correspondent angle θ around their centroid ($\bar{x}$ and $\bar{y}$ of Equation (2.19)), we obtained the desired rotation invariance (Figure 4.3b shows the polygon after the rotation).

Figure 4.3 – (a): Polygon generated from Figure 4.2. (b): Polygon from image (a) after rotation.

After achieving the rotation invariance, we performed an Orthogonal Procrustes Analysis (ROHLF; SLICE, 1990) in all the landmark configurations, achieving the scale and translation invariance as follows:

$$\mathbf{L}' = \frac{(\mathbf{I} - \mathbf{N})\mathbf{L}}{s}, \quad (4.5)$$

where

$$s = \sqrt{\text{tr}((\mathbf{I} - \mathbf{N})\mathbf{L}\mathbf{L}^t(\mathbf{I} - \mathbf{N}))} \quad (4.6)$$

and $\mathbf{L}$ is the matrix with the (x, y) coordinates of all the landmarks after rotation, $\mathbf{I}$ is an $|L| \times |L|$ identity matrix, $\mathbf{N}$ is an $|L| \times |L|$ matrix with all elements equal to $\frac{1}{|L|}$ and $\text{tr}(A)$ refers to the sum of all the principal diagonal elements of A .

The next features to be calculated are the Principal Warps (BOOKSTEIN, 1989), a tool to shape variation analysis that has been used to extract features for bee species recognition successfully (MEULEMEESTER et al., 2012; FRANCOY; FRANCO; ROUBIK, 2012). This feature is calculated as in Section 2.3.2.3, and the weight matrix is used as feature (Equation 2.25).

The last morphometric feature is the Relative Warps Score (BOOKSTEIN, 1991), that is calculated as indicated on Section 2.3.2.4.

4.2.2 Pixel-based Features

After calculating the angle of rotation in Equation (4.4), the entire image was rotated in the same way as the landmarks. This same procedure is repeated for all images on the training set, thus ensuring that all images are in the same rotation after this step.

After this procedure, the minimal bounding box that holds all the landmarks was calculated with a 5 pixel margin, then the image was cropped using the bounding box as a mask. The resulting image was divided into 256 quadrants, as illustrated in Figure 4.4, and the following values were extracted for each quadrant to be used as features:

Figure 4.4 – Quadrants and landmarks.

$$m_i = \overline{N_i} \quad \text{and} \quad (4.7)$$

$$std_i = \sigma_{N_i} \quad (4.8)$$

Where

$$N = I - \overline{I}. \quad (4.9)$$

I and $\overline{I}$ are, respectively, the matrix of values of one channel of an image and the mean of all those values, i.e., N is a matrix of deviations from the mean value of a given image channel. N_i is the matrix of all pixels inside quadrant i . σ_{N_i} is the standard deviation and $\overline{N_i}$ is the mean value of those pixels.

Values from Equation (4.7) and Equation (4.8) must be calculated for all three channels of the image (Red, Green and Blue).

4.2.3 Resulting Feature Vector

We want to evaluate if the addition of pixel-based features can improve classification performance, therefore, we have built two data sets for each training set described

in Chapter 4.1, the first one with only morphometric features (Section 4.2.1), and the second one with morphometric and pixel-based features (Section 4.2.2).

The first feature vector consists of the (x_i, y_i) position of each aligned landmark from Equation (4.5), the centroid size CS from Equation (2.18), each element of the weight matrix from Equation (2.25) and the relative warps scores (extracted from R' of Equation (2.31)). We have a total of 97 features for training set 1, and 103 features for training set 2.

$$\begin{aligned} FeatureVector = [x_1, y_1, x_2, y_2, \dots, x_{|L|}, y_{|L|}, CS, wx_1, wx_2, \dots, wx_{|L|-3}, \\ wy_1, wy_2, \dots, wy_{|L|-3}, rwx_1, rwx_2, \dots, rwx_{|L|-3}, rwy_1, rwy_2, \dots, rwy_{|L|-3}]. \end{aligned}$$

The second feature vector includes pixel-based features, i.e., has all features from the previous one, and values from Equation (4.7) and Equation (4.8) (for all three channels). Totalizing 1633 features for training set 1 and 1639 features for training set 2.

$$\begin{aligned} FeatureVector = [x_1, y_1, x_2, y_2, \dots, x_{|L|}, y_{|L|}, CS, wx_1, wx_2, \dots, wx_{|L|-3}, \\ wy_1, wy_2, \dots, wy_{|L|-3}, rwx_1, rwx_2, \dots, rwx_{|L|-3}, rwy_1, rwy_2, \dots, rwy_{|L|-3}, \\ mRed_1, mRed_2, \dots, mRed_{256}, stdRed_1, stdRed_2, \dots, stdRed_{256}, mGreen_1, \\ mGreen_2, \dots, mGreen_{256}, stdGreen_1, stdGreen_2, \dots, stdGreen_{256}, mBlue_1, \\ mBlue_2, \dots, mBlue_{256}, stdBlue_1, stdBlue_2, \dots, stdBlue_{256}]. \end{aligned}$$

4.3 Classification Algorithms

In order to choose the best classification algorithm, we have chosen 7 classifiers that have been successfully used to bee species identification or similar domains. As stated in Section 2.5, classifiers can be hampered by irrelevant or noisy features. Therefore, we have evaluated all classifiers both with and without Feature Selection.

