big data and cognitive computing

MDPI

Article

DeepWings©: Automatic Wing Geometric Morphometrics Classification of Honey Bee (*Apis mellifera*) Subspecies Using Deep Learning for Detecting Landmarks

Pedro João Rodrigues [1,*], Walter Gomes [1,2] and Maria Alice Pinto [3,*]

[1] Research Center in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal; walterbsg@gmail.com
[2] Ciência da Computação, DAINF—Universidade Tecnológica Federal do Paraná, Av. Monteiro Lobato, Ponta Grossa 84016-210, Brazil
[3] Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
* Correspondence: pjsr@ipb.pt (P.J.R.); apinto@ipb.pt (M.A.P.)

Abstract: Honey bee classification by wing geometric morphometrics entails the first step of manual annotation of 19 landmarks in the forewing vein junctions. This is a time-consuming and error-prone endeavor, with implications for classification accuracy. Herein, we developed a software called DeepWings© that overcomes this constraint in wing geometric morphometrics classification by automatically detecting the 19 landmarks on digital images of the right forewing. We used a database containing 7634 forewing images, including 1864 analyzed by F. Ruttner in the original delineation of 26 honey bee subspecies, to tune a convolutional neural network as a wing detector, a deep learning U-Net as a landmarks segmenter, and a support vector machine as a subspecies classifier. The implemented MobileNet wing detector was able to achieve a mAP of 0.975 and the landmarks segmenter was able to detect the 19 landmarks with 91.8% accuracy, with an average positional precision of 0.943 resemblance to manually annotated landmarks. The subspecies classifier, in turn, presented an average accuracy of 86.6% for 26 subspecies and 95.8% for a subset of five important subspecies. The final implementation of the system showed good speed performance, requiring only 14 s to process 10 images. DeepWings© is very user-friendly and is the first fully automated software, offered as a free Web service, for honey bee classification from wing geometric morphometrics. DeepWings© can be used for honey bee breeding, conservation, and even scientific purposes as it provides the coordinates of the landmarks in excel format, facilitating the work of research teams using classical identification approaches and alternative analytical tools.

Keywords: wing landmarks; deep learning; wing geometric morphometrics; honey bee classification; software

check for updates

Citation: Rodrigues, P.J.; Gomes, W.; Pinto, M.A. DeepWings©: Automatic Wing Geometric Morphometrics Classification of Honey Bee (*Apis mellifera*) Subspecies Using Deep Learning for Detecting Landmarks. *Big Data Cogn. Comput.* **2022**, *6*, 70. https://doi.org/10.3390/bdcc6030070

Academic Editor: Min Chen

Received: 2 June 2022
Accepted: 24 June 2022
Published: 27 June 2022

1. Introduction

The western honey bee (*Apis mellifera* L.) differentiated into 31 subspecies in its native range in Eurasia and Africa [1–4]. The great majority of this variation was recognized early by the father of honey bee taxonomy, Friedrich Ruttner, who identified 24 subspecies using 36 morphological traits derived from pilosity, pigmentation, length of different body parts, and wing venation. Analysis of the 36 traits still represents the gold standard of honey bee classification and is required for scientific applications, as the description of new subspecies [5]. However, measuring and analyzing all 36 traits is labor-intensive and involves expert knowledge, making classical morphometry unsuitable for colony identification for commercial, conservation, or even scientific purposes. To circumvent this limitation, efforts have recently been made to simplify and automate the identification of honey bee subspecies [6–9]. Yet, a tool for simple, fast, and inexpensive honey bee classification remains unavailable for use by the beekeeping community.

Here, we used machine learning tools to develop an open-source software, DeepWings©, to assist beekeepers, queen breeders, as well as researchers in identifying honey bee subspecies by wing geometric morphometrics in an entirely automatic, rapid, easy, and free-of-charge manner. The steps involved in the development of DeepWings© are fully detailed in the next sections, whereas the remainder of the Introduction section provides the state-of-the-art in honey bee identification by morphology-based methods, with an emphasis on wing venation traits. The Materials and Methods section starts with the modeling of the setup to obtain the final solution. Then, the wing images are described as well as the steps involved in the construction of the datasets used in the training, including data augmentation to deal with the problem of the small wing dataset used for classification training. The ensuing subsections detail the methods involved in the three major stages of the DeepWings© architecture. The first stage (preprocessing) encompasses the curation of the raw wing images using CNN, which entails the application of filters, wing detection, wing cropping, and wing size normalization. The overall approach allows for the analysis of images showing a wide range of variations regarding visual artifacts, noise, pose, and illumination, as well as containing a variable number of wings. The second stage (landmarks detection) entails detection (by the U-Net), extraction, and sorting of the taxonomically informative wing venation traits (landmarks) from the wing images segmented in the previous stage. In the third stage (classification), the extracted landmarks are subjected to Procrustes normalization, which handles translation, rotation, and scale wing invariances, before entering the SVM classifier. The performance of the methods involved in the three stages of the architecture is assessed in the Results and Discussion section, and DeepWings© is compared to other wing-based classification tools. This section also presents the implementation of DeepWings© as a Web service, its attributes, and its multiple applications in the real world. This work ends with the main findings in the Conclusions section.

Background

Classical morphometry has been replaced by labor-effective alternatives based on the forewing shape patterns, which are typically assessed on honey bee workers (infertile females). The forewings, specifically the vein junctions, carry high-information content and are therefore of great value for the identification of honey bee subspecies [5,10–12]. This feature, together with the quasi-two-dimensional structure of forewings, makes this body part well-suited for computerizing and automating honey bee classification using image analysis.

A suite of forewing traits is used, singly or in conjunction, in the identification of honey bee subspecies [5]. Among these are the cubital index, the hantel index, and the discoidal shift angle [13], which can be measured on wing images by the semi-automatic proprietary software CBeeWing (https://www.cybis.se/index.htm; accessed on 26 April 2022). This software is commonly employed by beekeepers engaged in the conservation of the endangered *Apis mellifera mellifera* subspecies in northern Europe [14]. However, by using a limited number of traits, CBeeWing does not take full advantage of the information content carried by the wing shape, making identification less accurate. On the other side of the spectrum is the very intensive DAWINO (Discriminant Analysis with Numerical Output) method, which requires measurements of 30 forewing characters encompassing vein angles, vein lengths, and indexes [5].

Wing shape analysis based on wing geometry, dubbed wing geometric morphometrics, offers an interesting alternative for honey bee identification [5,10]. Wing geometric morphometrics is widely used in insect taxonomy in general and was revealed as particularly useful for identifying bee species [15–19] and honey bee lineages and subspecies for a wide range of purposes [6,8,11,12,20–25]. Using this method, wing shape variations are captured by 19 data points, known as landmarks, acquired from the vein junctions annotated in images (Figure 1a,b). Given that the locations of the 19 landmarks are subspecies-specific, deviations in positional coordinates can be used in honey bee identification (Figure 1c,d).

Most of the 19 landmarks are also employed for calculating all vein lengths and angles by both the classical morphometry and the DAWINO methods [5]. To the best of our knowledge, the only system publicly available for honey bee identification by wing geometric morphometrics is implemented by the software IdentiFly [26]. The problem is that IdentiFly is a semi-automatic software requiring several steps for wing classification, often including manual correction of landmark annotations, making it very difficult for routine use by queen breeders or beekeepers.

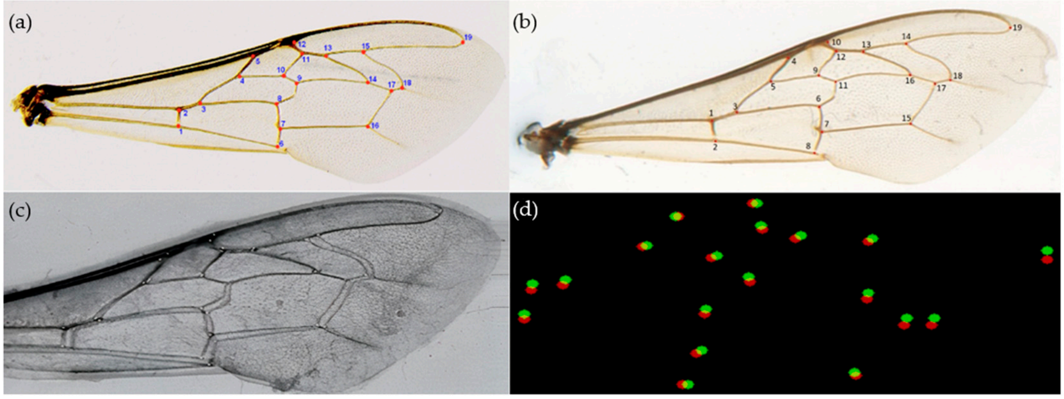


Figure 1. Right forewings of honey bee workers showing the 19 landmarks. (**a**) An example of a manually annotated image used for training the landmark segmenter. (**b**) Example of an image annotated automatically by the software DeepWings©. (**c**) Overlapping of the M-lineage subspecies *Apis mellifera mellifera* and *Apis mellifera iberiensis* right forewings, as extracted by the wing detector of DeepWings©. (**d**) Extracted landmarks of the two forewings, after Procrustes alignment, showing the positional deviations between these two honey bee subspecies.

The most recent advance in honey bee classification comes from the application of artificial intelligence through techniques of machine learning [9]. De Nart and colleagues [9] based their classification method on the entire wing and used Convolutional Neural Networks (CNN) to develop an end-to-end solution for resolving the images. Unfortunately, this new tool is limited to the classification of only seven honey bee subspecies and it was not made available for use by the scientific or beekeeping community.

