Ensemble Learning applied to Bee Species Identification using Wing Images

Felipe Leno da Silva[1], Ricardo de Souza Jacomini[1], Anna Helena Reali Costa[1]

Escola Politécnica - Universidade de São Paulo
f.leno@usp.br, ricardo.jacomini@usp.br,anna.reali@usp.br

Abstract. Bees have been extensively studied by biologist, mainly due to their great importance to agriculture. However, the discrimination of some bee species is difficult, requiring expertise and time. This difficulty has been hampering the conduction of new researches in this area. In the last years, the classification of bee species through morphometric features extracted from the bee wing was studied, leading to automated softwares to perform this task for some groups of bees. However, little effort was devoted on exploring different classification algorithms, and linear separation techniques were extensively used. In this article, we evaluate the classification performance of one ensemble learning and several classification techniques with a training set of real bee wings, from 26 different subspecies. The Stacking ensemble learning algorithm performed better than any individual classifier in our training set, proving that ensemble learning techniques can perform better for bee species identification, particularly when the training set is multi-class and class-imbalanced.

Categories and Subject Descriptors: I.2.6 [**Artificial Intelligence**]: Learning; I.5.4 [**Pattern Recognition**]: Applications

Keywords: Geometric Morphometrics, Machine Learning, Pattern Recognition, Stacking, Supervised Learning

1. INTRODUCTION

"One well-worn, and probably accurate, estimate says that one-third of the human diet can be traced directly, or indirectly, to bee pollination." [Delaplane et al. 2000]

Bees are major pollinators and, due to their great importance to agriculture [Klein et al. 2007], various researches have been conducted aiming at their study and conservation. However, some bee species are really difficult to distinguish, thus, this identification task has been hampering the progress of researches in this area, since the correct identification of some species requires time and specialized knowledge. A serious decline of the pollinator populations is being noticed since the end of the 20^{th} century [Buchmann et al. 1997], leading to a pressing need of these researches to identify means to preserve the population of pollinators.

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. 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; Francoy et al. 2008].The wing is an easily accessible part of the bee's body and the veins (*venation*) are easy to distinguish from the wing background in an image. This has allowed the

This work was partially funded by CNPq grant numbers 131238/2013-2 and 311058/2011-6 and CAPES. We wish to thank Tiago Mauricio Francoy, for providing the training set used in this article.

development of softwares that provided a higher level of automation [Roth et al. 1999; Tofilski 2004]. However, these softwares are either hyper-specialized to an specific group of bees or still require much manual interaction. There is still room to automated approaches that work for most of bee species and to new methods that improve the classification rate of wing images.

The current method to perform this task is to take a digital photo of the wing through a microscope and manually mark (usually with the assistance of a computational tool) the junctions of the veins in the wing. These marks are called *landmarks* and the resulting image for each sample will be similar to Figure 1. With the position of all landmarks, it is possible to extract features for classification (see Section 3.4).

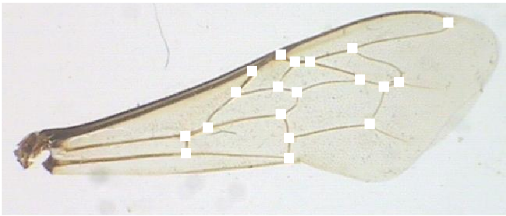

Fig. 1. *Apis mellifera adamii* forewing with landmarks (squares on the vein junctions).

Despite the development of new techniques and softwares, there is no systematic study defining which is the best classifier to bee species identification based on wing images. Previous works have stated that the use of ensemble learning methods can lead to better results than the use of a single classifier [Džeroski and Ženko 2004]. The use of ensemble methods also achieved a better accuracy in classification problems with class-imbalanced and multi-class training sets [Lin and Chen 2012; Sainin and Alfred 2012; Rokach et al. 2013]. Since the biologist's interest relies on classifying a given bee among several species, and some bee species are rarer than others (what can lead to imbalanced training sets), in this work we will evaluate the performance of ensemble learning techniques to bee species identification.

Thence, we chose classifiers that have been successfully used to bee identification or other similar domains to be analyzed, both in an ensemble learning algorithm and as classifier. We will provide references of sucessfull uses of each classifier in Section 2. The contribution of this article will be detailed in Section 3.