Finding the most appropriate Feature Selector for each classifier is not a trivial task, since each feature selection algorithm can be independent of the classifier and there is no algorithm that perform better in all situations. Also, performing an empirical evaluation of all combinations of algorithms is infeasible, since the number of combinations is too high to perform extensive experiments for all algorithms.

Therefore, we have defined a Feature Selection method for each one of the classifiers from Section 2.4 through a analysis of the literature. The most suitable pairs of feature selection techniques to be employed in our experiments are:

- **LDA and Fisher Separation Criterion:** While LDA can be used with any feature selection method cited in Section 2.5, the Fisher Separation Criterion was developed to be consistent with LDA, since both are based on the Fisher Criterion (LEI; LIAO; LI, 2012). Furthermore, this feature selection technique achieved a higher improvement of classification rate than other techniques for the LDA classifier (LEI; LIAO; LI, 2012).
- **Naïve Bayes and Correlation Feature Selection:** The Naïve Bayes classifier relies on the statistical independence of the chosen features (see Section 2.4.2). Because we cannot ensure the independence of the used features, we adopted the Correlation feature selector to work together with the Naïve Bayes. The correlation value for independent variables is 0. This will assure that highly correlated features will be removed from the feature vector before independent features are removed.
- **C4.5:** As explained in Section 2.4.5, the C4.5 classifier internally selects the best features through the Gain Ratio. Thus, no feature selection was applied to C4.5, which was only trained with the full dataset.
- **Logistic and Chi-Squared:** The performance of the Logistic classifier was evaluated with several feature selection techniques (SILVA et al., 2013), and the Chi-Squared achieved the best results; thus, we applied the same technique to our experiments.
- **MLP and Chi-Squared:** Both the Chi-Squared and Information Gain feature selectors achieved the same result in an evaluation of feature selectors with the MLP classifier (SILVA et al., 2013). Both of them could be used, and we adopted the Chi-Squared feature selector to work with the MLP.
- **KNN and Chi-Squared:** In a study comparing the effectiveness of feature selection techniques for various classifiers (ROGATI; YANG, 2002), the Chi-Squared feature selector performed better for the KNN classifier; thus, we used the same pair of feature selector and classifier.
- **SVM and Information Gain:** The information gain was reported to obtain better results for the SVM classifier in most of the experiments in a comparative study (ROGATI; YANG, 2002).

Some of these classifiers rely on manually set parameters. In this section we will discuss the chosen parameters, these combinations of parameters were empirically chosen through preliminary tests.

1. **Logistic:** The maximum number of iterations was set to 200. In our experiments this configuration led to the same accuracy as configuring the classifier without iteration limit, but with considerable better training time.
2. **KNN:** The adopted distance measure between neighbours was the Euclidian distance. The parameter k was set to 7 for experiments using Training Set 1 and 9 to experiments using Training Set 2.
3. **C4.5:** The used confidence factor was 0.25 when using Training Set 1 and 0.1 when using Training Set 2. The minimum number of instances per leaf was set to 3 in all experiments.
4. **MLP:** The MLP was trained with learning rate of 0.3 for Training Set 1 and 0.1 for Training Set 2. The number of neurons of the hidden layers was set to the sum of the number of classes and attributes divided by two. Empirically, those parameters achieved a good performance, however, other techniques may be used to determine the number of neurons in the hidden layer (STATHAKIS, 2009). The training was conducted in 100 epochs.
5. **SVM:** The SVM algorithm was evaluated with a polynomial kernel, an efficient kernel for multiclass classification (SANGEETHA; KALPANA, 2011).

4.4 Experimental Setup

In order to define the best classification algorithm for each situation, we have executed four independent experiments.

The first two experiments aim to define the best classifier for the first training set (Section 4.1), both using only Morphometric features and using a conjunction of morphometric and pixel-based features. The last experiments will perform the same procedure for the second training set. The results of these experiments will be the definition of the best combination of classifier and feature selector for training sets with different difficulty levels, which means that the appropriate algorithm can be chosen according to the similarity of the problem to be solve with our training sets.

Also, the comparison between the accuracies of the best algorithms in these four experiments will allow us to define if the inclusion of pixel-based features can improve the classification performance. Therefore, the results of our experiments will fulfill our objectives, since the best classification algorithm for both morphometric and pixel-based features will be defined.

The classification algorithms are evaluated in conjunction with a feature selection algorithm, as described in Chapter 4.3.

Each experiment was carried as illustrated in Figure 4.5. The first step is to extract features from the training set, after defining feature values for all samples, the following steps are executed for each classifier. A 10-fold Stratified Cross-Validation (Section 2.6.2) is performed without feature selection, then the confusion matrix is stored, a Feature Selection is performed removing approximately 5% of the total number of features and the loop goes back to the Cross-Validation.

Figure 4.5 – Illustration of the process to execute each experiment. This whole process is repeated for each classifier

After it is no longer possible to perform Feature Selection, the seed for the random split on the Cross-Validation is changed, all features are restored to the training set, and the experiment is carried again. This process is repeated 10 times (with 10 different seeds). After finishing all experiments, it is possible to evaluate the accuracy of classifiers with and without Feature Selection for all training sets. We extracted the correct classification percentage and we will present the mean accuracy percentage and the error margin defined by Standard Error (Section 2.6.3) for each experiment. The Wilcoxon Signed Rank Test (Section 2.6.4) was used to verify if differences between mean accuracies are statistically significant.