In a global world, maintaining the genetic integrity of native honey bee subspecies is becoming increasingly demanding. In this context, it is important that queen breeders engaged in honey bee conservation programs have tools at their disposal for colony identification. While molecular tools offer the most accurate solution for subspecies identification [27,28], they are not affordable for most queen breeders, in which case morphometry becomes the only option. Here, we filled a gap in the geometric morphometrics identification of honey bees by developing a fully automated software that is easy to use and is freely available as a Web service. To that end, we implemented machine learning techniques for (i) detecting wings using CNN, (ii) segmenting landmarks using U-Net, and (iii) classifying models using a support vector machine (SVM). Using this multi-step approach, we addressed three main questions: (i) Can the CNN-based wing detector handle images with multiple wings of varying orientations, ensuring the uniformity of the wing shape pattern at the entrance of the landmarks segmenter? (ii) Is the U-Net capable of extracting the taxonomically informative traits from the wing venation with a high level of precision, as required for accurate subspecies classification? (iii) How efficient is SVM at generalizing the subtle differences between the closely related honey bee subspecies and hence ensuring accurate classification?

2. Materials and Methods

2.1. Modelling of the Solution

Figure 2 depicts in four frames the modulation employed in this work to achieve the final solution. The top-left frame shows the organization of the datasets that were used to train the three machine learning models: the wing detector, the landmarks segmenter, and the wing classifier. These models were trained using the configurations displayed on the remaining frames and were fed with the data originating from the top-left frame. The bottom-left frame shows the organization of the wing detector, in which the wing images were the inputs and the positions of the bounding boxes (surrounding the manually annotated landmarks) were the targets for the CNN learning models. The bottom-right frame shows the organization of the landmarks segmenter, in which the U-Net inputs were loaded with the wing images and the U-Net outputs were loaded with the corresponding landmark images (masks). Finally, the top-right frame shows the intervening elements of the subspecies classifier. At this stage of the training, the landmarks previously annotated (bottom-right frame) were processed (PCA-Sorting-Procrustes) to assure their geometric stabilization before entering the SVM classifier. The One-Hot encoding was used in the SVM outputs, in order to represent the wings of the different subspecies.

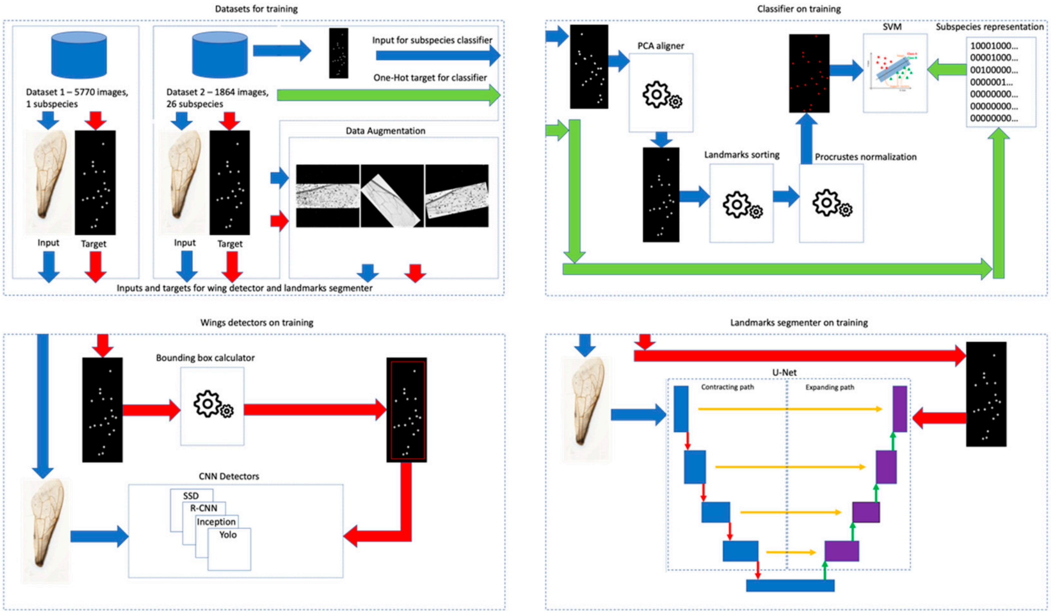


Figure 2. Modeling of the setup to obtain the final solution: the datasets (**top-left** frame), the wings detector (**bottom-left** frame), the landmarks segmenter (**bottom-right** frame), and the subspecies classifier (**top-right** frame).

2.2. Image Datasets

Two sets of images were used to train the system developed in this study. Dataset 1 comprised 5770 images of the right forewing of honey bee workers. The wings were photographed from mounts on microscopic slides (~10 wings per slide) using a stereomicroscope attached to a digital camera with a resolution of 1000 pixels per centimeter. These images were manually annotated for the 19 landmarks by one single operator, following the positional order portrayed in Figure 1a. Of the 5770 images, 3518 were collected from Iberian colonies of *Apis mellifera iberiensis* identified by genetic data [25,29] whereas 2252 were collected from Azorean colonies of mixed ancestry [25]. Dataset 1 was annotated for the population analysis carried out by Ferreira and colleagues [25].

Dataset 2 comprised 1864 images of the right forewing of honey bee workers obtained from the Morphometric Bee Data Bank in Oberursel, Germany, where the original specimens used by F. Ruttner [1] in his seminal taxonomic work are deposited. These images were taken from mounts on microscopic slides using a stereomicroscope attached to a digital camera with a resolution of 650 pixels per centimeter. Dataset 2 comprised wings belonging to 26 subspecies sampled from across the *A. mellifera* native range. The number of wings per subspecies varied among the four evolutionary lineages and subspecies. Specifically, the African lineage (A) included 116 *A. m. adansonii*, 30 *A. m. capensis*, 60 *A. m. intermissa*, 70 *A. m. lamarckii*, 55 *A. m. litorea*, 10 *A. m. major*, 80 *A. m. monticola*, 50 *A. m. ruttneri*, 20 *A. m. sahariensis*, 140 *A. m. scutellata*, 70 *A. m. unicolor*, and 133 *A. m. jemenitica* wing images. The eastern European lineage (C) included 150 *A. m. carnica*, 90 *A. m. cecropia*, 110 *A. m. ligustica*, 20 *A. m. macedonica*, and 10 *A. m. siciliana* wing images. The western European lineage (M) included 20 *A. m. iberiensis* and 140 *A. m. mellifera* wing images. Finally, the Oriental lineage (O) included 50 *A. m. adami*, 50 *A. m. anatoliaca*, 60 *A. m. armeniaca*, 120 *A. m. caucasia*, 40 *A. m. cypria*, 80 *A. m. meda*, and 90 *A. m. syriaca* wing images.

Contrary to the high quality of most images in dataset 1 (example in Figure 1a), dataset 2 contained numerous images with visual artifacts, noise, pose variations, illumination variations, and specular light reflections (examples in Figure 3), showing the difficulty of generalizing a solution for a reliable classification system. Furthermore, most images in both datasets contained more than one wing.

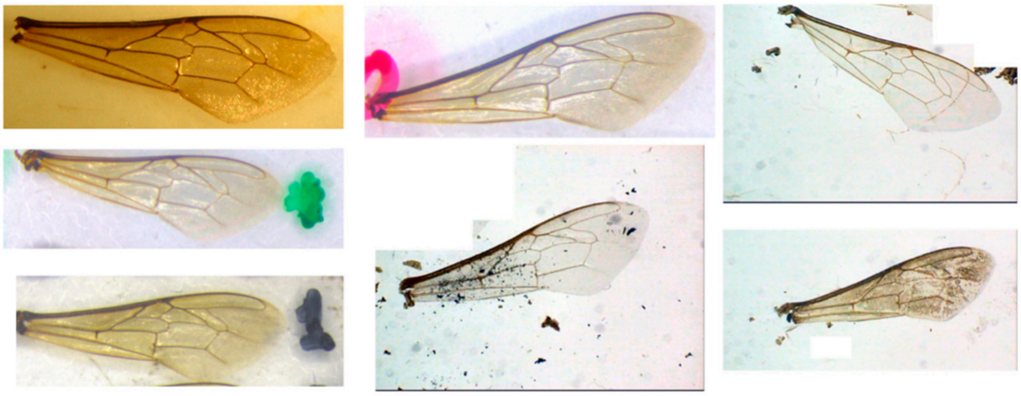


Figure 3. A sample of honey bee forewings showing the range of image quality.

The two datasets were aggregated and then split into three subsets according to the following ratios: 80% for the training subset, 10% for the validation subset, and 10% for the testing subset. These three new datasets were randomly built from the original aggregated dataset. The training dataset was used for the training process, the validation dataset was used to tune the parameters of the machine learning models, and the testing dataset was used to assess the final functional performance of the machine learning models.

2.2.1. Masks

Masks of the size of the input images were created by an algorithm based on the manually marked landmarks from dataset 1, with the pixels forming the landmarks denoted in white and the background in black. These masks were the targets of the U-Net neural network [30]. Figure 4 provides examples of the output masks created for the U-Net training.

2.2.2. Data Augmentation

Dataset 2 was considerably smaller than dataset 1 and comprised numerous images representing a large spectrum of visual variations (Figure 3), contrasting with the high image quality of dataset 1 (Figure 1a). To accommodate the problem of unbalanced datasets, dataset 2 was artificially expanded. Employment of a data augmentation approach [31] enabled the construction of a large database for increased precision of the automatic landmarks segmenter. To accommodate the problem of image variations, specific visual features were simulated, including dust, noise, and drastic angle changes (Figure 5).

