Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 273 (2022) 121058

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Reflectance spectroscopy and machine learning as a tool for the categorization of twin species based on the example of the *Diachrysia* genus

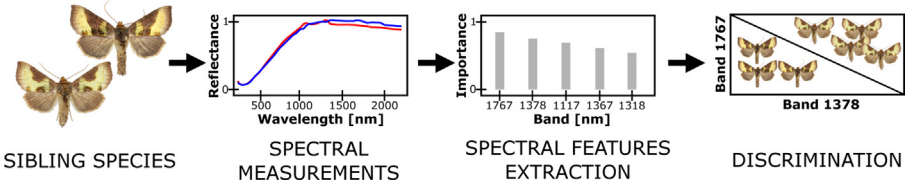
Krzysztof Dyba [a], Roman Wąsala [b], Jan Piekarczyk [c], Elżbieta Gabała [d], Magdalena Gawlak [d], Jarosław Jasiewicz [a,*], Henryk Ratajkiewicz [b,*]

[a] Institute of Geoecology and Geoinformation, Adam Mickiewicz University in Poznań, Poland
[b] Department of Entomology and Environment Protection, Poznań University of Life Sciences, Poland
[c] Institute of Physical Geography and Environmental Planning, Adam Mickiewicz University in Poznań, Poland
[d] Research Centre of Quarantine, Invasive and Genetically Modified Organisms, Institute of Plant Protection – National Research Institute, Poland

HIGHLIGHTS

- Spectroscopy and machine learning support taxonomic research.
- In reflectance spectra only small group of wavelengths provides relevant information.
- Features of wing scales are diagnostic for distinguishing between species.
- Morphological differences between moths are confirmed the features of chitin.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 December 2021
Received in revised form 11 February 2022
Accepted 14 February 2022
Available online 17 February 2022

Keywords:
Chemometry
LDA
Random Forest
Lepidoptera
Noctuidae

ABSTRACT

In our work we used noninvasive point reflectance spectroscopy in the range from 400 to 2100 nm coupled with machine learning to study scales on the brown and golden iridescent areas on the dorsal side of the forewing of *Diachrysia chrysitis* and *D. stenochrysis*. We used our approach to distinguish between these species of moths. The basis for the study was a statistically significant collection of 95 specimens identified based on morphological feature and gathered during 23 years in Poland. The numerical part of an experiment included two independent discriminant analyses: stochastic and deterministic. The more sensitive stochastic approach achieved average compliance with the species identification made by entomologists at the level of 99–100%. It demonstrated high stability against the different configurations of training and validation sets, hence strong predictors of *Diachrysia* siblings distinctiveness. Both methods resulted in the same small set of relevant features, where minimal fully discriminating subsets of wavelengths were three for glass scales on the golden area and four for the brown. The differences between species in scales primarily concern their major components and ultrastructure. In melanin-absent glass scales, this is mainly chitin configuration, while in melanin-present brown scales, melanin reveals as an additional factor.

1. Introduction

In recent years, reflectance spectroscopy (RS) has been applied in various disciplines of biology [1,2]. RS provides broad spectra, including the ultraviolet, near, and mid-infrared range. The advan-

* Corresponding authors.
E-mail addresses: jaroslaw.jasiewicz@amu.edu.pl (J. Jasiewicz), henryk.ratajkiewicz@up.poznan.pl (H. Ratajkiewicz).

https://doi.org/10.1016/j.saa.2022.121058
1386-1425/

tage of RS is the potential to quickly obtain a large amount of information about the studied object during a single and non-invasive measurement. Many studies refer to spectral properties of butterflies and moths but most of them have considered only reflectance in human-visible wavelengths (380–780 nm) in which coloration is perceivable. The color of the wing can be pigment-dependent and a separate phenomenon is the structural colors of the wing, which was the subject of in-depth research [3–5]. Few studies considered butterflies and moths wing reflectance properties beyond visible range of spectra [2,6]. In NIR (near infrared) and SWIR (shortwave infrared) ranges information on the structure, water and fat content, and the presence of chitin in the biological tissues is encrypted [7,8]. Nevertheless, only a few chemical compounds are identified, mainly bonds occurring in the structure. Combining information from SWIR and NIR was useful for insect species determination as well as physiological status, age and sex assessment [9–11]. However, to date there is no detailed study on species determination in the Lepidoptera order based on the wing reflectance spectrum in the full VNIR – SWIR spectrum (VNIR - visible near infrared).

There are scientific reports that the wing scales ultrastructure shapes the pattern of the reflectance spectrum even in closely related species [12]. Microscopic studies provide detailed information about the optical properties and structure of butterfly wing scales. For many years, the subject of such research has been the phenomena of light scattering, and diffraction on butterfly scales [4,13]. The wings of butterflies and moths have various structures and types of scales, which requires focusing on relatively small parts of the wing, e.g. eyespot [14]. We suppose that the microscopic spectral characteristics of selected areas of the wing are worth exploring over a wide spectrum and may provide new insight into taxonomic differentiation.

Remote sensing methods produce many features (spectral wavelengths), which means that rigorous numerical analysis is an indispensable step of the analytical procedure. Numerical analysis is necessary with all data sources, including genetics, RGB photos, and detailed hyperspectral measurements of various morphological characteristics such as shape, color, and wing pattern [15–17]. Several papers on entomology involving VIS and NIR spectroscopy applications coupled with numerical methods have been recently reported (see: Johnson and Naiker [18] for a comprehensive list). Kaya and Kayci [19] using neural networks trained on the RGB images and its texture created a classifier which distinguishes 14 species of butterflies with over 90% accuracy. Their collection covered only examples with a relatively different color pattern of wings. However, there exist many lepidopteran species which cause identification problems because wing coloration is not a diagnostic feature. Knowledge about the use of chemometric properties of scale groups in the VIS - SWIR range for distinguishing siblings species is a white card.

The Noctuidae family with about 12 000 known species, representing one of the most species-rich families of Lepidoptera. So, they make up a significant proportion of the world's animal diversity and impact on the environment and human welfare directly. Among them we found some of the most damaging agricultural pests in the world, like the fall armyworm *Spodoptera frugiperda* (Smith, 1797) and Old World bollworm *Helicoverpa armigera* (Hübner, 1808) [20].

Morphological examination of the genital organs and wing pattern is the basis of the classic method of determination of moths species [21] which is routinely performed on captured individuals. New insight into the taxonomy of species was made possible by research on allozyme analysis initiated by Hebert et al. [22] and Svensson et al. [23]. Further studies were continued with the use of the cytochrome oxidase subunit I [24,25].

Usually, genetic characteristics allow the different species of Lepidoptera to be distinguished very effectively. The research done by Janzen et al. [26] and Hausmann et al. [27] show that such separation achieves an efficiency between 97 and 99%. Unfortunately, a small number of species of Lepidoptera were found to share the same barcode sequence with other species but are morphologically distinct. For example, the pair of Phyllodesma ilicifolia (Linnaeus, 1758) - P. tremulifolia (Hübner, 1810) (Lasiocampidae) specimens can be discriminated unambiguously using genitalia and features in wing patterns. For another pair, like Plebejus argyrognomon (Bergsträsser, 1779) – P. idas (Linnaeus, 1761) with low sequence divergence (<1%), discrimination based on external appearance is challenging [27], but both the male and female genitalia are diagnostic [28].

We focused on Diachrysia chrysitis (Linnaeus, 1758) and Diachrysia stenochrysis (Warren, 1913) (Fig. 1), two species common in the Palearctic and other regions of the world. Both species provide important ecosystem services and are important to maintaining biodiversity in the face of increasing anthropopression. At the same time, these so-called siblings species represent one of the most interesting taxonomic puzzles in the Plusiinae subfamily [29]. Discrimination of *D. stenochrysis* and *D. chrysitis* on some individuals based on external appearance is challenging, and the genitalia are also similar. The minimum pairwise divergence of these two twin species is only 0.93% [16].

So far, the current methods used to distinguish *Diachrysia* siblings are one-character taxonomic approaches and therefore sometimes criticized [30]. In such a case, expert classification is also uncertain. Our research focuses on determining the scale of this uncertainty by juxtaposing it with another method based on a set of independent characteristics. In the absence of certain references, the high agreement of two independent classification methods makes it plausible that the morphological differences are not accidental and have deeper determinants.

As a result of expert classification, we divided the individuals into two morphological groups, "*D. chrysitis* type" and "*D. stenochrysis* type" denoted as *D. chrysitis* and *D. stenochrysis* in the rest of the paper. As an independent method, we used reflectance spectroscopy applied to the scales on the wings of moths.