All the experiments were executed with MATLAB (MATLAB, 2012). The WEKA API (HALL et al., 2009) was used for classifiers *Naïve Bayes*, *Logistic*, *MLP*, *KNN*, and *SVM* and for feature selectors *Chi-Squared* and *Information Gain*. The MATLAB implementation of *LDA* was used, and we implemented the *Correlation* and *Fisher Separation* methods on MATLAB.

The experiments were executed in a virtual machine hosted on the cloud services provided by University of Sao Paulo (USP Cloud), with an 8-core CPU of 2.3 GHz and 16GB of RAM.

5 RESULTS AND DISCUSSION

The results of each experiment will be presented and discussed in this chapter.

5.1 Experiment 1 - *Euglossa* training set using Morphometric Features

Our first experiment aims to define the best classification method for bee species identification using only morphometric features. Figure 5.1 presents the mean accuracies of each classifier in 10 executions with different Cross-Validation fold partitions. It is noteworthy that Feature Selection was able to increase the accuracy of almost all algorithms that were employing Feature Selection (actually, only Naïve Bayes was not benefited). It is noteworthy that C4.5 performance is represented by one point in all figures, since we did not employ other Feature Selection algorithm than the algorithm internal selection.

After Feature Selection, MLP was able to increase its accuracy from 71.81% (no Feature Selection) to 82.54% (after removing 88% of the original feature set). SVM improved its performance from 70.87% to 78.84% (after removing 84% of the feature set) and KNN increased its performance from 68.26% to 78.69% (removing 80% of the feature set). The highest improvement was for the LDA, which raised its accuracy from 49.49% to 74.71%. Those results show that Feature Selection can greatly improve the classification performance, and precisely LDA, that is the most used classification method in this domain, was greatly benefited by the remotion of irrelevant features. Finally, C4.5 built a tree considering 11 features (i.e., removed 88.18% of the feature set), achieving an accuracy of 61.52%.

Besides improving the classification performance, Feature Selection was also able to decrease the computation time needed for the training and classification for all classifiers. Figure 5.2 shows the mean time that each classifier took to classify one unlabeled instance. Note that the computation time decreases along with the number of features, as expected. This behavior was also observed with the training time, Figure 5.3 shows the mean elapsed time to perform one training and confirms the progressive reduction of computation time after Feature Selection.

Figure 5.4 shows the best accuracy achieved by each classifier both with and without Feature Selection. According to the Wilcoxon Signed Rank Test (the statistical results for this experiment is shown in Figure 5.5), the difference between the performance of MLP and Naïve Bayes is statistically significant. Therefore, for our first

Figure 5.1 – Mean of the correctly classified instances of 10 executions of the experiment 1.

Figure 5.2 – Mean of the Classification Time for each classifier on experiment 1.

Figure 5.3 – Mean of the time elapsed on one training for each classifier on experiment 1.

training set, the best classification method is the MLP, with accuracy of 82.54%. It is surprising that the best classification algorithms for this experiment were MLP and Naïve Bayes, since neither of them were previously used in this domain.

Figure 5.4 – Best result achieved by each classifier in experiment 1, the error margin is shown in yellow.

	C4.5	MLP	SVM	LDA	Naive Bayes	Logistic
KNN						
	C4.5					
		MLP				
			SVM			
				LDA		
					Naive Bayes	

Figure 5.5 – Wilcoxon Signed Rank Test results for experiment 1. Green cells mean that the difference between the best results for each classifier is statistically significant. Red cells mean that it is not possible to reject the null-hypothesis with a confidence level of 95%.

5.2 Experiment 2 - Euglossa training set using Morphometric and Pixel-based features

This experiment will allow us to directly compare the classification performance using only Morphometric features with classification using a conjunction of Morphometric and Pixel-based features.

Figure 5.6 shows that, in this experiment, every single classifier was benefited by the use of Feature Selection. The KNN classifier had a performance of 68.55% without Feature Selection and achieved an accuracy of 77.83% after the removal of 89% of the feature set. In turn, MLP increased its performance from 87.32% to 90.58% after removing 79% of the feature set. SVM improved its accuracy from 85.29% to 91.74% after removing 69% of the original features, and LDA raised its performance from 88.84% to 89.64%, after removing 30% of the feature set. Naïve Bayes achieved accuracy of 67.25% without Feature Selection and accuracy of 73.48% after removing 94% of the Feature Set. Finally, Logistic increased its performance from 85.29% to 91.45%, with 74% of the feature set removed. The tree of C4.5 algorithm was build with 11 features, what represents the remotion of 99.33% of the feature set.

This is the first experiment where the Naïve Bayes classifier was benefited by Feature Selection, what was an expected behavior, since we have included pixel values and standard deviations that are noisy and we cannot assure to be statistically independent. It is noteworthy that this algorithm, one of the top algorithms on the first experiments, was surpassed by almost all other algorithms. This is also explained by the introduction of noisy and possibly dependent features, that greatly hampers Naïve-Bayes accuracy. After the removal of most of irrelevant features, Naïve Bayes achieved its best results. LDA had an steep decrease of accuracy after the remotion of 74% of the feature set, indicating that after that point, much of the removed features are relevant, and Feature Selection would hamper the classification.