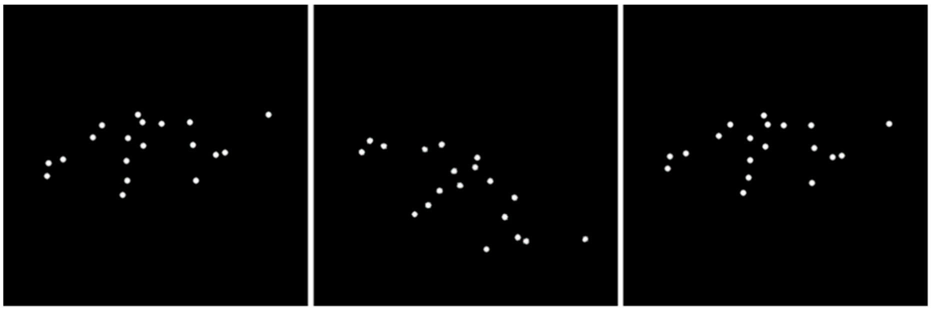


Figure 4. Examples of U-Net output masks.

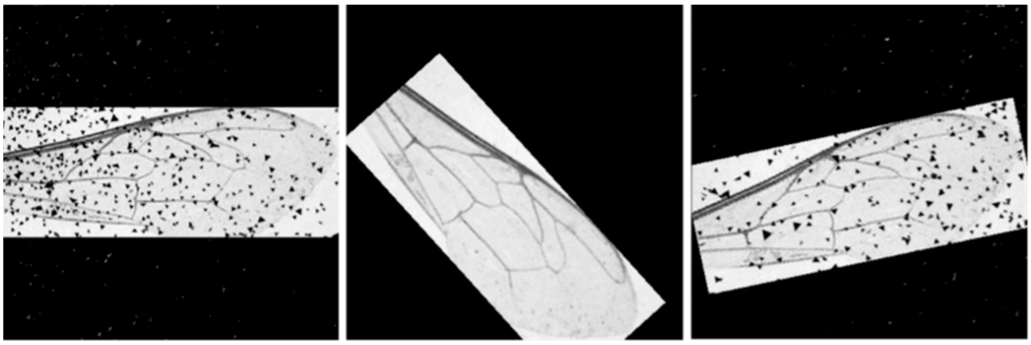


Figure 5. Examples of artificially created images to simulate visual variations.

2.3. Processing and Analyzing Wing Images

The processing and analysis of wing images encompassed three major stages: preprocessing, landmark extraction, and classification (Figure 6). The system must be capable of handling different variables of the images, such as wing pose, image size, and multiple wings in a single image.

2.3.1. Preprocessing

Preprocessing comprises several steps needed for curating the images for subsequent detection of the 19 landmarks. These steps included the application of filters, wing detection, wing cropping, and wing size normalization (Figure 6). After opening the images, two filters were applied: a CLHAE (Contrast Limited Adaptive Histogram Equalization) filter [32], to highlight important image features, and a Gaussian filter, to remove noise. The images were resized to a static value and a CNN-based [33] object detector, capable of perceiving the existence of each wing within an image, was used. This approach enabled image cropping in a normalized manner (Figure 7). Several object detector models were tested, including SSD MobileNet v1 FPN coco [34], Faster R-CNN NAS [35], Faster R-CNN Inception Resnet v2 Atrous Coco [36], and YoloV3 [37].

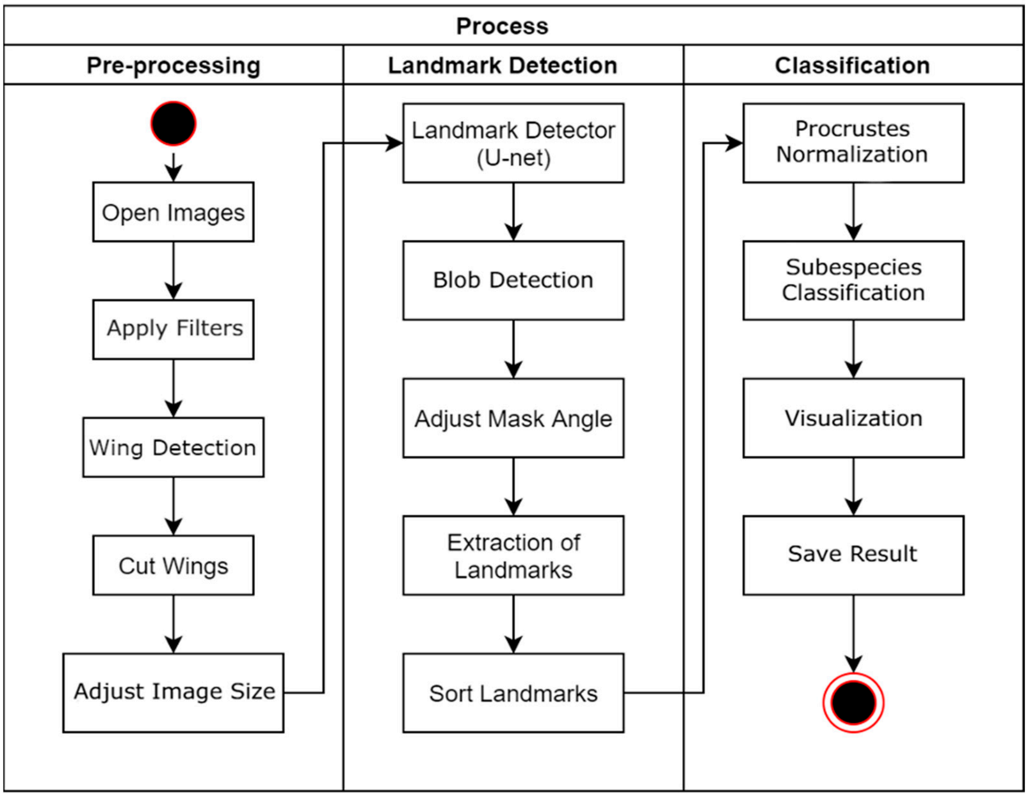


Figure 6. Pipeline developed for the processing of forewing images for honey bee classification.

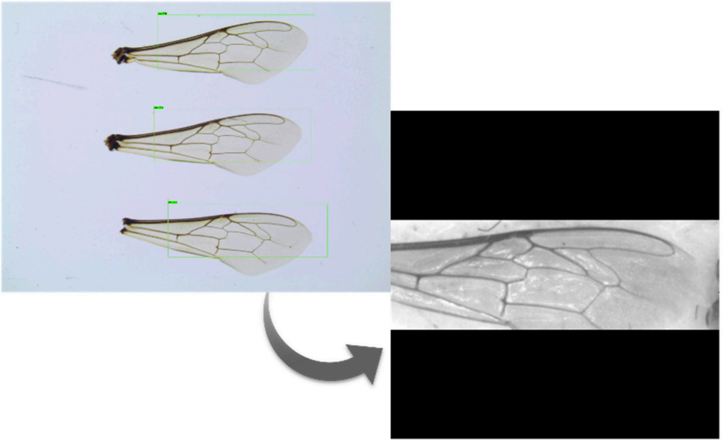


Figure 7. An example illustrating the end product of the preprocessing stage of a typical multi-wing image.

None of the datasets had annotated wing bounding boxes. However, it was possible to infer bounding boxes using the landmark coordinates of the training dataset (Figure 8). Prior to training the wing detection, the target landmarks were used to delineate the wing region and then the bounding boxes were inferred from the peripherical landmarks. A spatial margin around the landmarks ensured that they would fall inside the bounding box.

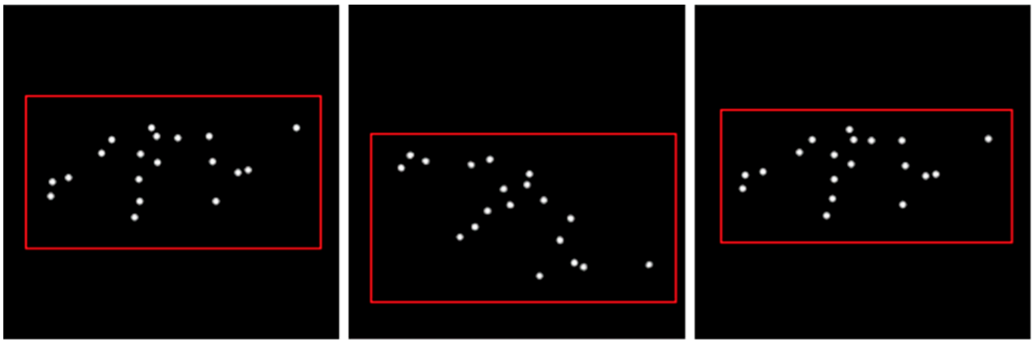


Figure 8. Bounding boxes, inferred from landmarks, used for training the wing detector.

2.3.2. Landmark Detection

Landmark detection comprises several steps, including implementation of the landmark detector, Blob detection, adjustment of mask angle, extraction and sorting of landmarks (Figure 6). The U-Net detector was revealed to be more precise in segmenting the 19 landmarks than a classification CNN and the classical approaches (e.g., adaptive binarization) available in the OpenCV library [38]. The U-Net architecture consists of a contracting path to capture context and a symmetric expanding path, which enables precise landmark positioning by extracting their mass centers (Figure 9). The U-Net was trained using a GTX1080ti GPU and the Keras deep learning framework [39].

