**Supplementary Table legends**

**Supplementary Table 1**. Taxonomic information of specimens included in the analysis.

**Supplementary Table 2.**  Training accuracy scores.

**Supplementary Table 3.** Summary of classification report: Number of classes found, accuracy, precision, recall, F1 score and support values of the classifiers tested.

**Supplementary Table 4.** ANOVA results for accuracy scores comparison.

**Supplementary Table 5.** Tukey multiple comparisons test for accuracy scores.

**Supplementary Table 6.** ANOVA results for F1 scores comparison.

**Supplementary Table 7.** Tukey multiple comparisons test for F1 scores

**Supplementary Tables**

| **Supplementary Table 2.** | |
| --- | --- |
| Linear Discriminant Analysis (LDA) | 0.747155 |
| Random Forest Classifier (RFC) | 0.640068 |
| Support Vector Machines (SVM) | 0.602331 |
| K-Nearest Neighbors (KNN) | 0.567963 |
| Naive Bayes (NB) | 0.439090 |
| Classification and Regression Trees (CART) | 0.433563 |
| Logistic Regression (LR) | 0.193987 |

| **Supplementary Table 3.** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Classes** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **Support** |
| Linear Discriminant Analysis (LDA) | 100 | 0.733715 | 0.75 | 0.73 | 0.73 | 1397 |
| Random Forest Classifier (RFC) | 92 | 0.642806 | 0.63 | 0.64 | 0.59 | 1397 |
| Support Vector Machines (SVM) | 92 | 0.614173 | 0.55 | 0.61 | 0.56 | 1397 |
| K-Nearest Neighbors (KNN) | 94 | 0.574087 | 0.58 | 0.57 | 0.55 | 1397 |
| Classification and Regression Trees (CART) | 99 | 0.428060 | 0.44 | 0.43 | 0.43 | 1397 |
| Naive Bayes (NB) | 92 | 0.418038 | 0.49 | 0.42 | 0.42 | 1397 |
| Logistic Regression (LR) | 92 | 0.193987 | 0.14 | 0.19 | 0.11 | 1397 |

| **Supplementary Table 4.** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **df** | **sum\_sq** | **mean\_sq** | **F** | **PR(>F)** |
| **C(variable)** | 6 | 2,178631 | 0,363105 | 547,8512 | 2,29E-52 |
| **Residual** | 63 | 0,041755 | 0,000663 |  |  |

| **Supplementary Table 5.** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **group1** | **group2** | **Diff** | **Lower** | **Upper** | **q-value** | **p-value** |
| **0** | LR | CART | 0,241004 | 0,205941 | 0,276067 | 29,60327 | 0,001 |
| **1** | LR | KNN | 0,395191 | 0,360128 | 0,430254 | 48,54254 | 0,001 |
| **2** | LR | NB | 0,225108 | 0,190045 | 0,260171 | 27,65066 | 0,001 |
| **3** | LR | LDA | 0,560378 | 0,525315 | 0,595442 | 68,83297 | 0,001 |
| **4** | LR | SVM | 0,450823 | 0,41576 | 0,485886 | 55,37593 | 0,001 |
| **5** | LR | RFC | 0,465005 | 0,429942 | 0,500068 | 57,11795 | 0,001 |
| **6** | CART | KNN | 0,154187 | 0,119124 | 0,18925 | 18,93926 | 0,001 |
| **7** | CART | NB | 0,015896 | -0,01917 | 0,05096 | 1,952608 | 0,786122 |
| **8** | CART | LDA | 0,319374 | 0,284311 | 0,354437 | 39,2297 | 0,001 |
| **9** | CART | SVM | 0,209819 | 0,174755 | 0,244882 | 25,77265 | 0,001 |
| **10** | CART | RFC | 0,224001 | 0,188937 | 0,259064 | 27,51467 | 0,001 |
| **11** | KNN | NB | 0,170084 | 0,13502 | 0,205147 | 20,89187 | 0,001 |
| **12** | KNN | LDA | 0,165187 | 0,130124 | 0,20025 | 20,29043 | 0,001 |
| **13** | KNN | SVM | 0,055632 | 0,020568 | 0,090695 | 6,83339 | 0,001 |
| **14** | KNN | RFC | 0,069814 | 0,03475 | 0,104877 | 8,57541 | 0,001 |
| **15** | NB | LDA | 0,335271 | 0,300207 | 0,370334 | 41,1823 | 0,001 |
| **16** | NB | SVM | 0,225715 | 0,190652 | 0,260778 | 27,72526 | 0,001 |
| **17** | NB | RFC | 0,239897 | 0,204834 | 0,27496 | 29,46728 | 0,001 |
| **18** | LDA | SVM | 0,109556 | 0,074492 | 0,144619 | 13,45704 | 0,001 |
| **19** | LDA | RFC | 0,095374 | 0,06031 | 0,130437 | 11,71502 | 0,001 |
| **20** | SVM | RFC | 0,014182 | -0,02088 | 0,049245 | 1,742019 | 0,871379 |