As in the other experiments, Feature Selection was able to decrease both classification and training times for all classifiers, as shown in Figures 5.7 and 5.8.

Figure 5.6 – Mean of the correctly classified instances of 10 executions of the experiment 2.

Figure 5.7 – Mean of the Classification Time for each classifier on experiment 2.

Figure 5.8 – Mean of the time elapsed on one training for each classifier on experiment 2.

Figure 5.9 shows that the SVM classifier achieved the best mean accuracy value. However, the difference between the accuracy from SVM and Logistic is not statistically significant (Figure 5.10). Therefore, The best classifiers in our third experiment are SVM and Logistic, with accuracies of 91.74% and 91.45%, respectively. The MLP classifier, the best one in the experiment of this same training set with Morphometric Features (Experiment 1), remained one of the best options, with 90.58% of mean accuracy. However MLP was surpassed by the SVM algorithm. It is noteworthy that the inclusion of pixel-based features enabled an improvement in mean accuracy from 82.54% (MLP on Experiment 1) to 91.74% (SVM in this experiment).

Figure 5.9 – Best result achieved by each classifier in experiment 2, the error margin is shown in yellow.

	C4.5	MLP	SVM	LDA	Naive Bayes	Logistic
KNN						
	C4.5					
		MLP				
			SVM			
				LDA		
					Naive Bayes	

Figure 5.10 – Wilcoxon Signed Rank Test results for experiment 2. Green cells mean that the difference between the best results for each classifier is statistically significant. Red cells mean that it is not possible to reject the null-hypothesis with a confidence level of 95%.

5.3 Experiment 3 - Apis training set using Morphometric Features

This experiment will define the best classifier for a harder classification problem, with 26 possible classes with an unequal number of samples per class.

Figure 5.11 shows the mean accuracy achieved by each classifier on this experiment. Unlike the first experiment, all classifiers achieved less than 65% of accuracy, which demonstrates that this problem is really difficult to be solved through only morphometric features, and a different feature set is required to achieve better results. The accuracy of 37.14% of the C4.5 shows that this algorithm was greatly hampered by the noise in feature values, the unequal number of samples per class and the high number of classes. Similarly to the first experiment, only Naïve Bayes was not benefited by Feature Selection. The KNN classifier raised its performance from 49.89% to 51.27% (after removing 32% of the feature set), MLP, in turn, increased the accuracy from 57.13% to 57.77% (with removal of 28% of the feature set). SVM could improve its

performance from 59.50% to 60.09% (28% of feature set reduction), while LDA raised its accuracy from 59.07% to 62.93% (after removing 51% of the feature set). Finally, the Logistic classifier had an increase from 60.58% to 63.47%. C4.5 built its tree with 87 features, removing 15.53% of the feature set.

Figure 5.11 – Mean of the correctly classified instances of 10 executions of the experiment 3.

The improvements in accuracy were much lesser than in the first experiment, indicating that Feature Selection is not enough to raise the performance to a good level, since only morphometric features were not enough to adequately represent all classes. Nevertheless, besides a small contribution in performance, Feature Selection maintained its property of reducing training and classification times for all classifiers, as shown in Figures 5.12 and 5.13.

Figure 5.14 shows the best accuracy for each classifier. The MLP, the best algorithm on the first experiment, was greatly hampered by this training set with a greater number of classes, but Naïve Bayes maintained a good performance, with the second greatest accuracy mean. Although Logistic had achieved the best mean accuracy, the difference between the mean performance of Logistic and Naïve Bayes is not statistically significant (Figure 5.15). Thus, the best classifiers in our third experiment are the Naïve Bayes and Logistic classifiers.

Figure 5.12 – Mean of the Classification Time for each classifier on experiment 3.

Figure 5.13 – Mean of the time elapsed on one training for each classifier on experiment 3.

Figure 5.14 – Best result achieved by each classifier in experiment 3, the error margin is shown in yellow.

	C4.5	MLP	SVM	LDA	Naive Bayes	Logistic
KNN						
	C4.5					
		MLP				
			SVM			
				LDA		
					Naive Bayes	

Figure 5.15 – Wilcoxon Signed Rank Test results for experiment 3. Green cells mean that the difference between the best results for each classifier is statistically significant. Red cells mean that it is not possible to reject the null-hypothesis with a confidence level of 95%.

5.4 Experiment 4 - Apis training set using Morphometric and Pixel-based features

Our last experiment aims to evaluate the classifiers with our second training set after including pixel-based features on the training set.

Figure 5.16 shows that, unlike on the other experiments, Feature Selection was not able to improve the accuracy of most classifiers. This is an unexpected outcome, since our pixel-based features are expected to be noisy, and Feature Selection should benefit the classification performance. The difficulty of this training set hampered the capacity of defining how to classify among the 26 classes with a reduced feature set, specially for classifiers that can deal with a high number of features, as SVM and Logistic. Two classifiers were strongly benefited by Feature Selection, though. LDA had a terrible performance without Feature Selection (10.87%), and was able to reach a performance of 77.98% after removing 85% of the feature set. In turn, Naïve Bayes improved its

performance from 38.91% without Feature Selection to 61.44%, after the removal of 95% of the feature set. Finally, C4.5 built a tree with 167 features, removing 89.81% of the feature set.