In our approach, we assume that the chemometric features of the scales are not related to the macroscopic features assessed by an expert, namely external genitalia. Therefore, we ask the question: are these groups: D. chrysitis and D. stenochrysis would be confirmed by the chemometric analysis? Machine learning-driven chemometrics can make taxonomic research reproducible, and sustainable, thus more efficient. It has been applied to determine sex, age, color diversity or taxonomical differentiation [31,32]. In this approach, we see great potential for discovering species-specific features. Such features are part of the body's various structural and building materials and are beyond the reach of classical methods.

Surprisingly, no attempt has yet been made to distinguish species complexes of Lepidoptera by studying parts of a wing using the full spectral range and advanced numerical techniques. In our work, we present the experiment results, comparing the detailed expert classification of individuals of moths of *D. chrysitis* and *D. stenochrysis* with results of the chemometric analysis based on the full range of reflectance spectrum between 400 and 2100 nm, obtained under the microscope from two distinct groups of scales on the forewing which have completely different melanin content.

2. Methods

2.1. Specimen sampling and depository

The research was carried out on 95 individuals of *D. chrysitis* (43) and *D. stenochrysis* (52). Male and female imagines were

Fig. 1. The dorsal part of a wing of *Diachrysia chrysitis* and *D. stenochrysis* and their forms with well-marked color pattern. a *D. chrysitis*. b *D. stenochrysis*.

caught in light traps between 1995 and 2018 in Poland. Determinative features and the location and trapping date of *D. chrysitis* and *D. stenochrysis* individuals, are presented in Supplementary Table S1. All sampled specimens were deposited in the Department of Entomology and Environmental Protection of the Poznań University of Life Sciences.

In Poland, the populations of eastern and western moths occur sympatrically, constantly mixing, and therefore, we can exclude that the moths came from a few homogeneous populations. This is often a problem in studies involving relatively small areas for species that extend over much of the Palearctic. In addition, moths have been collected over a period of more than 20 years, which further diversifies the risk of "selective choice". For these reasons we examined the potential influence of the age of specimens in collection on the quality of discrimination of species. Similarly, we assessed whether male and female individuals differ spectrally. For this purpose, we used the same discriminant analyzes described later in the paper.

2.2. Expert examination

The species of each individual was determined on the basis of the following features: morphology of genitalia and color pattern of the front pair of wings. In a laboratory, the body parts and the external genitalia were dissected in a standard way for each individual. The abdomen was first removed and dipped for 24–36 h in 10% caustic potash (KOH). Genitalia were then removed from the softened surrounding tissues of male and female individuals. The phallus was removed, and the external genitalia was partially dehydrated with ethanol and mounted on glycerine between the microscope slides and cover slips. Latest available publication of Ronkay et al. [29], and also the drawings of the genitalia shown in the several papers served as a key is the division of individuals into two morphologically different groups [33–35].

The color pattern of the wings is an ambiguous feature. *D. chrysitis* differs externally from *D. stenochrysis* by its unbroken brown median area and more indistinct subterminal line [29]. Fig. 1a and 1b present the individuals with the typically marked species features on the forewings of *D. chrysitis* and *D. stenochrysis*. The results of this part served as the reference for the spectral and chemometric procedures.

2.3. Spectral measurements

Spectra were measured with a system consisting of the ASD FieldSpec 3 spectrophotometer (FieldSpec Analytical Spectral Devices, Inc., Boulder, Colorado, USA) attached by optical fiber to microscope NU 2 (VEB Carl Zeiss Jena, Jena, Germany). The spectrophotometer recorded the reflected electromagnetic radiation in the wavelength range from 350 to 2500 nm with a spectral sampling of 1.4 nm from 350 to 1000 nm and 2 nm from 1000 to 2500 nm. The spectral resolution in VIS was 3 nm and at 1400 and 2100 nm was 10 nm. The spectrophotometer was calibrated with a Spectralon (Labsphere) white standard before each measurement series. A plan apochromat objective (25x) and a coaxial illumination with a halogen lamp were used in the microscope.

Spectral measurements were carried out on the two dominant color areas on the wings, brown and golden iridescent on the dorsal side of the forewing, which are typical of *D. chrysitis* and *D. stenochrysis*. In the case of the brown area the reflectance was measured from the median area at the front edge of the forewing while golden iridescent included subterminal area. In measurement location wings are covered with scales of two types. Cover scales are visible from the outside and brown ground scales below them (Supplementary Fig. S1). Brown area is covered with brown melanin-pigmented cover scales. While on the shimmering area, cover scales are actually colorless melanin-deprived and referred to as glass scales. As a result of light interference and diffraction, they form a physical color ranging from bluish to dominant gold to copper. Glass scales differ significantly in structure from brown scales, which are perforated and higher in cross-section (Supplementary Fig. S1 and S2). The base brown scales are also perforated. The scales are arranged in many layers on the wing.

The spectra at wavelengths below 400 nm and above 2100 nm exhibited high levels of noise, and they were not used in further analysis. The measurements were made in triplicate on the dorsal forewing of each of the 95 individuals and then the values have been averaged for each item. Results of spectra measurements are presented in Fig. 2. During the measurements, the reference was a Spectralon diffuse reflectance standard (Labsphere), and therefore the non-diffusively reflecting surfaces of the wing could cause the relative reflectance values to be greater than 1. The spectra of both species appear to be similar. The reflectance from the glass scales in the entire range of the analyzed spectrum is higher than from the brown scales. In both cases, the reflectance in the UV range was low. The spectrum of the glass scales shows a distinct peak below 600 nm. The presence of iridescent golden color manifests it. The curve then descends to a local minimum between 750 and 800 nm, and the reflectance increases again. The brown scales absorb radiation in the VNIR range due to melanin, which is lower reflectance than that of colorless cover scales providing iridescent golden color.