This article is organized as follows: Section 2 defines classification and each one of the classifiers used in this article. Section 3 defines our proposal and how we conducted the experiments, while Section 4 brings the results and the discussion about them. Finally, Section 5 outlines the main findings of this article and the open questions to be analyzed in further works.

2. CLASSIFICATION

In the general case of supervised learning, we want to classify a collection of observations into one of a set of predefined classes. The data needed to perform the classification consist in a set of feature values measured somehow from a set of samples, that are associated with a class label defined by an expert [Tarca et al. 2007]. This data is called training set and will be used to train classifiers.

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 Section 3.4). Each classifier has an unique procedure to learn how to classify new observations from a training set. The classifiers used in this research are described in the next subsections.

2.1 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], what is 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 have been extensively used to bee species identification [Tofilski 2008; Koca and Kandemir 2013; Roth et al. 1999].

2.2 Naïve Bayes

The Naïve Bayes classifier estimates the probability of new observations belong to 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.

The probability of a feature assume a given value is usually estimated assuming that all features values for a given class have a gaussian distribution, these distributions are defined for each class in the training phase.

This classifier assumes the independency between features in order to act in a relaxed problem and, consequently, reduce the computacional 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].

2.3 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 being of a given class is determined in the classification phase by applying the logistic function that have been defined in the training phase [Witten et al. 2011].

2.4 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 by: Defining the point corresponding to the observation in the same cartesian space of the samples; Defining the k nearest neighbors (samples) to the observation.; and labeling the observation with the most frequent class among the defined k neighbors.

In order to perform classification, a metric must be defined to calculate the distance between two points, e.g., the Euclidian distance [Russell and Norvig 2003]. This classifier has been used in various applications, obtaining, for example, good results in parasitic species recognition [Shinn et al. 2000].

2.5 C4.5

This classifier builds a decision tree in a training phase to 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 [Quinlan 1993]. This classifier was successfully applied to assist in the decision-making of pig farming [Kirchner et al. 2004].

2.6 Multilayer Perceptron

An Artificial Neural Network (ANN) is a computacional 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 and Hajmeer 2000]. For this type of ANN, the signals are propagated from the input to the output layer through a hidden layer, and each neuron in the hidden layer associates an weight to each input. 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].

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].

2.7 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*.

However, it can be impossible to perform a linear separation between classes; in this case, the feature vector $\boldsymbol{x}$ of p dimensions is transformed into a vector of N dimensions ($N > p$), so that a separation can be found in a higher dimension. A function $\phi : \Re^p \rightarrow \Re^N$ is chosen to transform the feature vector, enabling to define a hyperplane for non-linear data [Cortes and Vapnik 1995].

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. This classifier was successfully used for bee species identification in [Roth et al. 1999]. We used the polynomial kernel in this article.

2.8 Bayesian Network

The Bayesian Network (BayesNet) builds a network that represents the data in the training set. This network can be used to represent the probability of an observation being of a given class. To build a network, two components must be defined: a function for evaluating a given network based on the data and a method for searching through the space of possible networks. These components allow the construction of a network that represents properly the data in the training set and will be used to classify new observations [Witten et al. 2011].

3. PROPOSAL AND EXPERIMENTS

In a real situation where a biologist would build a training set to train a software, some classes would have more samples than others, since some subspecies are harder to find than others. The outcome of this will be a class-imbalanced training set, exactly the same scenario where ensemble learning achieved better classification accuracy [Lin and Chen 2012]. However, there is no published work of our knowledge evaluating the performance of ensemble learning techniques to bee species identification.

The contribution of this article is the evaluation of if an ensemble learning technique can perform better than single classifiers when applied to bee species identification by wings. We will perform this evaluation with a realist training set, composed by real wing images, to ensure an evaluation as better as possible. The evaluation will be carried out with features that are familiar to biologists and achieved the best result so far (Section 3.4). On the next subsection we will describe Ensemble Learning in general and the Stacking algorithm, that was used in this research. The cross-validation technique used to extract metrics for this evaluation will be described in Section 3.2.

3.1 Ensemble Learning

The idea of ensemble learning is to employ multiple learners and combine their predictions. There is no definitive taxonomy. In [Jain et al. 2000] eighteen classifier combination schemes are summarized, while in [Witten et al. 2011] several methods for combining multiple models are detailed. This method has been shown to be an effective tool for solving multi-class classification tasks [Rokach et al. 2013].