Figure 5.16 – Mean of the correctly classified instances of 10 executions of the experiment 4.

As in all other experiments, Feature Selection reduced the training and classification times for all classifiers (Figures 5.17 and 5.18)

However hard the classification problem imposed by this training set, some classifiers achieved good results after the inclusion of pixel-based features. As shown in Figure 5.19 the Logistic classifier achieved the best result, with accuracy of 85.25%. As shown in Figure 5.20, all differences between accuracy means are statistically significant. Comparing with results from Experiment 3, the addition of pixel-based features allowed an improvement in accuracy from 63.47% (Logistic classifier on Experiment 3) to 85.25% (Logistic in this experiment).

Figure 5.17 – Mean of the Classification Time for each classifier on experiment 4.

Figure 5.18 – Mean of the time elapsed on one training for each classifier on experiment 4.

Figure 5.19 – Best result achieved by each classifier in experiment 4, the error margin is shown in yellow.

	C4.5	MLP	SVM	LDA	Naive Bayes	Logistic
KNN						
	C4.5					
		MLP				
			SVM			
				LDA		
					Naive Bayes	

Figure 5.20 – Wilcoxon Signed Rank Test results for experiment 4. Green cells mean that the difference between the best results for each classifier is statistically significant. Red cells mean that it is not possible to reject the null-hypothesis with a confidence level of 95%.

5.5 Concluding Remarks

Our four experiments showed that the inclusion of pixel-based features improves the classification performance both for easy (training set 1) and hard (training set 2) classification problems within the bee species identification domain. Table 5.1 shows the best classification algorithm for each experiment, where the accuracy achieved when using pixel-based features is significantly better than only using morphometric features.

Using only morphometric features, the MLP classifier achieved the best accuracy in the first training set, while the second best was Naïve Bayes with a slightly worst accuracy. For the second training set, Logistic and Naïve Bayes were the best classifiers, while MLP achieved a poorer performance. Judging by the great performance of Naïve Bayes in both training sets, which was one of the two best classifiers for both training sets, with low differences comparing with the first one, the best classifier for bee species identification using only morphometric features is the Naïve Bayes classifier.

After the inclusion of pixel-based features, the overall accuracy increased for both training sets. The best classification algorithms for the first training set were SVM and Logistic, while the Logistic classifier was the best one for the second training set. Since the Logistic classifier was the best one on the first experiment, and the difference between SVM and Logistic was not statistically relevant on the first experiment, the best classification algorithm for bee species identification using a conjunction of Morphometric and Pixel-based features is the Logistic classifier.

Table 5.1 – Best classifier for each experiment. When the difference between two accuracies was not statistically significant, both classifiers were added to the table. *Morphometric* stands for the experiments using only Morphometric Features and *Pixel* stands for experiments with both Morphometric and Pixel-based Features. *Log.* is the Logistic classifier, and *N. Bayes* is the Naïve Bayes classifier

	Morphometric		Pixel	
	Best Classifier	Accuracy	Best Classifier	Accuracy
Training Set 1	MLP	82.54%	SVM and Log.	91.74%, 91.45%
Training Set 2	Best Classifier	Accuracy	Best Classifier	Accuracy
	Log. and N. Bayes	63.47%, 62.96%	Logistic	85.25%

6 CONCLUSION AND FURTHER WORK

Our research evaluated the use of seven different classification algorithms in the bee species identification domain. The classifiers were chosen from the state-of-art algorithms that had already been successfully used in our domain or a similar one (Section 2.4). We also chose Feature Selection algorithms that could be used together with each classifier, as explained in Section 4.3. Four experiments were carried in order to define the best classifier for bee species identification both using only Morphometric Features and using a conjunction of Morphometric and Pixel-based features.

For the best of our knowledge, this is the first work that evaluated whether the joint use of Morphometric and Pixel-based features could lead to better results than only using Morphometric Features, since previous researches use one of these two feature sets, and not both together.

The definition of the best classifier to bee species identification is also a novel contribution, since most of the researches use Linear Discrimination methods, and the ones who use other classification algorithms did not evaluate various classifiers. Our research also evaluated the benefits of Feature Selection, that had not been considered in most of previous works.

Two training sets with real wing images were provided by an expert for our experiments (Section 4.1). Both training sets were built trying to simulate real life classification problems, with different difficulty levels.

We defined that the best classifier using only Morphometric features is the Naïve Bayes classifier. This result can be useful to biologists or other researchers that will use softwares that calculate only Morphometric features (e.g. tpsRelw (ROHLF, 2010b)), while there is no software that uses both morphometric and pixel-based features. In turn, the Logistic classifier together with the Chi-Squared Feature Selector was the best classifier using Morphometric and Pixel-based features. This outcome can guide new researches trying to improve the bee species identification process and must be considered for the development of new softwares for this domain.

It was also defined that the use of a conjunction of Morphometric and Pixel-based features lead to better results than only using Morphometric features. We were able to significantly improve the classification performance in difficult classification problems after the addition of Pixel-based features (Experiment 4).

Further researches can focus on automating the definition of landmark points. Tofilski proved that it is possible to automatically detect landmark points (Section 3.1), however, his method must be improved to work for most bee species.