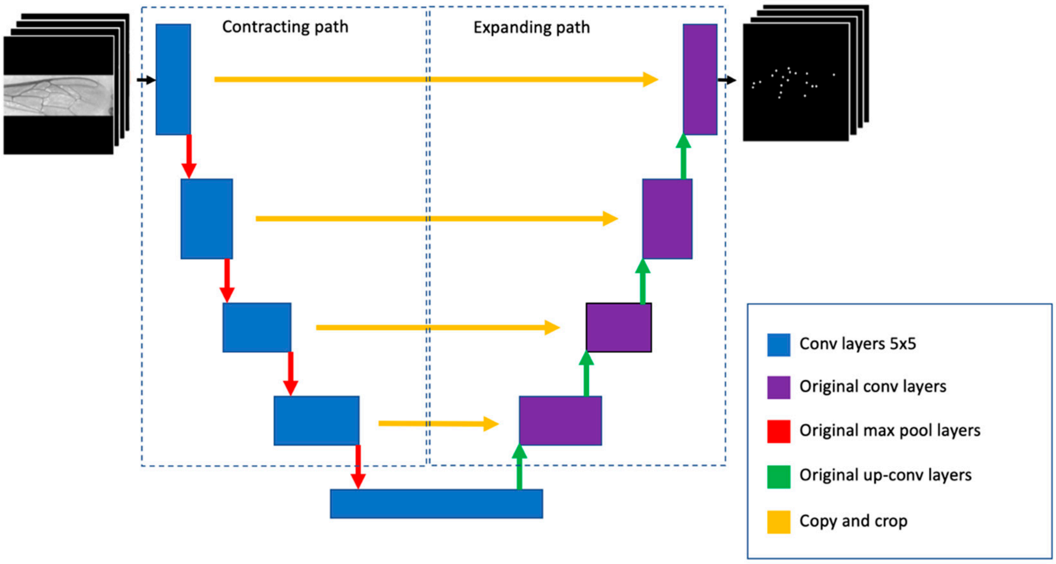


Figure 9. U-Net architecture used for segmentation of the landmarks. It is based on CNN with a contracting path to capture context and a symmetric expanding path to enable precise landmark positioning.

The images for the U-Net training were converted to greyscale, as the searched landmark patterns do not depend on color. This helped save computational memory and reduce the "curse of dimensionality" of the neural model. Because the wing images had variations in the positional angles, the cropped images tended to exhibit different sizes (Figure 8). Moreover, the U-Net input size needed to be static (400 × 400 inputs) and the image width–height ratio of the wing could not be altered in order to avoid pattern deformations. Therefore, the wing image was stacked on a black background where the horizontal axis of the wing image was scaled to the limits of that black layer (Figure 10).

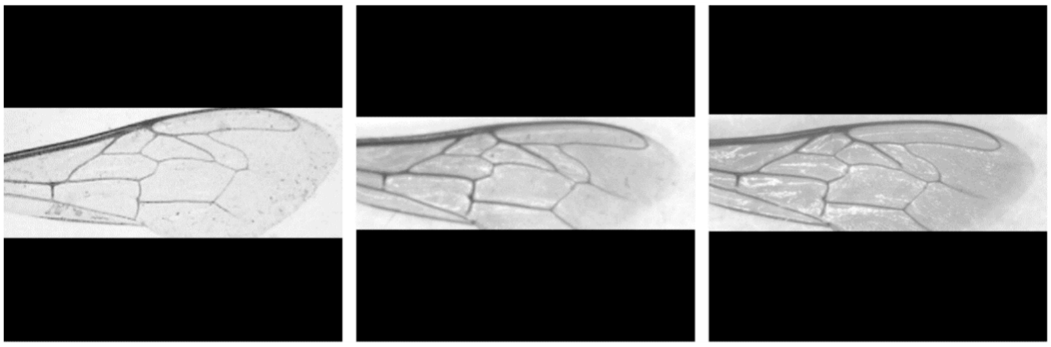


Figure 10. Examples of mounted images presented to the U-Net input.

During the U-Net training, it became evident that using only one pixel to represent each landmark was not a suitable solution. Therefore, to increase the signal-to-noise ratio, the U-Net received a small circle (synthetic landmark) for each landmark, reinforcing the landmark as the target region for the U-Net. The synthetic landmarks taken by the U-Net generated the regions illustrated in Figure 11, after the training. As expected, these regions are not a perfect circle, contrary to the U-Net target masks shown in Figure 4.

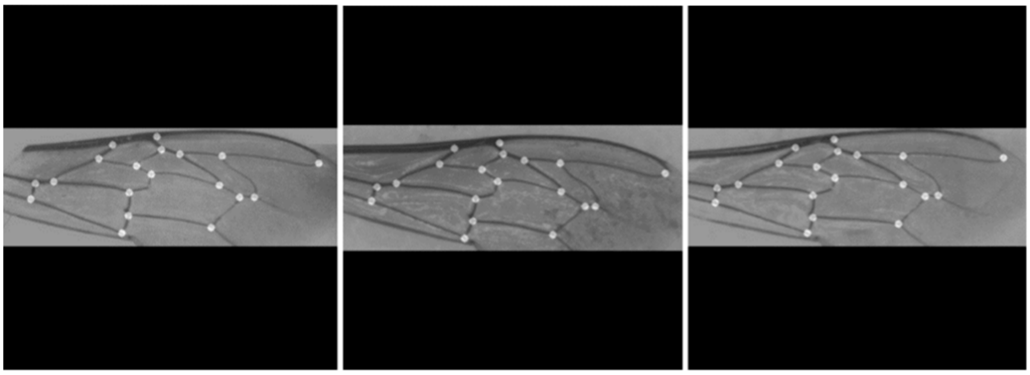


Figure 11. Landmarks superimposed on original wing images.

When the U-Net segments the 19 landmarks, it does not identify each one in a differentiated manner. This is a critical issue because the classifier can only make the right decision if the order of the landmarks displayed in Figure 1b is kept during the processing runs. The first step to ensure a standardized landmarks extraction was based on Principal Components Analysis [40], which identifies the largest eigenvector that gives the angle of the landmarks' distribution. Then, the rotation of the entire mask following that angle allows having the landmarks rotationally aligned to the horizontal axis. This procedure simplified the mechanism of extracting the landmarks in a standard order because their patterns could be scanned in a consistent pose.

From the masks generated by the U-Net, it was then required to extract the mass center of the detected landmark regions. The most straightforward way to do this was to compute the center of the segmented regions using the Blob Detector implemented in the OpenCV library. The total number of expected landmarks (19) was used to validate segmentation; if that number was not confirmed, the process would fail. Following detection, all 19 landmarks were sorted out according to their positional relationship (Figure 1b) to enable accurate classification. This step was accomplished by placing the 19 landmarks in a list sorted in ascending order through the x-axis. After sorting, if there were values on the x-axis that were very close to each other (distance of 5 pixels), the Y-axis was used to eliminate the doubt. This approach ensured the standard positional order of the 19 landmarks portrayed in Figure 1b, which does not coincide with the order entered manually (Figure 1a).

The positional precision of the 19 landmarks was assessed by comparing the landmark patterns obtained from the U-Net output with the manually annotated landmarks (Figure 1). The precision of a given computed landmark was calculated inside the Procrustes space by comparing it with its corresponding ground-true landmark. The Euclidian distances between the landmarks' pairs were then summed up and normalized. The maximum precision of 1 was achieved when the sum of the distances was equal to zero.

The information content carried by each landmark was calculated using the well-known information gain ratio criterion [41]. This criterion measures the uncertainty in how the data are separated based on a specific feature. The value of the information gain ratio was calculated for each feature, allowing assessment of their individual impact on subspecies classification.

2.3.3. Classification

The classification stage involved a Procrustes normalization (Figure 6), which allowed the ignoring of transposition, mirroring, rotation, and scale of the different landmark patterns [42]. The Generalized Procrustes Analysis (GPA) computed an invariant mean of landmark patterns generated from the training dataset. The linear projection of the new landmark pattern in the invariant space creates a pattern of information where the invariation rules are achieved.

Forewings classification was performed using a support vector machine (SVM) [43] by employing the Scikit-Learn [44] library. The SVM presents a high generalization capacity and is well-suited for small datasets, as is the case of dataset 2, used for training the classifier. Loading SVM with the geometric features invariant to translation, rotation, and scale was critical to achieving good classification results. After classification, the results were visualized and saved.

3. Results and Discussion

3.1. Wing Detector

The wing detector is essential to extract several wings from one image and normalize the image aspect, allowing for the stable loading of the CNN inputs. Several models of object detectors were trained on dataset 1 using transfer learning [45], which stabilized the training from the initial iterations. The criteria for model selection were accuracy and speed, which are both shown for each tested model in Table 1.

Table 1. Wing detector models with corresponding mAP and speed.

Model	Coco mAP@0.5	Images per Second
SSD MobileNet v1 FPN coco	0.975	24
Faster R-CNN NAS	0.950	0.6
Faster R-CNN Inception Resnet v2 Atrous Coco	0.950	1.8
YoloV3	0.900	18

While all trained models achieved mAP levels > 0.9, SSD MobileNet v1 FPN coco [34] was revealed to be the most accurate (0.975) and at the same time the fastest model (24 images per second). This was an expected result because MobileNet presents a simpler architecture that requires fewer images to be trained than other models. Moreover, its simpler architecture can be associated with its generalization capability, reducing the "curse of dimensionality". This generalization capability ensures correct wing detection even when the quality of wing images is poor. This characteristic of MobileNet will be critical for future users of the pipeline developed here as the images under analysis will likely exhibit a wide range of visual diversity (e.g., specular reflections, varying contrast) and artifacts (Figure 3).

Given the demonstrated superior performance, MobileNet was chosen for the final implementation of the system. This CNN-based wing detector proved to be essential for assuring the uniformity of the wing shape pattern at the entrance of the U-Net, allowing the extraction of several wings from the same image even when they exhibited different orientations.

3.2. Size of Synthetic Landmarks for Training

The size of the synthetic landmarks, denoted as circles, directly influenced the effectiveness of the U-Net at the training stage. Circles with a radius varying from one to five pixels were examined. The most appropriate circles had three and four pixels, with the latter performing better than the former (Table 2). Circles with a radius smaller than three pixels were too small and were virtually ignored by the neural network. On the other hand, circles with a radius larger than four pixels were too large, compromising the positional accuracy of the detected points. Furthermore, in several cases, the large radius size led to the merging of different landmark regions into a single one, preventing the detection of all 19 landmarks.