The generalization ability of an ensemble is usually much stronger than that of a single learner. For the following reasons: (1)The training data might not provide sufficient information for choosing a single best learner; (2)The search processes of the learning algorithms might be imperfect; (3)The hypothesis space being searched might not contain the true target function, while ensembles can give some good approximation.

In this article we applied the method called Stacked Generalization (Stacking), which is an ensemble method that combines the output from different classifiers to achieve greater predictive accuray. It involves training a learning algorithm to combine the predictions of several other classifiers. In typical stacking implementation, a number of first-level individual learners are generated from the training set by employing different learning algorithms. Then all individual learners are combined by a second-level learner which is called as meta-learner. This is done by training the meta-learner with the output of each individual learners, rather than with the original training set. Stacking typically yields to a better performance than any of the individual models[Wolpert 1992].

For this article, the Stacking was used with SVM as meta-learner and the first-level classifiers were used as follows: SVM, Logistic, Naïve Bayes and BayesNet. Other combinations of classifiers were also tested in the first-level, but this combination achieved a better performance, thus we will present the results of this setup.

3.2 Cross-Validation

As explained in Section 2, 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 excelent performance to classify the training set, however does not keep the same performance classifying new, previously unknown, observations [Russell and 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.

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.

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.

Since we have an uneven training set, we have used the stratified cross-validation in order to avoid leaving folds without samples of some classes in the training set division.

3.3 Experimental setup

In order to verify if ensemble methods can improve the performance of bee species identification, we used a training set provided by an taxonomist with marked landmarks and the respective label of 1821 wing images of 26 subspecies of *Apis mellifera* to build the training set. The number of examples per class is different, since some subspecies are harder to find than others. The number of examples per class ranges from 10 to 150.

In order to calculate a measure to define the best classification method, we have performed a 10-fold stratified cross-validation 20 times with different randomic splits. We then compared the observed performance for all the classifiers in these experiments, calculating an error margin with the standard error and testing the statistic relevance of the results with a Wilcoxon Signed Rank Test [Wilcoxon 1945], both with 95% confidence factor.

All the experiments were executed on MATLAB [MATLAB 2012] and the WEKA API [Hall et al. 2009] was used as the implementation of the classifiers *Naïve Bayes*, *Logistic*, *MLP*, *KNN*, *BayesNet*, *Stacking*, and *SVM*. The MATLAB implementation of the *LDA* was used.

3.4 Feature Extraction

We extracted landmark-based features from Geometric Morphometrics (reported to be the best method to extract features from landmarks [Tofilski 2008; Koca and Kandemir 2013; Roth et al. 1999]) that have already been sucessfully used for bee species identification [Tofilski 2008; Koca and Kandemir 2013; Francoy et al. 2008; Kandemir et al. 2011; Santana et al. 2014]. Since all features rely on the landmarks' position, firstly we chose 19 landmarks as shown in Figure 1 using the tpsDig software [Rohlf 2010a]. Then, we extracted the features: Aligned Coordinates, Centroid Size, Weight Matrix of the Principal Warps, and the Relative Warps scores in the tpsRelw software [Rohlf 2010b].

The Aligned Coordinates consist in the x and y coordinates of all landmarks (as in Figure 1) after a normalization to remove discrepancies regarding to translation, scale and rotation [Bookstein 1991]. The centroid size is the square root of the quadratic distance between each landmark and their centroid [Bookstein 1991]. The Weight Matrix of the Principal Warps and the Relative Warps scores are tools to shape variation analysis, and (for 19 landmarks) correspond to 34 ratios each one [Bookstein 1989; 1991] .

Thus, our resulting feature vector has 107 features and the class label.

4. RESULTS AND DISCUSSION

Figure 2 shows the hit ratio achieved by each classifier in the experiments, where the values represent a mean of the 20 observed Cross-Validation performances (as explained in Section 3). The error margin is defined by the standard error and, according to the Wilcoxon Signed Rank Test with 95% confidence factor, all the differences between classifiers have statistical meaning. The results confirm our hypothesis that Ensemble Learning methods could perform better than individual classifiers, since the *Stacking* was the best classification algorithm in our experiments.