The Pixel-based features used in this research could also be improved for more sophisticated features. Our features are only based in pixel values, that are not invariant to changes and illumination and other sources of noise. For our research all images were taken from the same equipment using the same level of illumination, but in different situations this is a noise source that must be considered. The ABIS system (Section 3.2) extracted various features from pixels, but it is not clear in their articles all the used features and how to extract them.

Some unexpected results of our experiments point toward another research opportunity. Theoretical studies must be done to investigate the effect of different parameter definitions for both Classification and Feature Selection techniques, in order to avoid situations where the Feature Selection is not able to improve classification performances, such as observed in our experiment 4 (Section 5.4).

Finally, the development of a new software using the concepts discussed in this work is crucial. Since many researchers in this area are biologists without programming skills, the development of a new integrated system to easily apply Morphometric and Pixel-based feature extraction, Logistic classifier and Chi-Squared Feature Selection would contribute to spread these techniques to all researchers interested in bee species identification. A mobile application would be specially useful for this domain, since the possibility of identifying bee species with a regular cellphone would help biologists in field researches. However, limitations imposed by these devices must be taken in account, such as memory and processing limitations. Thus, further studies must be done to define if our current mobile technology allows the mobile implementation of the methods discussed in this dissertation.

REFERENCES

- AYŞEGÜL GÖNÜLŞEN. *Feature Extraction of Honeybee forewings and hindlegs using image processing and active contours*. Tese (Doutorado) — THE MIDDLE EAST TECHNICAL UNIVERSITY, Turkey, 2004.
- BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, v. 43, p. 3–31, 2000.
- BOOKSTEIN, F. L. Principal warps: thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 11, n. 6, p. 567–585, Jun 1989.
- BOOKSTEIN, F. L. *Morphometric tools for landmark data: geometry and biology*. Cambridge, USA: Cambridge University Press, 1991.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, set. 1995.
- COSTANZA, R.; d'Arge, R.; Groot, R.; FARBER, S.; GRASSO, M.; HANNON, B.; LIMBURG, K.; NAEEM, S.; ONEILL, R. V.; PARUELO, J.; RASKIN, R. G.; SUTTON, P.; Belt, M. The value of the world's ecosystem services and natural capital. *NATURE*, v. 387, n. 6630, p. 253–260, maio 1997.
- DALY, H. V.; BALLING, S. S. Identification of africanized honeybees in the western hemisphere by discriminant analysis. *Journal of the Kansas Entomological Society*, v. 51, n. 4, p. 857–869, 1978.
- DELAPLANE, K.; MAYER, D.; MAYER, D. *Crop Pollination by Bees*. [S.l.]: CABI Publishing, 2000.
- DRAUSCHKE, M.; STEINHAGE, V.; VEGA, A. P. de la; MÜLLER, S.; FRANCOY, T. M.; WITTMANN, D. Reliable biometrical analysis in biodiversity information systems. In: FRED, A. L. N.; JAIN, A. K. (Ed.). *PRIS*. [S.l.]: INSTICC PRESS, 2007. p. 27–36.
- DUAN, K.; KEERTHI, S. S. Which is the best multiclass SVM method? An empirical study. *Proceedings of the Sixth International Workshop on Multiple Classifier Systems*, p. 278–285, 2005.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, n. 7, p. 179–188, 1936.
- FRANCOY, T.; FONSECA, V. I. A morfometria geométrica de asas e a identificação automática de espécies de abelhas. *Oecologia Australis*, v. 14, n. 1, p. 317–321, 2010.
- FRANCOY, T.; FRANCO, F. F.; ROUBIK, D. Integrated landmark and outline-based morphometric methods efficiently distinguish species of euglossa (hymenoptera, apidae, euglossini). *Apidologie*, Springer-Verlag, v. 43, n. 6, p. 609–617, 2012.

FRANCOY, T.; GRASSI, M.; IMPERATRIZ-FONSECA, V.; MAY-ITZA, W. J. Geometric morphometrics of the wing as a tool for assigning genetic lineages and geographic origin to *melipona beecheii* (hymenoptera: Meliponini). *Apidologie*, Springer-Verlag, v. 42, n. 4, p. 499–507, 2011.

FRANCOY, T.; SILVA, R.; NUNES-SILVA, P.; MENEZES, C.; IMPERATRIZ-FONSECA, V. et al. Gender identification of five genera of stingless bees (Apidae, Meliponini) based on wing morphology. *Genetics and Molecular Research*, FUNPEC-EDITORIA, v. 8, n. 1, p. 207–214, 2009.

FRANCOY, T. M.; GONÇALVES, L. S.; JONG, D. D. Rapid morphological changes in populations of hybrids between africanized and european honey bees. *Genetics and Molecular Research*, v. 11, n. 3, p. 3349–3356, 2012.

FRANCOY, T. M.; WITTMANN, D.; DRAUSCHKE, M.; MULLER, S.; STEINHAGE, V.; BEZERRA-LAURE, M. A. F.; JONG, D. D.; GONCALVES, L. S. Identification of africanized honey bees through wing morphometrics: two fast and efficient procedures. *Apidologie*, v. 39, n. 5, p. 488–494, 2008.

GALLAI, N.; SALLES, J.-M.; SETTELE, J.; VAISSIÈRE, B. E. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, v. 68, n. 3, p. 810–821, January 2009.