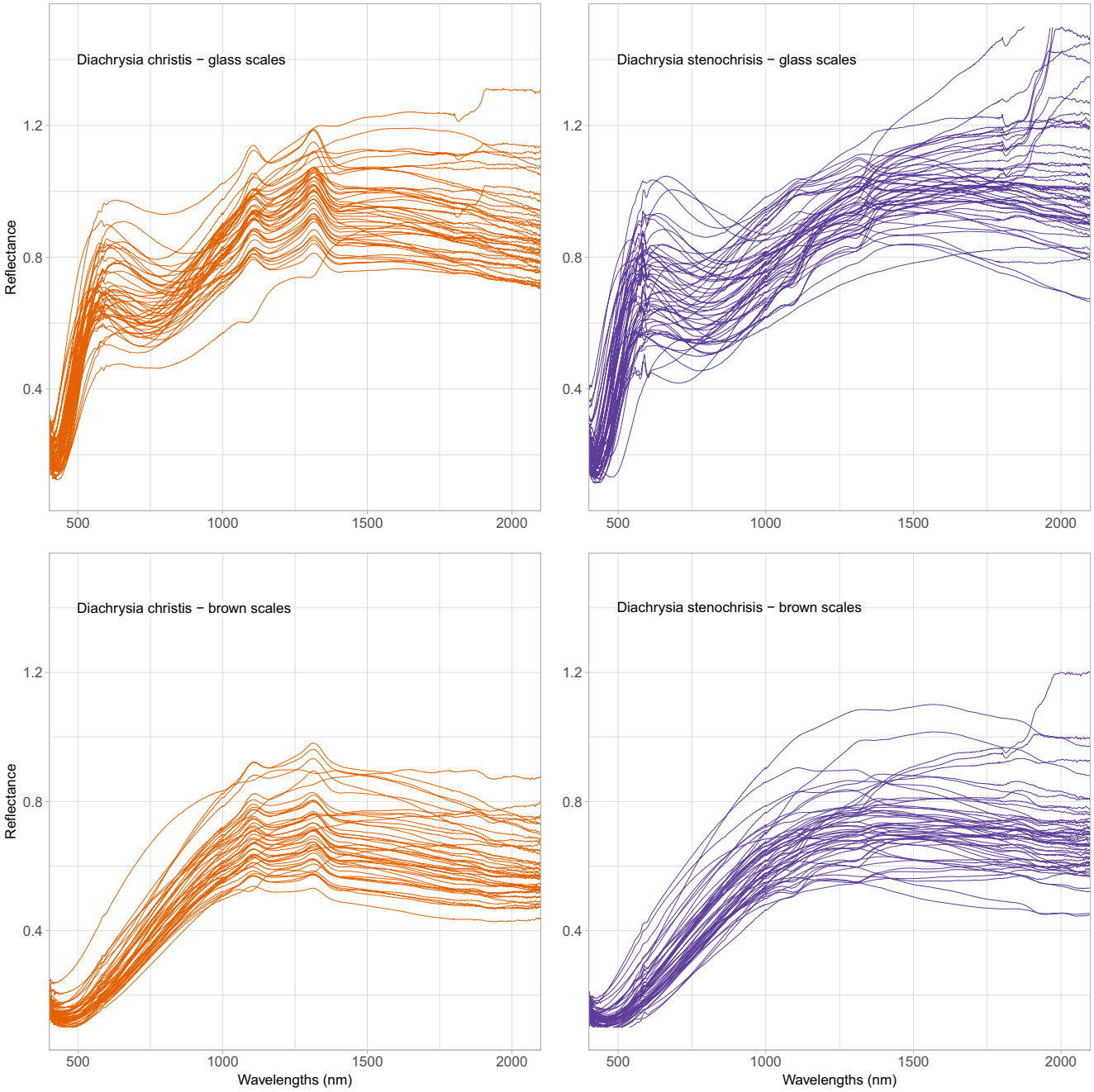


Fig. 2. Raw spectra for *D. chrysitis* and *D. stenochrysis* species.

2.4. Numerical analysis

In the first step, to highlight the spectra shape descriptors, data was transformed using Savitzky–Golay filter [36]. The filter removes unnecessary effects like baseline shifts and noise resulting from the non-ideal sampling process. The SG filter requires three parameters and by experiments we found that configuration: **differentiation order = 2, polynomial order = 2, and window size = 5** provides the best results. 3402 spectral bands after SG transformation were input features (predictors) for all classifiers. As labels, we used the experts' species identifications (either *D. chrysitis* or *D. stenochrysis*), and the entire dataset included 95 cases.

Typical machine learning procedures require dividing the collected data into at least two sets: the training - used to train the model, and the testing set - used for an independent evaluation of the model accuracy. With a small number of cases and many features, the role of features in the trained model may vary significantly for various training and testing sets. The consequence could be a potential over-fitting of the model to the provided data [37]. In our work, we attempt to reduce over-fitting and detect essential predictors by two independent methods. The first is fully stochastic both at the sampling and training level and works with a complete spectrum. The second is fully deterministic, and it first selects a limited subset of potentially useful spectral features and then discriminates *D. chrysitis* and *D. stenochrysis* by an exhaustive search inside this subset only. Both approaches have their own disadvantages, especially on small datasets. Stochastic approach, despite its utility, may result in random, non-optimal configuration, difficult to reproduce. Deterministic approach, through dimensionality reduction processes may make it difficult to find

less obvious but more optimal solutions. For this reason, the species discrimination process will be controlled by two independent algorithms.

In the procedure of discrimination, we analyzed popular machine learning algorithms. The performance results of tuned algorithms are shown in Table 1. The Support Vector Machine (SVM) and Random Forest (RF) achieved the highest performance, close to 1, for accuracy, specificity, and sensitivity. In the case of SVM, we interpret high efficiency as the effect of the existence of a clear hyperplane that separates both species in the feature space. The other tested algorithms did not achieve satisfactory performance. The latter is mainly the effect of an imbalance between the number of features and cases, and excludes the K-Nearest Neighbors (KNN). Overall, the low number of cases affects Back-Propagation Neural Network (BPNN) and eXtreme Gradient Boosting (XGBoost) performance. Due to the possibility of analyzing the role of individual features, we decided to use RF, both for discrimination and feature selection.

Random Forest is a stochastic algorithm and its results may change between each recurrence. We used the default model hyperparameters, i.e. number of trees = 500, tree depth = unlimited, mtry = 58, minimal node size = 1. Moreover, the standard procedure of machine learning training requires to hold part of the cases for the testing set. It means that the model is trained based on 60–80% of the entire dataset. Machine learning models gain their highest performance when the distributions of selected predictors (features) in the training and validation sets are close to each other. With large datasets this is not the case, but when the dataset is small, such an assumption is difficult to fulfill. It means that there is a risk that depending on the composition of the training set, the selected predictors will not be relevant for the cases included in the testing set. It is mainly related to high efficiency on the training set and low efficiency on the testing set.

Random Forest is considered to be an algorithm with a low risk of over-fitting. To determine the risk of over-fitting resulting from a small size of training set, we applied a bootstrapping procedure (Fig. 3-B), including 300 iterations. If such over-fitting risk is high, we can expect that the testing set's performance will vary from low to high between iterations. If such a risk is low, each iteration shall return with similar performance. Moreover, if the trained model will select the same predictors at each iteration, we can expect that correct and incorrect classification will include the same cases each time. At each iteration, the entire dataset is randomly divided into the training set including 70% (66 individuals) of cases and the testing set with remaining 30% (29 individuals). It means that we have 300 different training sets assessed by 300 different testing sets and each case had a chance to be included in the testing set 60 times on average. It is a large enough number to assess the stability of the process of discrimination (Figs. 3-B3). The RF algorithm also provides "feature importance" - a Gini index describing how each variable decreases the impurity of RF internal splits (Figs. 3-B2). The index varies between 0 and 1, where 0 denotes that the variable cannot increase the purity, while 1 means that variable allows to split the dataset into pure subgroups.

Table 1
Performance comparison of machine learning algorithms, calculated for testing sets. For abbreviations see text.

Model	Accuracy	Sensitivity	Specificity
SVM	0.99	1	0.99
RF	0.99	0.99	0.99
KNN	0.69	0.33	0.98
BPNN	0.89	0.85	0.93
XGBoost	0.88	0.85	0.91

In the second procedure (Fig. 3-C) we first attempted to define a group of the most significant spectrum bands for the distinction of *D. chrysitis* and *D. stenochrysis*. As "the most important", we define a limited, possibly small number of predictors that successfully separate the specimens under investigation. First, all bands were sorted in descending importance order. We used the value of two-sample Kolmogorov-Smirnov (K-S) test as a variable importance index usefulness of this statistic to feature selection results from the absence of the assumption on the form of the distributions of compared sets. For each wavelength D-statistic was calculated using equation (1):

$$D_s = sup|F_{dc}(x) - F_{ds}(x)| \quad (1)$$

where: F_{dc} and F_{ds} empirical distribution of *D. chrysitis* and *D. stenochrysis* subsets, and *sup* is supremum function i.e. the choice of the largest operand value. D-statistic is 0 when two distributions fully overlap, values between 0 and 1 when partially overlap, and 1 if two distributions are completely disjoint. All wavelengths were sorted (Figs. 3-D1) from the highest to lowest D-statistics, it means from the most to least separating features. Next, we searched for the minimal combination of predictors (Figs. 3-D2) providing perfect separation between *D. chrysitis* and *D. stenochrysis*. To evaluate the separation we used an accuracy of Linear Discriminant Analysis (LDA), which is considered as one of the best tools to find a minimal effective combination of spectral features [38]. The same procedure (Figs. 3-D1 and 3-D2) was applied to the best results of bootstrapping (Fig. 3-B).

2.5. Software

We used the R programming language [39] for the analysis; the Savitzky-Golay filter from prospectr [40] package to transform spectral curves; the ranger [41] package to train the random forest models; linear discrimination from the MASS [42] package.

3. Results

3.1. Stochastic discriminant analysis

In this procedure, all spectra, both of golden iridescent and brown fragments of the wing surfaces, were used simultaneously for classification. The classification achieved average compliance with the species identification made by entomologists at the level of 99% ± 1%. 91 individuals out of 95 were correctly classified during each of 300 testing iterations, and the remaining four individuals with over 98% correctness (two cases of both *D. chrysitis* and *D. stenochrysis* were classified incorrectly). It demonstrates the high stability of models trained against different training sets.

A one hundred percent reproducibility on the validation set for almost all cases indicates the presence of a group of relatively strong predictors among 3402 bands. The predictors are effective regardless of the configuration of the training sample. For each iteration of the simulation, the values of the importance index of all variables were recorded. After performing all iterations, the average value of the indexes was calculated. For the top 20 variables, listed in Table 2, the index value ranged from 0.49 for G1767 (glass scale, wavelength 1767 nm) to 0.22 for G1319 and only for the first 7 indicators the value of the index exceeded 0.3.