Surprisingly, the Naïve Bayes (that, to the best of our knowledge, has been cited in only one research on this domain [Santana et al. 2014]) performed better than the widely used LDA. This outcome indicate that the introduction of new classification algorithms can be beneficial to bee species identification, and, in the case of training samples with uneven number of samples per species, the use of ensemble learning techniques can lead to a better performance than the use of individual classifiers. Thus, when the accuracy is the priority, the use of Stacking is recommended over single classifiers. The increase in computational costs demanded by ensemble techniques was not explored in this article, thus, this counter-balance must be evaluated if the biologist has time restrictions to perform the classification task. In some situations, the Naïve Bayes can be a better option due to its lower computational costs and for having a performance only marginally worse than Stacking.

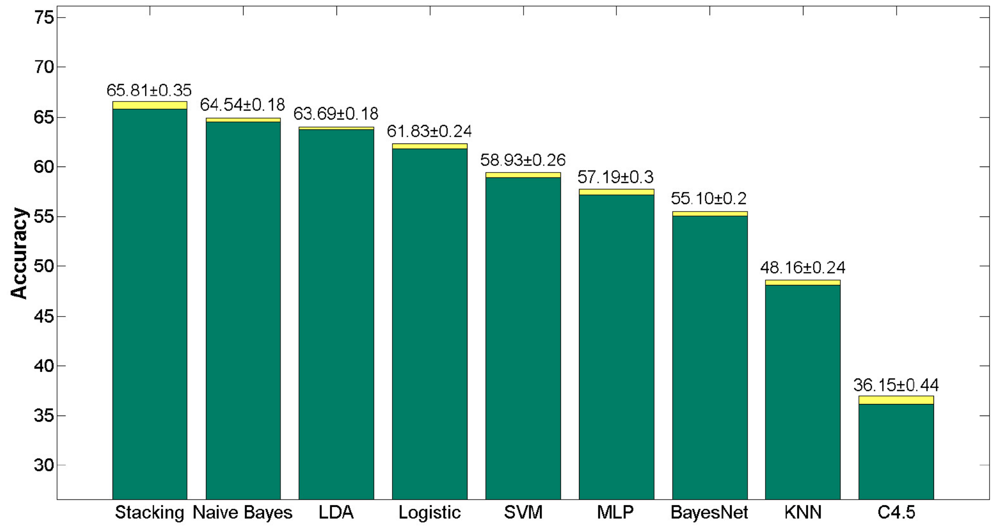

Fig. 2. Mean of the cross-validation performance achieved by each classifier in 20 runs, the error margin is shown in yellow. Accuracy values are in percentage.

5. CONCLUSION

In this article we outlined the importance of bee species identification and the need of better techniques and algorithms to perform this task. We have evaluated an Ensemble Learning technique and 8 state-of-the-art classifiers in an experiment with real wing images. The Ensemble Learning technique performed better than individual classifiers and, surprisingly, the Naïve Bayes (scarcely used in this domain) also performed better than the widely used LDA, thus indicating that there is still much more room to improvements regarding the classification techniques for bee species identification.

Further works can focus on the definition of the best classifier to pixel-extracted features, that have been proved to increase the performance of classification when used with the appropriate classification algorithm [Santana et al. 2014]. It is also possible to evaluate the performance of other ensemble learning techniques, such as Bagging, Radom Forest, Boosting, and Vote [Witten et al. 2011]. Other open question is the compromise between accuracy and computational costs (ensemble techniques demand more computation times), since in this article we did not define when it is better to use Stacking or Naïve Bayes in situations that have time restrictions.

REFERENCES

Basheer, I. A. and Hajmeer, M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods* vol. 43, pp. 3–31, 2000.

Bookstein, F. L. Principal warps: thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 11 (6): 567–585, Jun, 1989.

BOOKSTEIN, F. L. *Morphometric tools for landmark data: geometry and biology.* Cambridge University Press, Cambridge, USA, 1991.

BUCHMANN, S., NABHAN, G., AND MIROCHA, P. *The Forgotten Pollinators.* A Shearwater Book Series. Island Press, 1997.

CORTES, C. AND VAPNIK, V. Support-vector networks. *Machine Learning* 20 (3), Sept., 1995.

DELAPLANE, K., MAYER, D., AND MAYER, D. *Crop Pollination by Bees.* CABI, 2000.