Table 2. Functional performance of the U-Net landmarks segmenter in detecting exactly 19 landmarks when variating a set of factors (Radius, Altered images, Dust, Kernel, Weights).

	Radius	Altered Images	Dust	Kernel	Weights	Accuracy (%)
1	2	No	No	3 × 3	No	68.1
2	3	No	No	3 × 3	No	70.4
3	3	Yes	No	3 × 3	No	76.3
4	3	Yes	No	5 × 5	No	78.7
5	3	Yes	Yes	5 × 5	No	81.9
6	3	Yes	Yes	5 × 5	Yes	88.2
7	4	Yes	Yes	5 × 5	Yes	91.8
8	4	Yes	Yes	7 × 7	Yes	83.1

3.3. U-Net Optimization

The U-Net was revealed to be an essential solution for detecting the 19 landmarks. While we tried classical object detectors (e.g., Faster R-CNN, YoloV3), these were incapable of returning good precision in finding the correct position of the landmarks. Furthermore, as a segmenter, the U-Net was designed to make pixel-oriented annotations, and that precision was well-suited to our system needs.

Several parameter changes were made to the U-Net to improve its functional performance. The original U-Net implementation uses convolution layer kernels of 3 × 3 elements, but that version had a greater focus on detecting edges [30]. Kernels of sizes 3 × 3, 5 × 5, and 7 × 7 were tested. Larger kernels could deliver better results due to a broader view of the landmark context. The best segmentation quality was obtained for size 5 × 5 (Table 2). Although no new layers were added to the U-Net, changing the size of the kernels had a substantial impact on memory and speed performance. For instance, increasing the kernels from 3 × 3 to 5 × 5 made the neural network run three times slower.

The training results of the U-Net were further optimized by implementing different approaches, including (i) early stopping [46], (ii) reduction on plateau [39], and (iii) a weight loss function. The early stopping ends the model's training when the validation loss starts growing relative to the previous training iterations. By interrupting the process soon after the model converged (generalization state), it was possible to avoid overfitting. The reduction on plateau decreases the training learning rate when the loss figure stops improving. Implementation of these two approaches allowed a better level of convergence and landmark segmentation. The weight loss function, which allows differentiating the importance of different training classes, was implemented to emphasize the landmarks relative to the background. To illustrate the problem, a 400 × 400 image has 160,000 pixels and contains 19 circles with a radius of four pixels. Therefore, only 955 of those pixels constitute landmarks to be segmented, representing <1% of the total pixels. Accordingly, weighting the two classes by using the weight loss function was crucial for achieving segmentation success. Several weight configurations were tested. The one that produced the best results kept the weights of the background class at one and increased by 50 the weight of the landmark class.

3.4. Evaluation of Landmarks Segmentation

In this section, the parametrization and the results that produce a suitable generalization regarding the landmarks segmentation are analyzed regarding (i) the segmentation capabilities of the U-Net (ensuring that the 19 landmarks are detected), (ii) the positional precision of the 19 landmarks, and (iii) the processing robustness in dealing with image dust and noise.

Table 2 shows the capability of the U-Net in annotating exactly the 19 landmarks, using several configurations in the validation dataset. Radius represents the radius in pixels of the landmarks used during the training. Altered images indicate whether the images were rotated or displaced during the process of data augmentation in the training dataset. Dust corresponds to artifacts and noise artificially added to the images. Kernel corresponds to the U-Net kernel size. Weights represent the use of weight loss function during training. Accuracy indicates the success of the U-Net model in detecting all 19 landmarks. A summary of a larger grid of parameter combinations is shown in Table 2. Lines 1 and 2 show an improvement in the U-Net performance when the circle radius is increased from 2 to 3. Lines 2 and 3 show the improvement resulting from data augmentation regarding affine geometric transformations (Altered Images). Lines 3 and 4 reveal the benefit of using a kernel size of 5 × 5 versus 3 × 3. The improvement achieved from adding noise is evidenced when comparing lines 4 and 5. A comparison between lines 5 and 6 underscores the benefit of using the weight loss function to train the U-Net. Lines 6, 7 and 8 show that the radius of size four outperforms the radius of size three and that the kernel size of 5 × 5 outperforms 7 × 7. When combining the best parameters, the accuracy of landmark segmentation reaches 91.8%.

The U-Net model learned how to handle problems of dust and angle modifications, as illustrated for several examples in Figure 12a,b. However, when the images were excessively corrupted by visual artifacts, some landmarks went undetected (Figure 12c). In addition to distorting the detection of landmarks, artifacts in the image may also create false positives.

The positional precision of the 19 computed landmarks was assessed on the testing dataset by comparing the U-Net-generated landmarks with the manually annotated landmarks (Figure 1). Obtaining a high-precision landmarks segmenter is critical for honey bee classification when using wing geometric morphometrics. This is because the locations of the 19 landmarks are subspecies specific, as illustrated in Figure 1c,d for two genetically close subspecies [5], and any positioning error will impact classification accuracy. As shown in Table 3, precision varied among landmarks, with the lowest value obtained for landmark 9 (0.900) and the highest for landmark 19 (0.975). Curiously, landmark 19 is the single one located outside of a vein junction. The average precision obtained for the 19 computed landmarks was 0.943 ± 0.020.

While all 19 landmarks are used by the subspecies classifier, the information content carried by each one is variable. Table 4 shows the information gain ratio obtained for each x and y input feature associated with the 19 landmark coordinates. The most important features (information gain ratio > 0.2) for subspecies classification were 13-x, 17-y, 15-y, 13-y, and 8-y, which exhibited a precision ranging from 0.933 (landmark 15) to 0.958 (landmark 17; Table 3). Interestingly, these four landmarks are implicated in the calculations of the cubital index, hantel index, and discoidal shift angle, which are used by many queen breeders engaged in *A. m. mellifera* conservation [14]. This finding suggests that the pipeline is using, with good precision, the features that are well-known as correlated to the subspecies. Moreover, the use of wing landmarks in honey bee classification enables a very affirmative criterion for automatically excluding images, as classification is aborted when the number of extracted landmarks is different from 19.

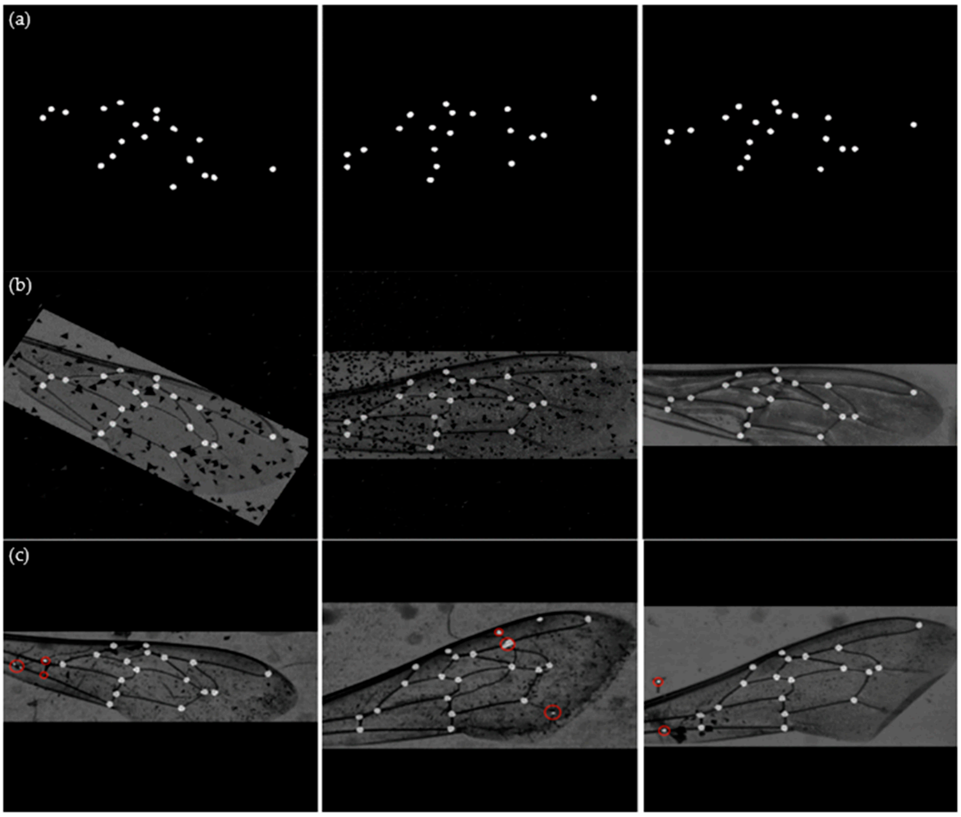


Figure 12. (**a**) The output masks of the U-Net after training. (**b**) Masks superimposed on the corresponding original images. (**c**) Examples of errors, denoted by the red circles, in low-quality images.

Table 3. Positional precision of the 19 detected landmarks. The landmarks correspond in nomination to the sequence of annotated numbers shown in Figure 1b.

Landmark	Precision	Landmark	Precision
1	0.968	11	0.924
2	0.970	12	0.926
3	0.963	13	0.937
4	0.954	14	0.945
5	0.911	15	0.933
6	0.932	16	0.931
7	0.939	17	0.958
8	0.950	18	0.962
9	0.900	19	0.975
10	0.937	Average ± SD	0.943 ± 0.020

In summary, the essays carried out here, using complex images from dataset 2 and images from external datasets (data not shown), revealed a high capacity of the U-Net for generalization in detecting the landmarks. The U-Net model was not only capable of individually detecting each landmark but also capable of relating the landmarks in a pattern, enabling inference of missing landmarks. Furthermore, the U-Net showed precision capability and robustness in dealing with new and visually corrupted images.