| **Supplementary Table 6.** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **df** | **sum\_sq** | **mean\_sq** | **F** | **PR(>F)** |
| **C(variable)** | 6 | 2,158777 | 0,359796 | 551,4847 | 1,87E-52 |
| **Residual** | 63 | 0,041102 | 0,000652 |  |  |

| **Supplementary Table 7.** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **group1** | **group2** | **Diff** | **Lower** | **Upper** | **q-value** | **p-value** |
| **0** | LR | CART | 0,246162 | 0,211374 | 0,28095 | 30,47608 | 0,001 |
| **1** | LR | KNN | 0,395191 | 0,360403 | 0,429979 | 48,92669 | 0,001 |
| **2** | LR | NB | 0,225108 | 0,19032 | 0,259896 | 27,86948 | 0,001 |
| **3** | LR | LDA | 0,560378 | 0,525591 | 0,595166 | 69,37769 | 0,001 |
| **4** | LR | SVM | 0,450823 | 0,416035 | 0,485611 | 55,81416 | 0,001 |
| **5** | LR | RFC | 0,460923 | 0,426136 | 0,495711 | 57,06466 | 0,001 |
| **6** | CART | KNN | 0,149029 | 0,114242 | 0,183817 | 18,45061 | 0,001 |
| **7** | CART | NB | 0,021054 | -0,01373 | 0,055842 | 2,606597 | 0,521343 |
| **8** | CART | LDA | 0,314217 | 0,279429 | 0,349004 | 38,90161 | 0,001 |
| **9** | CART | SVM | 0,204661 | 0,169873 | 0,239449 | 25,33808 | 0,001 |
| **10** | CART | RFC | 0,214762 | 0,179974 | 0,24955 | 26,58858 | 0,001 |
| **11** | KNN | NB | 0,170084 | 0,135296 | 0,204871 | 21,0572 | 0,001 |
| **12** | KNN | LDA | 0,165187 | 0,130399 | 0,199975 | 20,45101 | 0,001 |
| **13** | KNN | SVM | 0,055632 | 0,020844 | 0,090419 | 6,887468 | 0,001 |
| **14** | KNN | RFC | 0,065732 | 0,030944 | 0,10052 | 8,137974 | 0,001 |
| **15** | NB | LDA | 0,335271 | 0,300483 | 0,370059 | 41,50821 | 0,001 |
| **16** | NB | SVM | 0,225715 | 0,190927 | 0,260503 | 27,94467 | 0,001 |
| **17** | NB | RFC | 0,235816 | 0,201028 | 0,270604 | 29,19518 | 0,001 |
| **18** | LDA | SVM | 0,109556 | 0,074768 | 0,144343 | 13,56354 | 0,001 |
| **19** | LDA | RFC | 0,099455 | 0,064667 | 0,134243 | 12,31303 | 0,001 |
| **20** | SVM | RFC | 0,010101 | -0,02469 | 0,044888 | 1,250506 | 0,9 |