Our collection has a limited number of cases but a very large number of features. For this reason, we need to consider whether the very high accuracy, which we achieved in our experiment, is accidental; namely, the classifier would find distinctive features for other macroscopic divisions. In order to falsify this thesis, we run an experiment consisting of multiple random divisions of our collection of moths into two equal groups and then training and

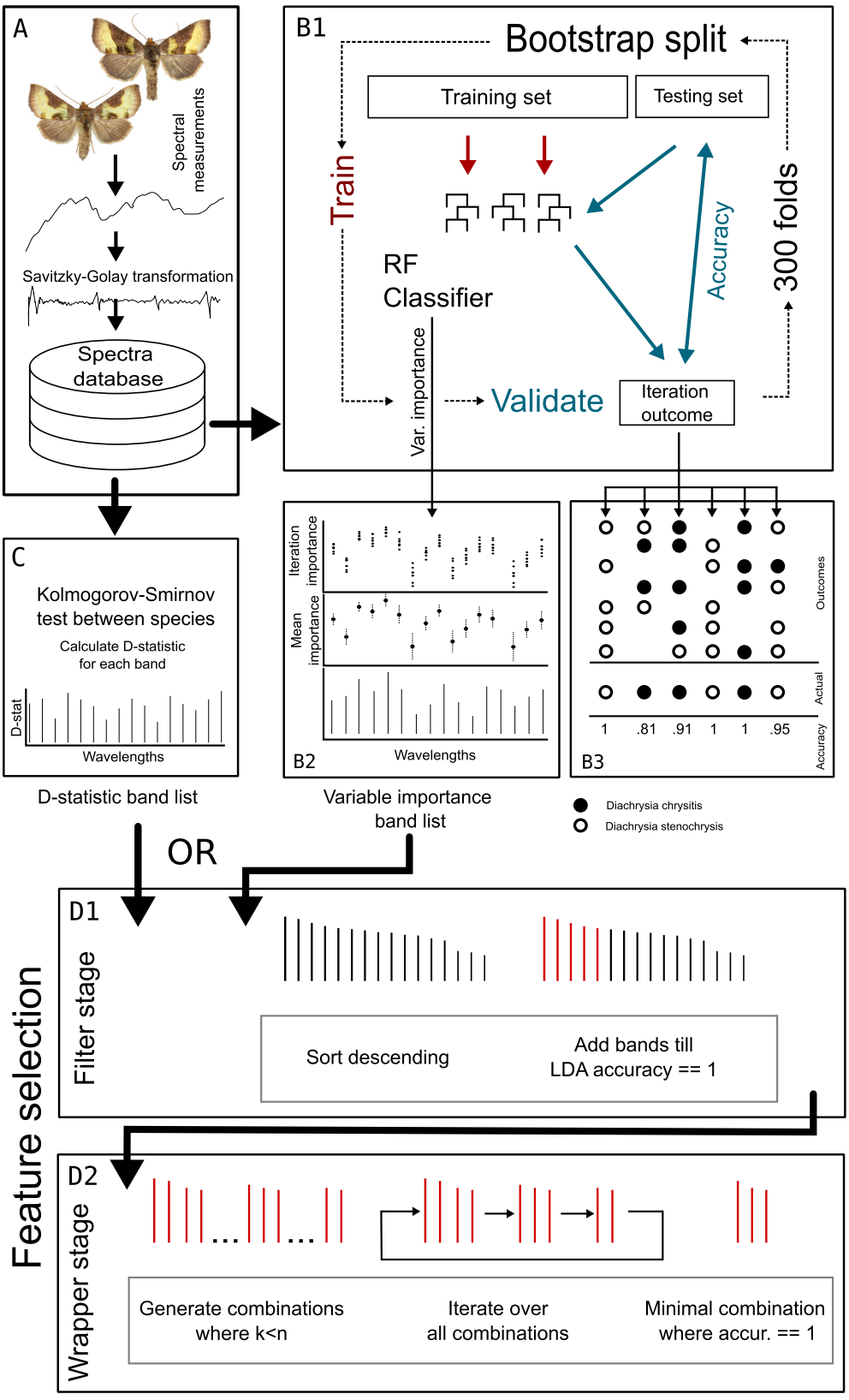


Fig. 3. Block diagram of the procedure. A total of 300 iterations of model training and testing were performed, and the final results were averaged.

testing the RF classifier on independent training and testing subsets. During 1000 iterations, the accuracy on the testing set varied from 0.38 to 0.74, with the average being 0.53. Such results show that in all iterations, the classifier cannot find significant features that would be present in both the training and the testing sets. By comparing the accuracy 0.98–1 with the average being close

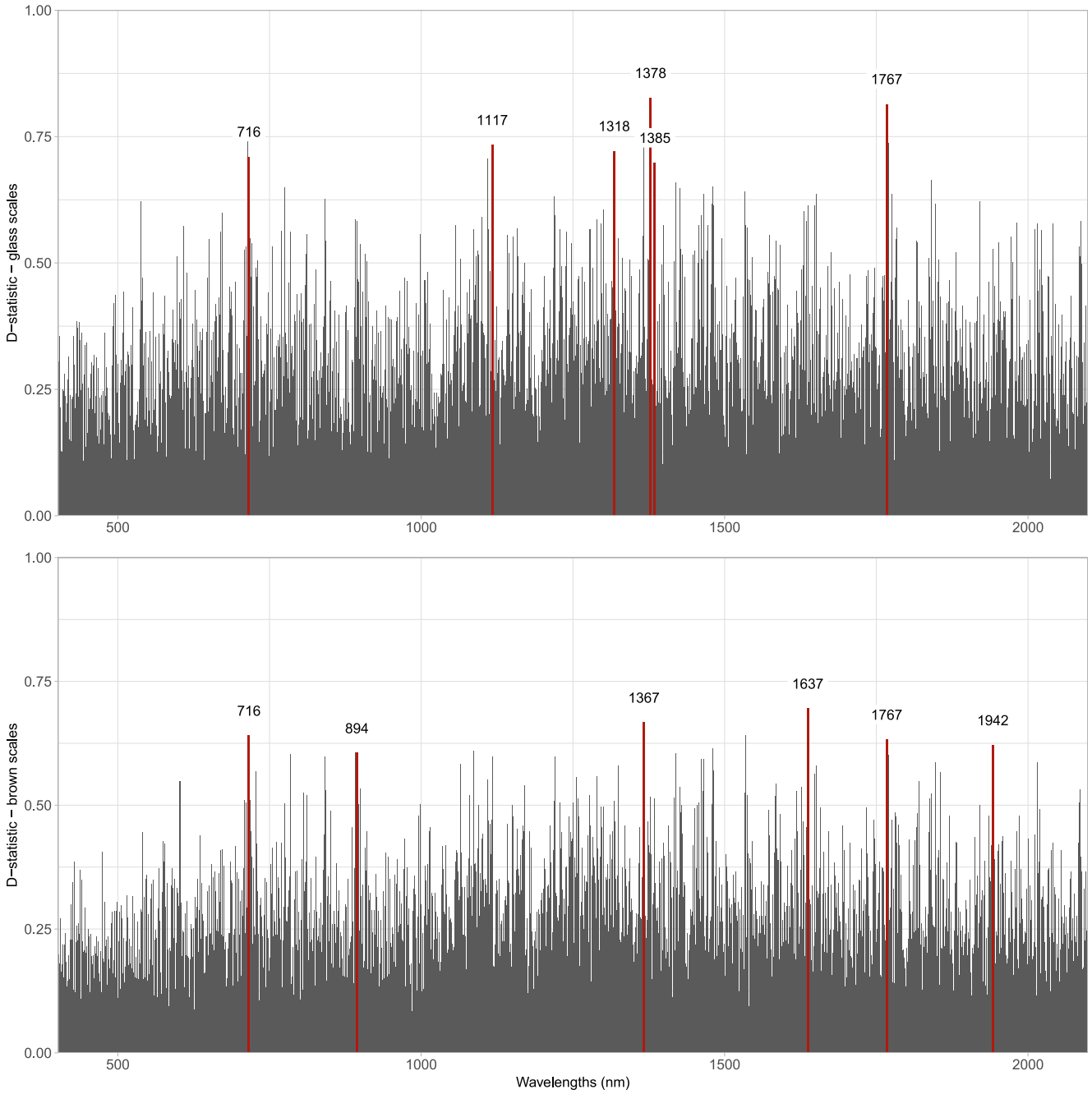


Fig. 4. The importance of spectral bands for specimens classification. The higher the value, the greater the suitability for distinguishing between *D. chrysitis* and *D. stenochrysis*. Most important bands are marked red. Notice that two bands 716 and 1767 nm appear in both types of scales.

to 1 achieved in our experiment for the division made by experts, we can reject the hypothesis that the high efficiency of our method may be coincidental.