GOLUB, G.; KAHAN, W. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, SIAM, v. 2, n. 2, p. 205–224, 1965.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 2nd. ed. [S.l.]: Pearson Education, 2002.

GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. *Digital Image Processing Using MATLAB*. 2nd. ed. [S.l.]: Gatesmark Publishing, 2003. Hardcover.

GRIMSON, W. E. L. *Object recognition by computer: the role of geometric constraints*. Cambridge, MA, USA: MIT Press, 1990. 25–27 p.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, ACM, New York, NY, USA, v. 11, n. 1, p. 10–18, nov. 2009.

HALL, M. A.; HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 15, n. 6, p. 1437–1447, nov. 2003.

JOHN, G.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, p. 338–345, 1995.

- KANDEMIR, I.; OZKAN, A.; FUCHS, S. Reevaluation of honeybee (*apis mellifera*) microtaxonomy: a geometric morphometric approach. *Apidologie*, Springer-Verlag, v. 42, n. 5, p. 618–627, 2011.
- KIRCHNER, K.; TOLLE, K. H.; KRIETER, J. Decision tree technique applied to pig farming datasets. *Livestock Production Science*, v. 90, n. 2-3, p. 191 – 200, 2004.
- KLEIN, A. M.; VAISSIÈRE, B. E.; CANE, J. H.; STEFFAN-DEWENTER, I.; CUNNINGHAM, S. A.; KREMEN, C.; TSCHARNTKE, T. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society of London B: Biological Sciences*, v. 274, n. 5, p. 303–313, 2007.
- KOCA, A. O.; KANDEMIR, I. Comparison of two morphometric methods for discriminating honey bee (*apis mellifera* L.) populations in turkey. *Turkish Journal of Zoology*, v. 37, p. 205–210, 2013.
- KOHAHI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. [S.l.]: Morgan Kaufmann, 1995. p. 1137–1143.
- LEI, Z.; LIAO, S.; LI, S. Efficient feature selection for linear discriminant analysis and its application to face recognition. In: *Pattern Recognition (ICPR), 2012 21st International Conference on*. [S.l.: s.n.], 2012. p. 1136–1139.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.
- MATLAB. *version 7.14.0 (R2012a)*. Natick, Massachusetts: The MathWorks Inc., 2012.
- MEULEMEESTER, T. D.; MICHEZ, D.; AYTEKIN, A. M.; DANFORTH, B. N. Taxonomic affinity of halictid bee fossils (Hymenoptera: Anthophila) based on geometric morphometrics analyses of wing shape. *Journal of Systematic Palaeontology*, p. 1–10, 2012.
- MICHEZ, D.; MEULEMEESTER, T. D.; RASMONT, P.; NEL, A.; PATINY, S. New fossil evidence of the early diversification of bees: *Paleohabropoda oudardi* from the french paleocene (hymenoptera, apidae, anthophorini). *Zoologica Scripta*, Blackwell Publishing Ltd, v. 38, n. 2, p. 171–181, 2009.
- MIGUEL, I.; BAYLAC, M.; IRIONDO, M.; MANZANO, C.; GARNERY, L.; ESTONBA, A. Both geometric morphometric and microsatellite data consistently support the differentiation of the *apis mellifera* m evolutionary branch. *Apidologie*, v. 42, n. 2, p. 150–161, 2011.
- NIELSEN, D.; EBERT, P.; HUNT, G.; GUZMAN-NOVOA, E.; KINNEE, S.; PAGE, R. Identification of africanized honey bees (Hymenoptera: Apidae) incorporating morphometrics and an improved polymerase chain reaction mitotyping procedure. *Annals of the Entomological Society of America*, Entomological Society of America, v. 92, n. 2, p. 167–174, 1999.

- PARLETT, B. N. *The Symmetric Eigenvalue Problem*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1998.
- POTTS, S. G.; BIESMEIJER, J. C.; KREMEN, C.; NEUMANN, P.; SCHWEIGER, O.; KUNIN, W. E. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution*, v. 3, n. 12, p. 1–9, fev. 2010.
- QUINLAN, J. R. *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- RINDERER, T. E.; SYLVESTER, H. A.; BROWN, M. A.; VILLA, J. D.; PESANTE, D.; COLLINS, A. M.; SPENCER, R.; KLEINPETER, S.; LANCASTER, V. Field and simplified techniques for identifying africanized and european honey bees. *Apidologie*, v. 17, n. 1, p. 33–48, 1986.
- RISH, I. An empirical study of the naive Bayes classifier. *IJCAI-01 workshop on Empirical Methods in Artificial Intelligence*, p. 41–46, 2001.
- ROGATI, M.; YANG, Y. High-performing feature selection for text classification. In: *Proceedings of the Eleventh International Conference on Information and Knowledge Management*. [S.l.: s.n.], 2002. (CIKM '02), p. 659–661.
- ROHLF, F.; BOOKSTEIN, F. *Proceedings of the Michigan Morphometrics Workshop*. [S.l.]: University of Michigan Museum of Zoology, 1990. (Special publication, vol. 1).
- ROHLF, F.; SLICE, D. Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, v. 39, n. 1, p. 40–59, 1990.
- ROHLF, F. J. Relative warp analysis and an example of its application to mosquito wings. In: MARCUS, L. F.; BELLO, E.; GARCIA-VALEDCASAS, A. (Ed.). *Contributions to Morphometrics*. [S.l.]: Museo Nacional de Ciencias Naturales, 1993. p. 131–159.
- ROHLF, F. J. *tpsDig, Version 2.16*. Department of Ecology and Evolution, State University of New York, Stony Brook, NY, USA. 2010.
- ROHLF, F. J. *tpsRelw, Version 1.49*. Department of Ecology and Evolution, State University of New York, Stony Brook, NY, USA. 2010.
- ROHLF, F. J. *Software by F. James Rohlf*. 2014. <<http://life.bio.sunysb.edu/ee/rohlf/software.html>>. Accessed: 2014-05-01.
- ROTH, V.; VEGA, A. P.; STEINHAGE, V.; SCHRODER, S. Pattern recognition combining feature and pixel-based classification within a real world application. *DAGM-Symposium*, Springer, p. 120–129, 1999.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 2nd. ed. [S.l.]: Prentice-Hall, Englewood Cliffs, NJ, 2003.
- SANGEETHA, R.; KALPANA, B. Performance evaluation of kernels in multiclass support vector machines. *International Journal of Soft Computing and Engineering*, v. 1, n. 26, p. 138–145, 2011.