Table 4. Input feature regarding the 19 landmarks (x, y) coordinates ranked by information content.

Landmark (x or y) Component	Information Gain Ratio	Landmark (x or y) Component	Information Gain Ratio
13 (x)	0.267	18 (x)	0.095
17 (y)	0.248	17 (x)	0.088
15 (y)	0.240	6 (x)	0.087
13 (y)	0.236	1 (y)	0.087
8 (y)	0.203	7 (x)	0.071
15 (x)	0.166	3 (y)	0.068
4 (y)	0.150	2 (y)	0.062
10 (x)	0.147	12 (x)	0.053
14 (x)	0.146	6 (y)	0.053
3 (x)	0.141	11 (y)	0.051
9 (y)	0.137	19 (y)	0.043
5 (y)	0.135	11 (x)	0.042
16 (x)	0.134	4 (x)	0.040
12 (y)	0.131	9 (x)	0.039
5 (x)	0.131	7 (y)	0.025
10 (y)	0.130	1 (y)	0.020
18 (y)	0.127	16 (y)	0.019
14 (y)	0.109	2 (x)	0.017
8- (x)	0.107	1 (x)	0.016

This study provides new insights into machine learning research by offering an alternative solution for problems demanding annotation of landmarks with a high level of precision (e.g., reference points in aerial images, facial points used for biometric recognition, markers at body joints employed for motion acquisition, anatomical landmarks in medical images, fingerprint minutiae for person recognition). To the best of our knowledge, our approach is unique in the usage of the U-Net for the segmentation of landmarks. Remarkably, the U-Net allowed a precise and robust (regarding the noise) segmentation process when using a kernel of 5 × 5 and landmark masks with a radius of 4 pixels. Furthermore, employing a weight cost function was revealed to be essential to emphasize the landmarks relative to the background, further contributing to the success of the final solution.

3.5. Classification

The classifier receives the image features already treated and simplified in the previous stages of the pipeline (Figure 6). For instance, the U-Net translates the wing image into 19 stable data points, and the Procrustes method ensures positional independence (translation, scale, rotation), to avoid a training dataset with those variations and allow future classification of images with a range of variations.

The images were classified using SVM [43], a model that does not need a large hyperparametrization, simplifying the training phase. The validation dataset was employed to find the SVM configuration that returned the most accurate classification. The SVM C factor was 30, the gamma value was 0.17, and the best kernel was the RBF (Radial Base Function). The model input consisted of the 38 values (x, y landmarks coordinates) normalized by the Procrustes method. The model output codified the subspecies under classification.

The final SVM configuration was able to classify the forewings with 86.6% ± 6.9 average (±SD) accuracy across the 26 subspecies represented in the testing dataset (Table 5), despite the relatively small number of images used in the classification training and the low quality of many of them (Figure 3). Nawrocka and colleagues [8] found a similar average accuracy (88.4%) for a similar manually annotated dataset (25 instead of 26 subspecies of the Oberursel collection), further validating the performance of the pipeline developed herein. These authors [8] also employed the geometric morphometrics method, although they classified the 25 subspecies using the linear canonical variate analysis as opposed to the non-linear SVM approach. In another study, Da Silva and colleagues [7] used manually annotated landmarks on the Oberursel wing images to compare seven different classifiers, including SVM. Unexpectedly, the authors reported poor performance of the SVM classifier (60.04%), as compared to this study, with the best classifier, Naïve Bayes, achieving only 65.15% accuracy in cross-validation.

Table 5. Classification accuracy obtained for the 26 *A. mellifera* subspecies.

Lineage	Average (± SD) Accuracy (%)	*A. mellifera* Subspecies	Accuracy (%)	*A. mellifera* Subspecies	Accuracy (%)
A	75.0 ± 7.1	*adansonii*	72.4	*monticola*	80.0
		capensis	77.9	*ruttneri*	66.1
		intermissa	75.2	*sahariensis*	82.7
		lamarckii	69.3	*scutellata*	71.8
		litorea	60.9	*unicolor*	87.9
		major	78.8	*jemenitica*	77.2
M	92.2 ± 3.3	*iberiensis*	88.7	*mellifera*	95.3
C	88.1 ± 7.3	*carnica*	89.6	*macedonica*	85.7
		cecropia	96.4	*siciliana*	75.3
		ligustica	93.1		
O	91.2 ± 4.1	*adami*	90.3	*cypria*	82.7
		anatoliaca	93.9	*meda*	94.5
		armeniaca	92.1	*syriaca*	95.3
		caucasia	88.2		

At the subspecies level, the lowest accuracy was observed for the African *A. m. litorea* (60.9%) and the highest for the eastern European *A. m. cecropia* (96.4%; Table 5). At the lineage level, the lowest average accuracy was observed for the African (75.0%) and the highest for the western European (92.2%; Table 5). The poorer performance of the classifier for African subspecies is consistent with the findings of [8] and is explained by the closer morphological proximity among subspecies from central and southern Africa. Accordingly, when the subspecies of African ancestry were excluded from the analysis, classification accuracy increased up to 90.5% ± 1.7.

The performance of the pipeline was further improved (95.8% accuracy; Table 6) by training another model with only five subspecies chosen for their commercial value (*A. m. ligustica*, *A. m. carnica* and *A. m. caucasia*) or conservation status (*A. m. mellifera* and its sister *A. m. iberiensis*) [47]. Except for *A. m. iberiensis*, the remaining four subspecies were amongst the best represented in dataset 2, used for classification training, with *A. m. carnica* having the largest number of wing images (n = 150). While excluding subspecies of African ancestry led to greater classification accuracy (this study and [8]), it is possible that the improved performance allowed by the five subspecies model is also related to the larger sample size used during training. Although not directly comparable with our system, the end-to-end solution developed from the entire wing (as opposed to the landmarks) by De Nart and colleagues [9] achieved accuracy values of 99% in cross-validation from a much larger (n = 9887) proprietary wing image dataset representing one hybrid and seven subspecies (*A. m. ligustica*, *A. m. carnica*, *A. m. caucasia*, *A. m. anatoliaca*, *A. m. siciliana*, *A. m. iberiensis*, *A. m. mellifera*) using CNN.

Training the system with a larger number of wing images per subspecies may lead to a more accurate classification. Yet, this effort can only be achieved with wing images obtained from newly collected specimens outside of the Oberursel collection, as was carried out by De Nart and colleagues [9]. The problem is that this effort involves the identification of the specimens using the full set of morphological traits, which is a time-consuming endeavor requiring expert knowledge that is not always available [5]. Moreover, the classification pipeline developed herein was based on the explicit choice of training the system with the original wing collection used by Ruttner [1] in the delineation of the *A. mellifera* subspecies, in spite of the low number of wing images and their often-low quality. Nonetheless, if needed, the system can be trained with new examples to improve classification rates on all types of wing images.

Table 6. Classification accuracy obtained for five *A. mellifera* subspecies.

***A. mellifera* Subspecies**	**Accuracy (%)**
carnica	98.9
caucasia	97.7
iberiensis	91.1
ligustica	96.4
mellifera	95.0
Average ± SD	95.8 ± 2.7