3.2. Deterministic discriminant analysis

Fig. 4 shows the importance of each band by means of the D-statistic between *D. chrysitis* and *D. stenochrysis* for given wavelength for glass and brown scales. D-statistics range from 0.07 to 0.83, but associate p-values show that only D-statistics larger than 0.4 allow to accept the null hypothesis that distributions inside *D. chrysitis* and *D. stenochrysis* for the given band are different. This condition was true for no more than 500 bands. Only nine bands, all measured from the glass scales, have D-statistic greater than 0.7, with a maximum slightly exceeding 0.82 for band 1378 nm. Twenty wavelengths with the highest discriminating potential

Table 2
List of 20 best wavelengths indicated by Random Forest feature importance and Kolmogorov-Smirnov D-statistics.

Method	Wavelengths: G – glass scales, B – brown scales
Random Forest	G1767, G1378, G1117, G1367, G1318, G1770, G1385, G714, G1110, G716, G1419, B1637, G1841, G1480, B1100, G1651, B1399, B1367, B1248, G1319
Kolmogorov-Smirnov D-statistics	G1378, G1767, G1367, G714, G1770, G1117, G1318, G716, G1110, G1385, B1637, B1367, G1841, G1419, G1480, G775, G1427, B716, G1533, B1535

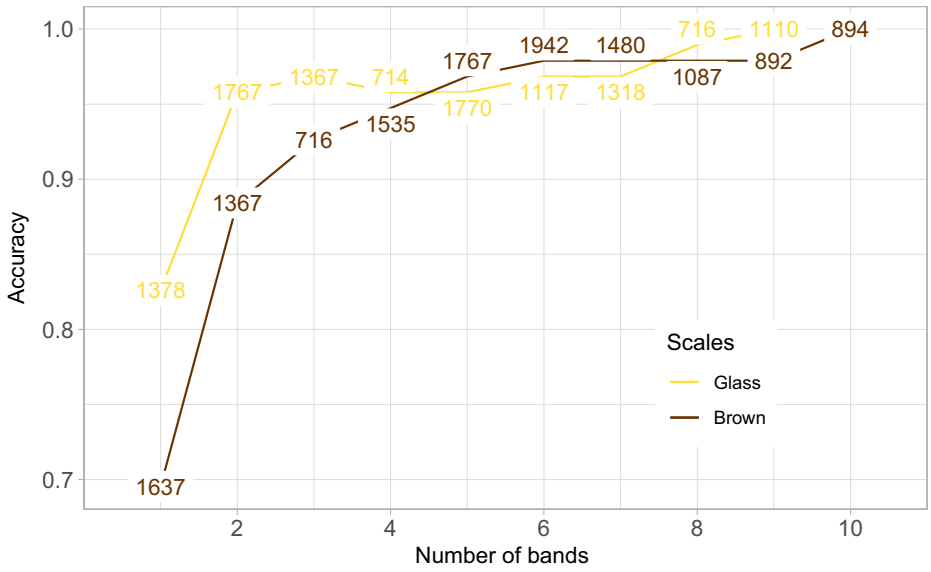


Fig. 5. Relation of the LDA accuracy of *D. chrysitis* and *D. stenochrysis* discrimination to the number of the consecutive spectral bands. Notice that the values on the Y axis do not start from 0.

are also listed in Table 2. It is worth pointing out that 10 best wavelengths indicated by D-statistics are the same as 10 best resulting from the stochastic procedure, but in slightly different order. All these 10 wavelengths were measured from glass scales.

There is no single spectral feature which fully distinguishes *D. chrysitis* and *D. stenochrysis*. It means that we must search for a minimal number of bands from the entire spectrum whose linear combination allows such distinction. The discrimination process was undertaken by sorting features, starting from the most important. To achieve this goal, we added the next band from the list in each step, checking by how much the accuracy would increase. After adding the top 4 bands, the accuracy stabilizes at 0.95. These changes are shown in Fig. 5. We found that linear combination of the first nine most important wavelengths (all from glass scales) perfectly differentiate *D. chrysitis* and *D. stenochrysis* in LDA space, thus limiting the number of possible band combinations (Fig. 5). We also tested the potential of spectra derived from the brown scales only. The discriminatory potential of the individual wavelengths is smaller compared to those sensed from glass scales, but their linear combination already shows similar performance to the bands from brown scales (Fig. 5).

3.3. Minimal discriminatory combination

In the last step of the analysis, we searched for the smallest possible group of wavelengths for each of the scales separately, which allow for full separation of *D. chrysitis* and *D. stenochrysis*. It should be emphasized that this is not a performance of LDA classifier measured on an independent testing set but only a group of bands for which there is a hyperplane in the multidimensional space completely separating *D. chrysitis* and *D. stenochrysis* in the collected dataset. There exist such four independent subsets for glass scales - one subset with three and three subsets with four wavelengths, and two subsets for brown scales - one with four and one subset with five wavelengths. All are presented in Table 3. Other, higher dimensional configurations are just supersets of those mentioned above. In that way we can indicate the most important wavelengths - for glass scales there are: 1378, 1767, 716, 1385, 1117, 1318, nm and for brown: 1637, 1367, 894, 1942, 716, 1767 nm. Band 1535 nm is omitted from the brown scale list because it does not contribute to any subset that fully separates *D. chrysitis* and *D. stenochrysis*. The isolation of the wavelengths of the greatest diagnostic importance prompts us to insight into the biophysical and biochemical reasons for their relevance. The above thesis becomes all the more important because both the stochastic and deterministic analysis showed the existence of the same 10 most important bands for glass scales. Moreover, the LDA revealed the significance of two wavelengths (716, 1767 nm) in both brown and glass scales.

4. Discussion

Previous studies on the spectral features of wings in moths of the genus *Diachrysia* focused on explaining the phenomenon of golden iridescence in *D. chrysitis* [43,44]. This arises as a result of interference, scattering and absorption in the structures of the glass scales. Based on our results we can apply the same for *D. stenochrysis*, whose wings have not been the subject of spectral studies so far. It should be noted that in the visible range there is a large variation in the iridescent color formed on colorless scales in the population of both species, from pale blue to pale copper. These color forms of moth wings were distinguished at the beginning of the 20th century [45]. In the course of the reflectance curves from glass scales of both species above 600 nm after the maximum reflectance caused by interference, a distinct decrease is visible. The primary building material of glass scales is chitin. Reeves [46] showed that the maximum chitin absorption is around 730 nm in this range. The overall absorption in this area may be influenced by melanin in the scales directly underneath the glass scales (Supplementary Fig. S1a). The brown scales lack this absorption region because the radiation absorption by melanin increases proportionally with the shortening of the wavelength. Using the reflected light spectrum recorded in the VNIR-SWIR range, we have collected other, potentially more useful information than that provided by the visible range. The light transmission in the wing structure drops significantly with distance and the radiation is strongly reflected. Based on that, we recognized that the spectrum recorded by the spectrometer was shaped by the scales structures on the upper side of the wing.