- SANTANA, F. S.; COSTA, A. H. R.; TRUZZI, F. S.; SILVA, F. L.; SANTOS, S. L.; FRANCOY, T. M.; SARAIVA, A. M. A reference process for automating bee species identification based on wing images and digital image processing. *Ecological Informatics*, v. 24, p. 248–260, 2014. ISSN 1574-9541.
- SCHRÖDER, S.; WITTMANN, D.; DRESCHER, W.; ROTH, V.; STEINHAGE, V.; CREMERS, A. The new key to bees: automated identification by image analysis of wings. *Pollinating bees—the Conservation Link Between Agriculture and Nature*, Ministry of Environment, Brasília, p. 209–218, 2002.
- SHINN, A.; KAY, J.; SOMMERVILLE, C. The use of statistical classifiers for the discrimination of species of the genus *gyrodactylus* (monogenea) parasitizing salmonids. *Parasitology*, v. 120, n. 3, p. 261–269, 2000.
- SILVA, F. L.; JACOMINI, R. S.; COSTA, A. H. R. Ensemble learning applied to bee species identification using wing images. In: *II Symposium on Knowledge Discovery, Mining and Learning*. [S.l.: s.n.], 2014. p. 1–8.
- SILVA, F. L.; SELLA, M. L. G.; FRANCOY, T. M.; COSTA, A. H. R. Evaluating classification and feature selection techniques for honeybee subspecies identification by wing images. *Computers and Electronics in Agriculture*, 2015 (Under revision).
- SILVA, L. O. L. A.; KOGA, M. L.; CUGNASCA, C. E.; COSTA, A. H. R. Comparative assessment of feature selection and classification techniques for visual inspection of pot plant seedlings. *Computers and Electronics in Agriculture*, v. 97, p. 47–55, set. 2013.
- SMITH, S. W. *The Scientist and Engineer's Guide to Digital Signal Processing*. San Diego, CA, USA: California Technical Publishing, 1997.
- SOILLE, P. *Morphological Image Analysis: Principles and Applications*. 2. ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.
- STATHAKIS, D. How many hidden layers and nodes? *International Journal of Remote Sensing*, v. 30, n. 8, p. 2133–2147, 2009.
- STEINHAGE, V.; SCHRODER, S.; ROTH, V.; CREMERS, A. B.; DRESCHER, W.; WITTMANN, D. The science of “fingerprinting” bees. *German Research*, v. 28, n. 1, p. 19–21, 2006.
- TARCA, A. L.; CAREY, V. J.; CHEN, X.-W.; ROMERO, R.; DRAGHICI, S. Machine learning and its applications to biology. *PLoS Comput Biol*, Public Library of Science, v. 3, n. 6, p. 953–963, June 2007.
- TOFILSKI, A. Drawing, a program for numerical description of insect wings. *Journal of Insect Science*, v. 4, n. 17, p. 1–5, 2004.
- TOFILSKI, A. *Automated taxon identification in systematics: theory, approaches and applications*. [S.l.]: Boca Raton : CRC Press, 2008.

- TOFILSKI, A. Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie*, v. 39, n. 5, p. 558–563, 2008.
- VANENGELSDORP, D.; HAYES, J.; UNDERWOOD, R. M.; PETTIS, J. A Survey of Honey Bee Colony Losses in the U.S., Fall 2007 to Spring 2008. *PLoS ONE*, Public Library of Science, v. 3, n. 12, p. e4071+, dez. 2008.
- WEEKS, P.; GAULD, I.; GASTON, K.; O'NEILL, M. Automating the identification of insects: a new solution to an old problem. *Bulletin of Entomological Research*, Cambridge University Press, v. 87, n. 2, p. 203–212, 1997.
- WILCOXON, F. Individual comparisons by ranking methods. *Biometrics Bulletin*, JSTOR, p. 80–83, 1945.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*. 3. ed. Amsterdam: Morgan Kaufmann, 2011.
- ZHANG, L.; WANG, X.; QU, L. Feature reduction based on analysis of covariance matrix. In: *ISCST (1)*. [S.l.]: IEEE Computer Society, 2008. p. 59–62.