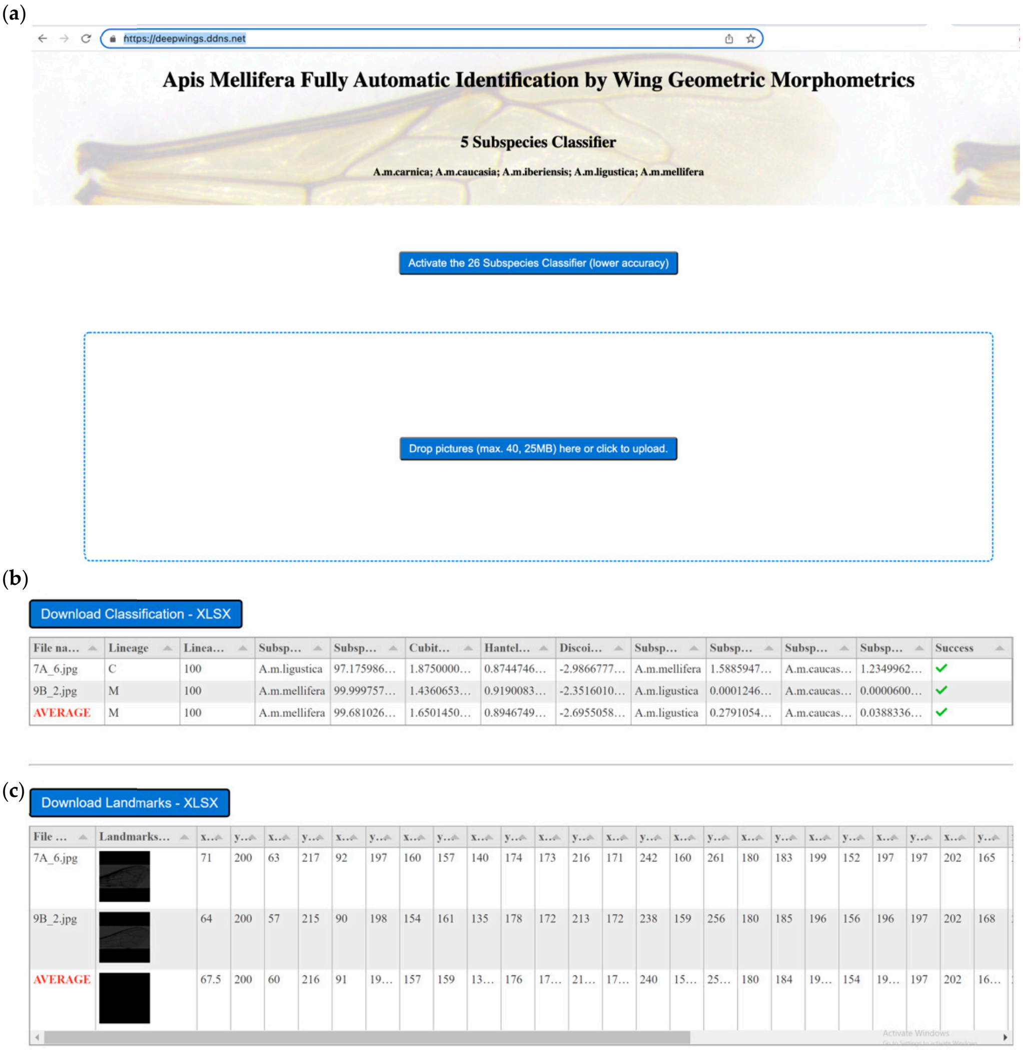
3.6. Computational Cost Analysis

The final implementation of the system, using a CNN MobileNet, a U-Net, and a SVM, presented a good speed performance, requiring 14 s to process 10 images, which is the minimum number recommended for colony-level identification [5]. The computational machine was based on an i5-9400@2.9 GHz six cores CPU and 16 GBytes of main memory. As the implemented solution is based on threads, the overall speed performance could be increased by using a CPU with a larger number of cores. The CPU clock also directly influences the speed performance because the main processing demands mathematical vector operations that are correlated to the vectorial instruction speed. The deep learning framework allows for running the code on a GPU to achieve an overall velocity of about 15 times as compared to the CPU.

3.7. DeepWings© as a Web Service

The pipeline developed herein for honey bee subspecies identification was registered as software named DeepWings© (*Registo de obra* n.° 3214/2019, *Inspeção-Geral das Atividades Culturais*). DeepWings© is implemented as a free Web service available at the URL https://deepwings.ddns.net, accessed on 26 April 2022. The Flask framework was used to program the Web service, which was constructed with parallel programming based on threads.

DeepWings© is a user-friendly software that only requires dragging wing images into a file drop zone (Figure 13a). After image processing, a table containing the probabilities of the top three subspecies is built on the Web page (Figure 13b). In addition to classification, DeepWings© computes several wing geometric parameters and the coordinates of the landmarks (Figure 13c). Classification probabilities, geometric parameters (cubital index, hantel index, discoidal shift angle) and landmark coordinates can be downloaded as excel files, facilitating data storage for further analysis or alternative applications. For instance, the coordinates of the landmarks can be used directly by other identification software, such as MorphoJ [48] or IdentiFly [26], or to calculate angles and lengths required by other methods such as the DAWINO or classical wing morphometry [5]. DeepWings© can therefore be used in conjunction with measurements of other body traits (e.g., pilosity, pigmentation, proboscis and lengths) for purposes requiring more intensive methods, such as identification of new subspecies [2–5]. However, more importantly, beekeepers and queen breeders now have a friendly tool for identifying their colonies for conservation or commercial purposes.


(a)
https://deepwings.ddns.net
Activate the 26 Subspecies Classifier (lower accuracy)
Drop pictures (max. 40, 25MB) here or click to upload.
(b)
Download Classification - XLSX
(c)
Download Landmarks - XLSX
Activate Windows

Apis Mellifera Fully Automatic Identification by Wing Geometric Morphometrics

5 Subspecies Classifier

A.m.carnica; A.m.caucasia; A.m.iberiensis; A.m.ligustica; A.m.mellifera

File na…	Lineage	Linea…	Subsp…	Subsp…	Cubit…	Hantel…	Discoi…	Subsp…	Subsp…	Subsp…	Subsp…	Success
7A_6.jpg	C	100	A.m.ligustica	97.175986…	1.8750000…	0.8744746…	-2.9866777…	A.m.mellifera	1.5885947…	A.m.caucas…	1.2349962…	✓
9B_2.jpg	M	100	A.m.mellifera	99.999757…	1.4360653…	0.9190083…	-2.3516010…	A.m.ligustica	0.0001246…	A.m.caucas…	0.0000600…	✓
AVERAGE	M	100	A.m.mellifera	99.681026…	1.6501450…	0.8946749…	-2.6955058…	A.m.ligustica	0.2791054…	A.m.caucas…	0.0388336…	✓

File …	Landmarks…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…	x…	y…
7A_6.jpg		71	200	63	217	92	197	160	157	140	174	173	216	171	242	160	261	180	183	199	152	197	197	202	165
9B_2.jpg		64	200	57	215	90	198	154	161	135	178	172	213	172	238	159	256	180	185	196	156	196	197	202	168
AVERAGE		67.5	200	60	216	91	19…	157	159	13…	176	17…	21…	17…	240	15…	25…	180	184	19…	154	19…	197	202	16…

Figure 13. Deep Wings© running as a Web service. (**a**) Interface showing the drop file zone where wing images are uploaded. Interface showing the output tables containing the (**b**) subspecies classification and (**c**) landmark coordinates.

It is increasingly recognized that sustainable beekeeping requires the use of native subspecies, as they are better adapted to local environments and show superior performance when compared with exotic subspecies [49]. The problem is that the genetic integrity of many honey bee subspecies is threatened after many generations of importation of exotic queens [47]. *A. m. mellifera* is the best example of such a situation, as in large tracts of its native distribution, this subspecies is severely introgressed or is even on the brink of extinction. This has led to increasing demand for native subspecies, especially for *A. m. mellifera*, and consequently for identification tools that can be used to certify the origin of the queens. Such certification is required for beekeepers for (i) moving their colonies to conservation areas, (ii) monitoring the efficiency of isolated mating stations, and (iii) receiving subventions according to local or European legislation. Additionally, beekeepers may be able to increase the market value of their stock by certifying the queens' origin. Highly accurate subspecies identification implies measuring 36 morphological traits [5] or genotyping genome-wide molecular markers [28]. However, these methods are not accessible to most beekeepers and queen breeders, as the former is too laborious and time-consuming and the latter is still too expensive. DeepWings© offers an alternative solution for the above applications, which often do not require the accuracy levels provided by those morphological or molecular methods.

4. Conclusions

The software DeepWings© developed here for the identification of honey bees by wing geometric morphometrics, combines a CNN as the wing detector, a deep learning U-Net as the landmarks segmenter, and a SVM as the subspecies classifier. This multi-step solution was revealed to be better suited than an end-to-end solution for our problem for three main reasons. First, by using models of low complexity (U-Net and SVM), it was possible to train the system with the small wing collection used by Ruttner [1] to identify the honey bee subspecies. Second, it allowed the neural model to search for the features (landmarks) that are used by the gold-standard method in honey bee classification based on wings: geometric morphometrics. Third, by introducing the Procrustes method between the landmark detection step and the classification step, the classifier could be trained with greater robustness, as the wing invariances (translation, rotation, and scale) do not need to be learned. Despite the apparent complexity, the system showed good speed performance, requiring 14 s to process 10 images, which is the minimum number recommended for colony-level identification [5].

While there is another wing geometric morphometrics tool (IdentiFly) available for honey bee classification, DeepWings© is the first to do so in a fully automated manner. The scientific novelty and greatest contribution of our solution to honey bee classification is related to the capability of DeepWings© to segment the wings from images containing varying artifacts, varying numbers of wings, and wings with different orientations, and then segment the landmarks with high precision.

Since the wing shape patterns differ slightly among subspecies, particularly when they belong to the same lineage, it became critical to use a segmenter that would allow high precision in detecting the 19 landmarks. This goal was achieved by using the U-Net, which showed good performance even when the images were very noisy.

The use of SVM in the classification was revealed to be a good solution with suitable generalization, given the weak separation among subspecies (particularly those of African ancestry) and the low number of images available for classification training (26 subspecies represented by only 1864 images). Despite these limitations, classification accuracy was 86.6% for the 26 subspecies and increased to 95.8% when the classifier was trained with only five subspecies. Higher accuracy would have been expected had the dataset used in the classification training been larger. However, despite the low number of wing images contained in the Oberursel collection and their often-low quality, we explicitly chose to train our system with the dataset that was originally used by Ruttner [1] in the delineation of the *A. mellifera* subspecies.

DeepWings© is available as free software for use in colony identification for multiple purposes (e.g., monitoring isolated mating stations, selection of queens in conservation apiaries) by beekeepers, queen breeders, and even scientists. In addition to classification, DeepWings© provides the coordinates of the 19 landmarks, and these can be processed by other software, facilitating data exchange between different scientific studies and research teams.

Author Contributions: Conceptualization, P.J.R. and M.A.P.; methodology, P.J.R. and W.G.; software, P.J.R. and W.G.; validation, P.J.R. and W.G. and M.A.P.; formal analysis, P.J.R. and W.G.; investigation, W.G.; resources, P.J.R. and M.A.P.; data curation, W.G.; writing—original draft preparation, P.J.R., M.A.P. and W.G.; writing—review and editing, P.J.R. and M.A.P.; supervision, P.J.R.; project administration, P.J.R.; funding acquisition, M.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support was provided through the program COMPETE 2020—POCI (Programa Operacional para a Competividade e Internacionalização) and by Portuguese funds through FCT (Fundação para a Ciência e a Tecnologia) in the framework of the project BeeHappy (POCI-01-0145-FEDER-029871). FCT provided financial support by national funds (FCT/MCTES) to CIMO (UIDB/00690/2020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: https://github.com/walterBSG/Beeapp-landmark-detection, accessed on 26 April 2022.