Our approach to species identification based on spectral characteristics of spotty-defined areas of the wing and machine learning is original and new. So far, microscopic spectra studies have focused on the identification of phenomena and comparative char-

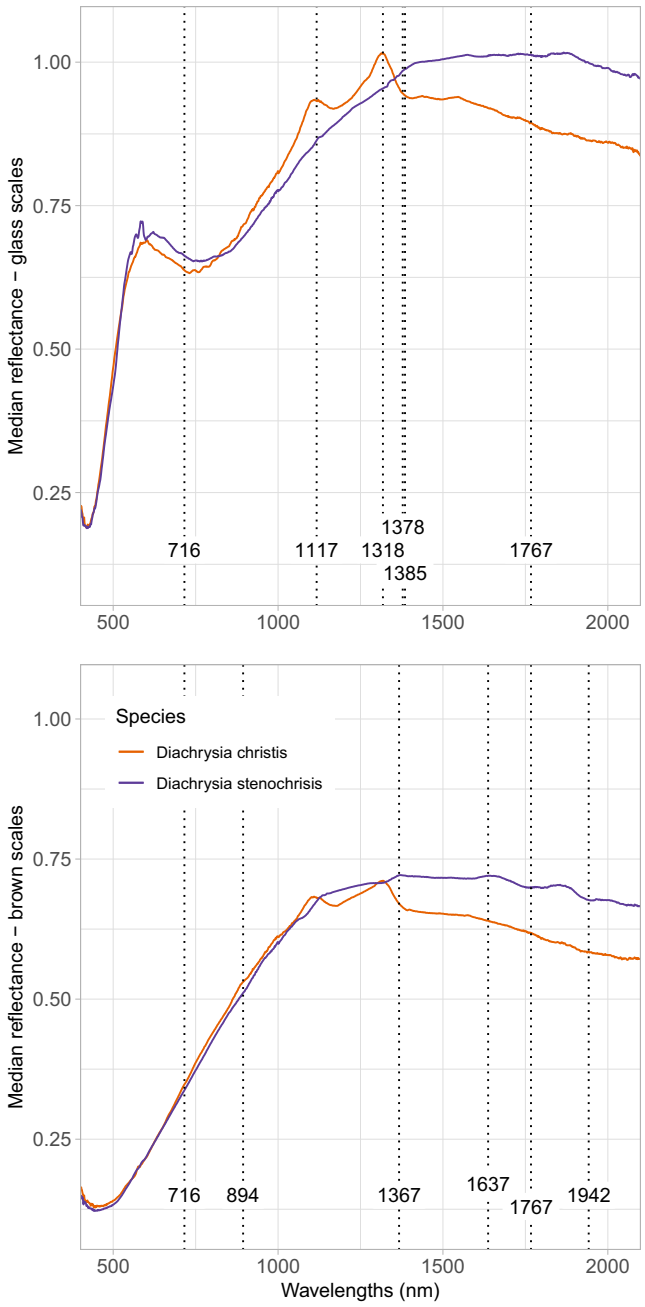


Fig. 6. The medians of the *D. chrysitis* and *D. stenochrysis* spectra for glass and brown scales with overprinted most important wavelengths.

acteristics of wing scales [4,13]. The analysis of the spectra is so complicated that even a trained expert can have problems distinguishing between species without using numerical methods. In our case, machine learning was crucial for decoding the information written in the form of the reflectance spectrum and to isolate information of significant importance for the determination of *Diachrysia* siblings. A comprehensive comparison between expert classification and iteratively trained stochastic RF classifiers revealed that models very rarely incorrectly assign labels to validation data, regardless of the changes of the training sets composition. This indicates that the diagnostic features favored by RF repeat in all individuals of the collection. It was also confirmed by Kolmogorov-Smirnov D-statistic which indicated the same group of predictors. This means that with the optimal size and quality of the training set, spectral measurements coupled with machine learning can be an effective tool not only for mass species differentiation but also can provide insight into the reasons for this variation.

Table 3
The minimal combination of wavelengths to achieve full discrimination of *D. chrysitis* and *D. stenochrysis* in LDA space. *Denote the D-statistic which is equivalent for a single band. The wavelengths marked in bold appear for the first time in the given subset.

Scales	Number of bands (dimensions)	List of bands (nm)	Accuracy
Glass	1	1378	0.826*
	2	**1767**, **716**	0.968
	3	**1385**, 1767, 716	1
	4	1378, 1767, 716, **1117**	1
		1378, 1767, 716, **1318**	
		1378, **1318**, 716, 1385	
Brown	1	1637	0.695*
	2	1637, **1535**	0.926
	3	1637, **1367**, **894**	0.968
		1637, **1367**, **1942**	
	4	1637, 1367, 1942, 894	1
	5	1637, 1367, 1942, **716**, **1767**	1

Since the identification of insects based on chemometric analysis is a new issue, an applicable spectral measurements library has not yet been developed. The only source of knowledge about the relationship between spectral bands and the properties of scales are previous studies. However, in-depth study of Lepidoptera wing scales that would explain the phenomena behind the meaning of the spectral features does not yet exist. In the developed model, the SWIR range turned out to be the most important in terms of species diagnostics. This is also seen in Fig. 6 where the visual differences between spectra *D. chrysitis* and *D. stenochrysis* are clearly visible. In the optimal model we have three or four unique bands subsets for glass scales and four or five unique subsets for brown (Table 3). Few bands are of particular importance: 1767, 716 nm and located in close proximity 1378/1367 nm appear in unique sets, both in glass and brown scales. Those three bands suggest the presence of features important for the separation of moths and independent of the type of scales. The remaining bands are important mainly for one type of scales. The amplitude of the spectra near these bands is small, indicating small differences between the *D. chrysitis* and *D. stenochrysis*.

The skeleton of the butterfly's scales is made of a composite material. There are three main components of Lepidoptera wing scales: chitin, proteins and melanin [47]. The overall characteristic of the spectrum of *Diachrysia* scales, especially glass scales, is very similar to the general spectrum of chitin presented by Apetroaei et al. [48]. Glass scales are colorless and melanin-deprived [49]. Studies on structural color analysis proved that the chitinous structure undoubtedly shapes the reflectance spectra which can be estimated from it [50,51]. Azofeifa et al. [51] observed that with the increasing complexity of the structures of the analyzed chitin material, the simulated spectra showed weakening of the wave pattern. Thus, not only the overall shape of the curve but also the subtle changes in monotonicity of the curve depend on the complexity of chitin layers. Fig. 6 presents location of the selected wavelengths indicated in "Minimal discriminatory combination" section of "Results" against medians of the raw spectra. All values correspond to the location of changes in monotonicity of the spectral curves. This means that wavelengths relate to subtle physicochemical differences between scales of *D. chrysitis* and *D. stenochrysis*, rather than the qualitatively identified presence or absence of selected chemical components.

Recently, scale ultrastructure can be described numerically on the basis of scanning electron microscope studies as shown by

Day et al. [14]. A detailed understanding of this structure could help answer the question that bothers us about the biophysical basis of the separateness of *Diachrysia* siblings. Scales ultrastructure is biochemically dependent on the remaining components of the scales. Melanin is the significant factor influencing the brown scales ultrastructure and response of our model. Moths' wings are usually dull as a consequence of high melanin content in brown or dark scales [5]. The same we can state on brown melanin-pigmented scales on *Diachrysia* wings. There exist several types of melanin and it strongly influences the color and ultrastructure of moth scales [52]. Brown scales on *Diachrysia* wings have a different spatial structure compared to glass ones (Supplementary Fig. S1 and S2). Perforated and pigmented brown scales scatter and absorb light more intensively than non-perforated [53]. Thus, melanin being one of the biochemical determinants of the formation of scale ultrastructure, may therefore indirectly influence its differentiation between siblings moths. Although we cannot clearly explain the role of all the indicated wavelengths, we consider that the features denoted as important for separation of *D. chrysitis* and *D. stenochrysis* relate to the scale ultrastructure and its major components, chitin and melanin. Moreover, it should be emphasized that two wavelengths: 1767 and 716 nm, were identified in both types of scales, which minimizes the risk that the indicated wavelengths are random.

5. Conclusions

The subject of our study were the twin species *D. chrysitis* and *D. stenochrysis* that cause identification problems for entomologists. The novelty of our approach is that the combination of reflectance spectroscopy and machine learning has been used to separate these two groups. We used microscopic spectroscopy to obtain pure spectra from scales on the forewing. Applied methods of specimens separation proved a decent accuracy. The advantage of the RF is a simplicity of procedure and no preliminary assumptions about the importance and distribution of individual predictors. Especially the latter means that none of the essential features will be omitted in the decision-making process. The feature selection procedure allows indicating a relatively small set (between three and six) of predictors with the highest discriminatory potential.

Spectral differences between *D. chrysitis* and *D. stenochrysis* are mainly the result of ultrastructure of scales influenced by chitin and melanin, the nature of which is not achievable for this spectral study. The differences are subtle and quantitative, nonetheless statistically significant, and allow, using just several predictors, to fully discriminate these two groups. At the same time, we showed that for other combinations of individuals, the classifier is not able to provide a separation beyond a random guess. Two of these wavelengths of 1767 and 716 nm appear in unique sets, both in glass and brown scales.