DŽEROSKI, S. AND ŽENKO, B. Is combining classifiers with stacking better than selecting the best one? *Machine learning* 54 (3): 255–273, 2004.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7 (7): 179–188, 1936.

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

HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I. H. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11 (1): 10–18, Nov., 2009.

JAIN, A. K., DUIN, R. P. W., AND MAO, J. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22 (1): 4–37, 2000.

KANDEMIR, I., OZKAN, A., AND FUCHS, S. Reevaluation of honeybee (apis mellifera) microtaxonomy: a geometric morphometric approach. *Apidologie* 42 (5): 618–627, 2011.

KIRCHNER, K., TOLLE, K. H., AND KRIETER, J. Decision tree technique applied to pig farming datasets. *Livestock Production Science* 90 (2-3): 191 – 200, 2004.

KLEIN, A. M., VAISSIÈRE, B. E., CANE, J. H., STEFFAN-DEWENTER, I., CUNNINGHAM, S. A., KREMEN, C., AND TSCHARNTKE, T. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society of London B: Biological Sciences* 274 (5): 303–313, 2007.

KOCA, A. O. AND KANDEMIR, I. Comparison of two morphometric methods for discriminating honey bee (apis mellifera l.) populations in turkey. *Turkish Journal of Zoology* vol. 37, pp. 205–210, 2013.

LIN, W.-J. AND CHEN, J. J. Class-imbalanced classifiers for high-dimensional data. *Briefings in bioinformatics*, 2012.

MATLAB. *version 7.14.0 (R2012a).* The MathWorks Inc., Natick, Massachusetts, 2012.

QUINLAN, J. R. *C4.5: programs for machine learning.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

RISH, I. An empirical study of the naive Bayes classifier. *IJCAI-01 workshop on Empirical Methods in Artificial Intelligence*, 2001.

ROHLF, F. J. tpsDig, Version 2.16. Department of Ecology and Evolution, State University of New York, Stony Brook, NY, USA, 2010a.

ROHLF, F. J. tpsRelw, Version 1.49. Department of Ecology and Evolution, State University of New York, Stony Brook, NY, USA, 2010b.

ROKACH, L., SCHCLAR, A., AND ITACH, E. Ensemble methods for multi-label classification. *arXiv preprint arXiv:1307.1769*, 2013.

ROTH, V., VEGA, A. P., STEINHAGE, V., AND SCHRODER, S. Pattern recognition combining feature and pixel-based classification within a real world application. *DAGM-Symposium*, 1999.

RUSSELL, S. AND NORVIG, P. *Artificial Intelligence: A Modern Approach.* Prentice-Hall, Englewood Cliffs, NJ, 2003.

SAININ, M. S. AND ALFRED, R. A direct ensemble classifier for imbalanced multiclass learning. In *4th Conference on Data Mining and Optimization.* IEEE, pp. 59–66, 2012.

SANTANA, F. S., COSTA, A. H. R., TRUZZI, F. S., SILVA, F. L., SANTOS, S. L., FRANCOY, T. M., AND SARAIVA, A. M. A reference process for automating bee species identification based on wing images and digital image processing. *Ecological Informatics* 0 (0): –, 2014.

SHINN, A., KAY, J., AND SOMMERVILLE, C. The use of statistical classifiers for the discrimination of species of the genus gyrodactylus (monogenea) parasitizing salmonids. *Parasitology* 120 (3): 261–269, 2000.

TARCA, A. L., CAREY, V. J., CHEN, X.-W., ROMERO, R., AND DRAGHICI, S. Machine learning and its applications to biology. *PLoS Comput Biol* 3 (6): 953–963, June, 2007.

TOFILSKI, A. Drawwing, a program for numerical description of insect wings. *Journal of Insect Science* 4 (17): 1–5, 2004.

TOFILSKI, A. Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie* 39 (5): 558–563, 2008.

WEEKS, P., GAULD, I., GASTON, K., AND O'NEILL, M. Automating the identification of insects: a new solution to an old problem. *Bulletin of Entomological Research* 87 (2): 203–212, 1997.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics bulletin*, 1945.

WITTEN, I. H., FRANK, E., AND HALL, M. A. *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, Amsterdam, 2011.

WOLPERT, D. H. Stacked generalization. *Neural networks* 5 (2): 241–259, 1992.