Acknowledgments: We are indebted to Helena Ferreira for manually annotating wings of dataset 1 and to Tiago M. Francoy for providing the wing images of dataset 2.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ruttner, F. *Biogeography and Taxonomy of Honeybees*; Springer: Berlin/Heidelberg, Germany, 1988; p. 284.
2. Sheppard, W.S.; Meixner, M.D. *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie* **2003**, *34*, 367–375. [CrossRef]
3. Meixner, M.D.; Leta, M.A.; Koeniger, N.; Fuchs, S. The honey bees of Ethiopia represent a new subspecies of *Apis mellifera-Apis mellifera simensis* n. ssp. *Apidologie* **2011**, *42*, 425–437. [CrossRef]
4. Chen, C.; Liu, Z.G.; Pan, Q.; Chen, X.; Wang, H.H.; Guo, H.K.; Liu, S.D.; Lu, H.F.; Tian, S.L.; Li, R.Q.; et al. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. *Mol. Biol. Evol.* **2016**, *33*, 1337–1348. [CrossRef] [PubMed]
5. Meixner, M.D.; Pinto, M.A.; Bouga, M.; Kryger, P.; Ivanova, E.; Fuchs, S. Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *J. Apic. Res.* **2013**, *52*, 1–28. [CrossRef]
6. Tofilski, A. DrawWing, a program for numerical description of insect wings. *J. Insect Sci.* **2004**, *4*, 17. [CrossRef]
7. Da Silva, F.L.; Sella, M.L.G.; Francoy, T.M.; Costa, A.H.R. Evaluating classification and feature selection techniques for honeybee subspecies identification using wing images. *Comput. Electron. Agric.* **2015**, *114*, 68–77. [CrossRef]
8. Nawrocka, A.; Kandemir, I.; Fuchs, S.; Tofilski, A. Computer software for identification of honey bee subspecies and evolutionary lineages. *Apidologie* **2018**, *49*, 172–184. [CrossRef]
9. De Nart, D.; Costa, C.; di Prisco, G.; Carpana, E. Image recognition using convolutional neural networks for classification of honey bee subspecies. *Apidologie* **2022**, *53*, 5. [CrossRef]
10. Bookstein, F.L. *Morphometric Tools for Landmark Data: Geometry and Biology*; Cambridge University Press: Cambridge, UK, 1992.
11. Francoy, T.M.; Wittmann, D.; Drauschke, M.; Muller, S.; Steinhage, V.; Bezerra-Laure, M.A.F.; De Jong, D.; Goncalves, L.S. Identification of Africanized honey bees through wing morphometrics: Two fast and efficient procedures. *Apidologie* **2008**, *39*, 488–494. [CrossRef]
12. Kandemir, I.; Ozkan, A.; Fuchs, S. Reevaluation of honeybee (*Apis mellifera*) microtaxonomy: A geometric morphometric approach. *Apidologie* **2011**, *42*, 618–627. [CrossRef]
13. Prabucki, J.S.; Samborski, J.; Chuda-Mickiewicz, B. The use of three taxonomic characters for race identification of Middle European bee. *J. Apic. Sci.* **2002**, *46*, 41–47.
14. Bouga, M.; Alaux, C.; Bienkowska, M.; Buchler, R.; Carreck, N.L.; Cauia, E.; Chlebo, R.; Dahle, B.; Dall'Olio, R.; De la Rua, P.; et al. A review of methods for discrimination of honey bee populations as applied to European beekeeping. *J. Apic. Res.* **2011**, *50*, 51–84. [CrossRef]
15. Bonatti, V.; Simoes, Z.L.P.; Franco, F.F.; Francoy, T.M. Evidence of at least two evolutionary lineages in *Melipona subnitida* (Apidae, Meliponini) suggested by mtDNA variability and geometric morphometrics of forewings. *Naturwissenschaften* **2014**, *101*, 17–24. [CrossRef] [PubMed]
16. Francoy, T.M.; Franco, F.D.; Roubik, D.W. Integrated landmark and outline-based morphometric methods efficiently distinguish species of *Euglossa* (Hymenoptera, Apidae, Euglossini). *Apidologie* **2012**, *43*, 609–617. [CrossRef]
17. Francoy, T.M.; Bonatti, V.; Viraktamath, S.; Rajankar, B.R. Wing morphometrics indicates the existence of two distinct phenotypic clusters within population of *Tetragonula iridipennis* (Apidae: Meliponini) from India. *Insectes Sociaux* **2016**, *63*, 109–115. [CrossRef]
18. Costa, C.P.; Machado, C.A.S.; Santiago, W.M.S.; Dallacqua, R.P.; Garófalo, C.A.; Francoy, T.M. Biome variation, not distance between populations, explains morphological variability in the orchid bee *Eulaema nigrita* (Hymenoptera, Apidae, Euglossini). *Apidologie* **2020**, *51*, 984–996. [CrossRef]
19. Rebelo, A.R.; Fagundes, J.M.G.; Digiampietri, L.A.; Francoy, T.M.; Biscaro, H.H. A fully automatic classification of bee species from wing images. *Apidologie* **2021**, *52*, 1060–1074. [CrossRef]
20. Francoy, T.M.; Prado, P.R.R.; Gonçalves, L.S.; Costa, L.d.F.; Jong, D.D. Morphometric differences in a single wing cell can discriminate *Apis mellifera* racial types. *Apidologie* **2006**, *37*, 91–97. [CrossRef]
21. Evin, A.; Baylac, M.; Ruedi, M.; Mucedda, M.; Pons, J.-M. Taxonomy, skull diversity and evolution in a species complex of *Myotis* (Chiroptera: Vespertilionidae): A geometric morphometric appraisal. *Biol. J. Linn. Soc.* **2008**, *95*, 529–538. [CrossRef]
22. Tofilski, A. Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie* **2008**, *39*, 558–563. [CrossRef]
23. 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* **2011**, *42*, 150–161. [CrossRef]

24. Oleksa, A.; Tofilski, A. Wing geometric morphometrics and microsatellite analysis provide similar discrimination of honey bee subspecies. *Apidologie* **2015**, *46*, 49–60. [CrossRef]
25. Ferreira, H.; Henriques, D.; Neves, C.J.; Machado, C.A.S.; Azevedo, J.C.; Francoy, T.M.; Pinto, M.A. Historical and contemporaneous human-mediated processes left a strong genetic signature on honey bee populations from the Macaronesian archipelago of the Azores. *Apidologie* **2020**, *51*, 316–328. [CrossRef]
26. Tofilski, A. IdentiFly Software, Version 0.31. Available online: http://drawwing.org/identifly (accessed on 4 April 2022).
27. Henriques, D.; Browne, K.A.; Barnett, M.W.; Parejo, M.; Kryger, P.; Freeman, T.C.; Muñoz, I.; Garnery, L.; Highet, F.; Jonhston, J.S.; et al. High sample throughput genotyping for estimating C-lineage introgression in the dark honeybee: An accurate and cost-effective SNP-based tool. *Sci. Rep.* **2018**, *8*, 8552. [CrossRef] [PubMed]
28. Momeni, J.; Parejo, M.; Nielsen, R.O.; Langa, J.; Montes, I.; Papoutsis, L.; Farajzadeh, L.; Bendixen, C.; Căuia, E.; Charrière, J.-D.; et al. Authoritative subspecies diagnosis tool for European honey bees based on ancestry informative SNPs. *BMC Genom.* **2021**, *22*, 101. [CrossRef]
29. Chavez-Galarza, J.; Henriques, D.; Johnston, J.S.; Carneiro, M.; Rufino, J.; Patton, J.C.; Pinto, M.A. Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: Maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. *Mol. Ecol.* **2015**, *24*, 2973–2992. [CrossRef]
30. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1, Lake Tahoe, Nevada, 3–6 December 2012; pp. 1097–1105.
32. Bradski, G.R.; Pisarevsky, V. Intel's computer vision library: Applications in calibration, stereo, segmentation, tracking, gesture, face and object recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000), Hilton Head, SC, USA, 15 June 2000; pp. 796–797.
33. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
34. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
35. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
36. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First Aaai Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
37. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
38. Kaehler, A.; Bradski, G.R. *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*, 1st ed.; O'Reilly Media: Sebastopol, CA, USA, 2017; 990 p.
39. Chollet, F. *Keras: The Python Deep Learning Library*; Astrophysics Source Code Library: Houghton, MI, USA, 2018; p. ascl:1806.1022.
40. Mudrová, M.; Procházka, A. Principal component analysis in image processing. In Proceedings of the MATLAB Technical Computing Conference, Prague, Czech Republic, 4–8 July 2005.
41. Quinlan, J.R. Simplifying decision trees. *Int. J. Man-Mach. Stud.* **1987**, *27*, 221–234. [CrossRef]
42. Gower, J.C. Generalized procrustes analysis. *Psychometrika* **1975**, *40*, 33–51. [CrossRef]
43. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
44. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
45. Pan, S.J.; Yang, Q.A. A Survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
46. Prechelt, L. Automatic early stopping using cross validation: Quantifying the criteria. *Neural Netw.* **1998**, *11*, 761–767. [CrossRef]
47. De la Rua, P.; Jaffe, R.; Dall'Olio, R.; Munoz, I.; Serrano, J. Biodiversity, conservation and current threats to European honeybees. *Apidologie* **2009**, *40*, 263–284. [CrossRef]
48. Klingenberg, C.P. MorphoJ: An integrated software package for geometric morphometrics. *Mol. Ecol. Resour.* **2011**, *11*, 353–357. [CrossRef]
49. Büchler, R.; Costa, C.; Hatjina, F.; Andonov, S.; Meixner, M.D.; le Conte, Y.; Uzunov, A.; Berg, S.; Bienkowska, M.; Bouga, M.; et al. The influence of genetic origin and its interaction with environmental effects on the survival of *Apis mellifera* L. colonies in Europe. *J. Apic. Res.* **2014**, *53*, 205–214. [CrossRef]