Biologists dealing with systematics and evolution of Lepidoptera are constantly looking for new sources of characteristics and methods to distinguish taxa. Since chitin-melanin cuticular skeleton occurs in Lepidoptera scales and are generally common in the world of insects, the reflectance spectroscopy supported by machine learning may be useful and effective to assist in the determination of species of Lepidoptera.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Mr. Jerzy Wrocławski, an expert in optics, for integrating the spectroscope with the microscope for the purposes of our research.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

H.R., R.W., J.P. developed the research concept. K.D. was responsible for the mathematical part, solved the research problem, wrote the code and prepared figures. R.W. collected individuals and made entomological evaluation. J.P., H.R. made spectral measurements under a microscope. J.J. described the methodological part, supervised the chemometric analysis and plotting the graphical results. K.D., H.R., J.J, J.P., R.W. wrote the main manuscript text. EG, MG took pictures with an electron microscope imaging technique. All authors contributed to the critical appraisal of the paper and approved the final version.

Data and code availability

The input data (legislative data and spectral measurements) with a description, and the results of this study are available in the GitHub repository (https://github.com/kadyb/diachrysia-classification) under the MIT license. The fully functional and reproducible codes are also available at this link.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2022.121058.

References

[1] T.D. Schultz, N.F. Hadley, Structural Colors of Tiger Beetles and Their Role in Heat Transfer through the Integument, Physiol. Zool. 60 (6) (1987) 737–745, https://doi.org/10.1086/physzool.60.6.30159990.
[2] C.-C. Tsai, R.A. Childers, N. Nan Shi, C. Ren, J.N. Pelaez, G.D. Bernard, N.E. Pierce, N. Yu, Physical and behavioral adaptations to prevent overheating of the living wings of butterflies, Nat. Commun. 11 (2020) 551, https://doi.org/10.1038/s41467-020-14408-8.
[3] S. Yoshioka, T. Nakano, Y. Nozue, S. Kinoshita, Coloration using higher order optical interference in the wing pattern of the Madagascan sunset moth, J. R. Soc. Interface. 5 (21) (2008) 457–464, https://doi.org/10.1098/rsif.2007.1268.
[4] B.D. Wilts, A.J.M. Vey, A.D. Briscoe, D.G. Stavenga, Longwing (Heliconius) butterflies combine a restricted set of pigmentary and structural coloration mechanisms, BMC Evol. Biol. 17 (2017) 226, https://doi.org/10.1186/s12862-017-1073-1.
[5] D.G. Stavenga, J.R.A. Wallace, E.J. Warrant, Bogong Moths Are Well Camouflaged by Effectively Decolourized Wing Scales, Front. Physiol. 11 (2020) 95, https://doi.org/10.3389/fphys.2020.00095.
[6] A. Krishna, X. Nie, A.D. Warren, J.E. Llorente-Bousquets, A.D. Briscoe, J. Lee, Infrared optical and thermal properties of microstructures in butterfly wings, Proc. Natl. Acad. Sci. 117 (3) (2020) 1566–1572, https://doi.org/10.1073/pnas.1906356117.
[7] R.H. Wilson, K.P. Nadeau, F.B. Jaworski, B.J. Tromberg, A.J. Durkin, Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization, J. Biomed. Opt. 20 (2015), https://doi.org/10.1117/1.JBO.20.3.030901 030901.
[8] H.-S. Song, K.-T. Lee, S.-M. Park, O.-J. Kang, H.-S. Cheong, Measurement of Deproteinization and Deacetylation of Chitin and Chitosan by Near Infrared Spectroscopy, Korean, J. Fish. Aquat. Sci. 36 (2003) 88–93, https://doi.org/10.5657/KFAS.2003.36.2.088.
[9] A. Gebru, M. Brydegaard, E. Rohwer, P. Neethling, Probing insect backscatter cross section and melanization using kHz optical remote detection system, J. Appl. Remote Sens. 11 (2017) 016015, https://doi.org/10.1117/1.JRS.11.016015.
[10] A. Gebru, S. Jansson, R. Ignell, C. Kirkeby, J.C. Prangsma, M. Brydegaard, Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight, J. Biophotonics. 11 (2018) e201800014, https://doi.org/10.1002/jbio.201800014.

[11] M. Brydegaard, S. Svanberg, Photonic Monitoring of Atmospheric and Aquatic Fauna, Laser Photonics Rev. 12 (2018) 1800135, https://doi.org/10.1002/lpor.201800135.
[12] A.J. Parnell, J.E. Bradford, E.V. Curran, A.L. Washington, G. Adams, M.N. Brien, S. L. Burg, C. Morochz, J.P.A. Fairclough, P. Vukusic, S.J. Martin, S. Doak, N.J. Nadeau, Wing scale ultrastructure underlying convergent and divergent iridescent colours in mimetic *Heliconius* butterflies, J. R. Soc. Interface. 15 (2018) 20170948, https://doi.org/10.1098/rsif.2017.0948.
[13] T.M. Trzeciak, B.D. Wilts, D.G. Stavenga, P. Vukusic, Variable multilayer reflection together with long-pass filtering pigment determines the wing coloration of papilionid butterflies of the nireus group, Opt. Express. 20 (2012) 8877, https://doi.org/10.1364/OE.20.008877.
[14] C.R. Day, J.J. Hanly, A. Ren, A. Martin, Sub-micrometer insights into the cytoskeletal dynamics and ultrastructural diversity of butterfly wing scales, Dev. Dyn. 248 (8) (2019) 657–670, https://doi.org/10.1002/dvdy.63.
[15] L.-Q. Zhu, Z. Zhang, Auto-classification of insect images based on color histogram and GLCM, in: 2010 Seventh Int. Conf. Fuzzy Syst. Knowl. Discov., IEEE, Yantai, China, 2010: pp. 2589–2593. http://doi.org/10.1109/FSKD.2010.5569848.
[16] J. Lim, J. Cho, T. Nam, S. Kim, Development of a classification algorithm for butterflies and ladybugs, in: TENCON 2006 - 2006 IEEE Reg. 10 Conf., IEEE, Hong Kong, China, 2006: pp. 1–3. http://doi.org/10.1109/TENCON.2006.344144.
[17] H. Yang, W. Liu, K. Xing, J. Qiao, X. Wang, L. Gao, Z. Shen, Research on insect identification based on pattern recognition technology, Sixth Int. Conf. Nat. Comput. IEEE, Yantai, China 2010 (2010) 545–548, https://doi.org/10.1109/ICNC.2010.5583156.
[18] J.B. Johnson, M. Naiker, Seeing red: A review of the use of near-infrared spectroscopy (NIRS) in entomology, Appl. Spectrosc. Rev. 55 (2020) 810–839, https://doi.org/10.1080/05704928.2019.1685532.
[19] Y. Kaya, L. Kayci, Application of artificial neural network for automatic detection of butterfly species using color and texture features, Vis. Comput. 30 (1) (2014) 71–79, https://doi.org/10.1007/s00371-013-0782-8.
[20] K.L. Keegan, J. Rota, R. Zahiri, A. Zilli, N. Wahlberg, B.C. Schmidt, J.D. Lafontaine, P.Z. Goldstein, D.L. Wagner, Toward a Stable Global Noctuidae (Lepidoptera) Taxonomy, Insect Syst. Divers. 5 (2021) 1, https://doi.org/10.1093/isd/ixab005.
[21] N.P. Kristensen, M.J. Scoble, O. Karsholt, Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity, Zootaxa. 1668 (2007) 699–747. http://doi.org/10.11646/zootaxa.1668.1.30.
[22] P.D.N. Hebert, A. Cywinska, S.L. Ball, J.R. deWaard, Biological identifications through DNA barcodes, Proc. R. Soc. Lond. B Biol. Sci. 270 (1512) (2003) 313–321, https://doi.org/10.1098/rspb.2002.2218.
[23] I. Svensson, P. Douwes, B.O. Stille, Are Diachrysia chrysitis (L.) and D. tutti (Kostrowicki) different species?, (Lepidoptera: Noctuidae), Insect Syst. Evol. 20 (1) (1989) 15–22, https://doi.org/10.1163/187631289X00474.
[24] A. Hille, M.A. Miller, S. Erlacher, DNA sequence variation at the mitochondrial cytochrome oxidase I subunit among pheromotypes of the sibling taxa Diachrysia chrysitis and D. tutti (Lepidoptera: Noctuidae), Zool. Scr. 34 (1) (2005) 49–56, https://doi.org/10.1111/j.1463-6409.2005.00171.x.
[25] P. Huemer, P. Hebert, DNA-Barcoding der Schmetterlinge (Lepidoptera) Vorarlbergs (Österreich) - Erkenntnisse und Rückschlüsse, Inatura - Forsch, Online 15 (2015) 1–36.
[26] D.H. Janzen, M. Hajibabaei, J.M. Burns, W. Hallwachs, E.d. Remigio, P.D.N. Hebert, Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding, Philos. Trans. R. Soc. B Biol. Sci. 360 (1462) (2005) 1835–1845, https://doi.org/10.1098/rstb.2005.1715.
[27] A. Hausmann, G. Haszprunar, A.H. Segerer, W. Speidel, G. Behounek, P.D. Hebert, Now DNA-barcoded: the butterflies and larger moths of Germany, Spixiana 34 (2011) 47–58.
[28] A. Segerer, Beitrag zur Genitaldiagnose einiger bayerischer Tagfalterarten unter besonderer Berueckschtigung der Weibchen, Beitr Bayer Entomofaun. 4 (2001) 5–25.
[29] L. Ronkay, G. Ronkay, G. Behounek, Plusiinae 1, Heterocera Press, Budapest, 2008.
[30] M.S. Engel, L.M.P. Ceríaco, G.M. Daniel, P.M. Dellapé, I. Löbl, M. Marinov, R.E. Reis, M.T. Young, A. Dubois, I. Agarwal, P. Lehmann A., M. Alvarado, N. Alvarez, F. Andreone, K. Araujo-Vieira, J.S. Ascher, D. Baêta, D. Baldo, S.A. Bandeira, P. Barden, D.A. Barrasso, L. Bendifallah, F.A. Bockmann, W. Böhme, A. Borkent, C. R.F. Brandão, S.D. Busack, S.M. Bybee, A. Channing, S. Chatzimanolis, M.J.M. Christenhusz, J.V. Crisci, G. D'elía, L.M. Da Costa, S.R. Davis, C.A.S. De Lucena, T. Deuve, S. Fernandes Elizalde, J. Faivovich, H. Farooq, A.W. Ferguson, S. Gippoliti, F.M.P. Gonçalves, V.H. Gonzalez, E. Greenbaum, I.A. Hinojosa-Díaz, I. Ineich, J. Jiang, S. Kahono, A.B. Kury, P.H.F. Lucinda, J.D. Lynch, V. Malécot, M. P. Marques, J.W.M. Marris, R.C. Mckellar, L.F. Mendes, S.S. Nihei, K. Nishikawa, A. Ohler, V.G.D. Orrico, H. Ota, J. Paiva, D. Parrinha, O.S.G. Pauwels, M.O. Pereyra, L.B. Pestana, P.D.P. Pinheiro, L. Prendini, J. Prokop, C. Rasmussen, M.-O. Rödel, M.T. Rodrigues, S.M. Rodríguez, H. Salatnaya, Í. Sampaio, A. Sánchez-García, M.A. Shebl, B.S. Santos, M.M. Solórzano-Kraemer, A.C.A. Sousa, P. Stoev, P. Teta, J.-F. Trape, C.V.-D. Dos Santos, K. Vasudevan, C.J. Vink, G. Vogel, P. Wagner, T. Wappler, J.L. Ware, S. Wedmann, C.K. Zacharie, The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches, Zool. J. Linn. Soc. 193 (2021) 381–387. http://doi.org/10.1093/zoolinnean/zlab072.
[31] S.M.N. Lazzari, F.C. Ceruti, J.I. Rodriguez-Fernandez, G. Opit, F.A. Lazzari, Intra and interspecific variation assessment in Psocoptera using near spectoscopy, Julius-Kühn-Arch. 425 (2010) 139–144, https://doi.org/10.5073/JKA.2010.425.250.
[32] S. Wu, C.-M. Chang, G.-S. Mai, D.R. Rubenstein, C.-M. Yang, Y.-T. Huang, H.-H. Lin, L.-C. Shih, S.-W. Chen, S.-F. Shen, Artificial intelligence reveals environmental constraints on colour diversity in insects, Nat. Commun. 10 (2019) 4554, https://doi.org/10.1038/s41467-019-12500-2.
[33] B. Goater, L. Ronkay, M. Fibiger, B. Goater, Catocalinae & Plusiinae, Entomological Press, Sorø, 2003.
[34] S. Kostrowicki, Studies on the palaearctic species of the subfamily Plusiinae (Lepidoptera, Phalaenidae), Acta Zool Cracov. 10 (1961) 367–472.
[35] E. Arnold, Untersuchung schwerbestimmbarer Nachtfalter anhand von Genitalpraeparation am Beispiel Plusia chrysitis und Plusia tutti, Facetta. 5 (1992) 1–27.
[36] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (8) (1964) 1627–1639, https://doi.org/10.1021/ac60214a047.
[37] I. Tsamardinos, E. Greasidou, G. Borboudakis, Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation, Mach. Learn. 107 (12) (2018) 1895–1922, https://doi.org/10.1007/s10994-018-5714-4.
[38] J. Piekarczyk, H. Ratajkiewicz, J. Jasiewicz, D. Sosnowska, A. Wójtowicz, An application of reflectance spectroscopy to differentiate of entomopathogenic fungi species, J. Photochem. Photobiol. B. 190 (2019) 32–41, https://doi.org/10.1016/j.jphotobiol.2018.10.024.
[39] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. https://www.R-project.org/.
[40] A. Stevens, L. Ramirez-Lopez, An introduction to the prospectr package, 2020.
[41] M.N. Wright, A. Ziegler, ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, J. Stat. Softw. 77 (2017). 10.18637/jss.v077.i01.
[42] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, Fourth, Springer, New York, 2002. http://www.stats.ox.ac.uk/pub/MASS4/.
[43] D. Pantelić, S. Savić-Šević, D.V. Stojanović, S. Ćurčić, A.J. Krmpot, M. Rabasović, D. Pavlović, V. Lazović, V. Milošević, Scattering-enhanced absorption and interference produce a golden wing color of the burnished brass moth, *Diachrysia chrysitis*, Phys. Rev. E. 95 (2017) 032405, https://doi.org/10.1103/PhysRevE.95.032405.
[44] S. Savić-Šević, D. Pantelić, B. Jelenković, B. Salatić, D.V. Stojanović, Golden moth-inspired structures with a synergistic effect of interference, absorption and scattering, Soft Matter. 14 (27) (2018) 5595–5603, https://doi.org/10.1039/C8SM00683K.
[45] J. Romaniszyn, F. Schille, Fauna motyli Polski (1929).
[46] J.B. Reeves, Effects of Water on the Spectra of Model Compounds in the Short-Wavelength near Infrared Spectral Region (14,000–9091 cm $^{-1}$ or 714–1100 nm), J. Infrared Spectrosc. 2 (4) (1994) 199–212, https://doi.org/10.1255/jnirs.46.
[47] S. Hunt, Composition of scales from the moth Xylophasia monoglypha, Experientia 27 (9) (1971) 1030–1031, https://doi.org/10.1007/BF02138861.
[48] M. Apetroaei, A.-M. Manea, G. Tihan, R. Zgarian, V. Schroder, I. Rău, Improved method of chitosan extraction from different crustacean species of Romanian Black Sea coast, UPB Sci. Bull. Ser. B Chem. Mater. Sci. 79 (2017) 25–36.
[49] D. Gomez, C. Pinna, J. Pairraire, M. Arias, J. Barbut, A. Pomerantz, C. Noûs, W. Daney de Marcillac, S. Berthier, N. Patel, C. Andraud, M. Elias, Transparency in butterflies and moths: structural diversity, optical properties and ecological relevance, Evol. Biol. (2020), https://doi.org/10.1101/2020.05.14.093450.
[50] J.P. Vigneron, J.-F. Colomer, N. Vigneron, V. Lousse, Natural layer-by-layer photonic structure in the squamae of *Hoplia coerulea* (Coleoptera), Phys. Rev. E. 72 (2005) 061904, https://doi.org/10.1103/PhysRevE.72.061904.
[51] D.E. Azofeifa, H.J. Arguedas, W.E. Vargas, Optical properties of chitin and chitosan biopolymers with application to structural color analysis, Opt. Mater. 35 (2) (2012) 175–183, https://doi.org/10.1016/j.optmat.2012.07.024.
[52] Y. Matsuoka, A. Monteiro, Melanin Pathway Genes Regulate Color and Morphology of Butterfly Wing Scales, Cell Rep. 24 (1) (2018) 56–65, https://doi.org/10.1016/j.celrep.2018.05.092.
[53] D.G. Stavenga, S. Stowe, K. Siebke, J. Zeil, K. Arikawa, Butterfly wing colours: scale beads make white pierid wings brighter, Proc. R. Soc. Lond. B Biol. Sci. 271 (1548) (2004) 1577–1584, https://doi.org/10.1098/rspb.2004.2781.