

Supplementary figure captions:

SFig. 1: The UV spectra (A: absorption of samples based on the extraction of different solvents; B: absorption of chloroform).

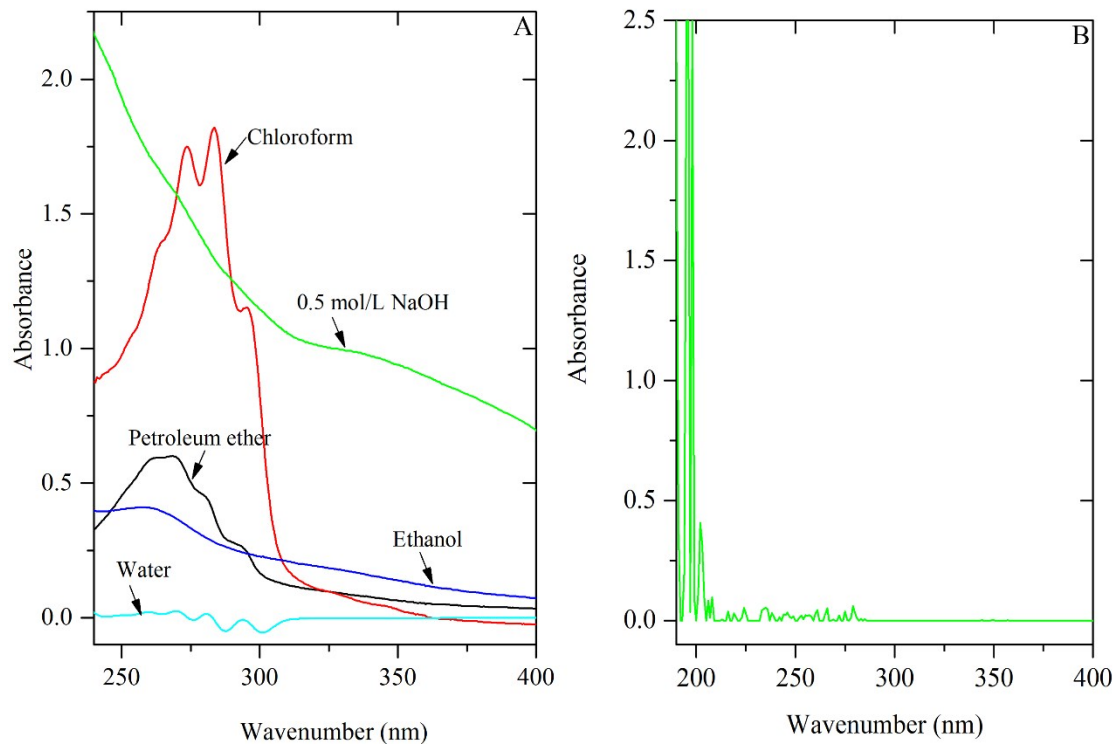
SFig. 2: A brief flow chart of data treatment.

SFig. 3: A simple illustration for explaining the operation principle of k-NN.

SFig. 4: The PCA score plots with 95% confidence ellipses for each dataset (A: UV-Cap; B: UV-Stipe; C: FTIR-Cap; D: FTIR-Stipe).

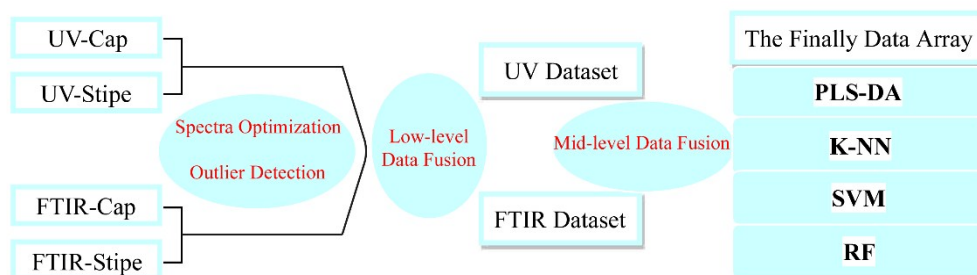
SFig. 5: The searching process of important parameters and the accuracy of test samples for SVM origin discrimination (A: Genetic algorithm; B: Accuracy of test samples based on c and g from genetic algorithm; C: Grid search method; D: Accuracy of test samples based on c and g from genetic algorithm).

SFig. 6: The RF model optimization procedure of n_{tree} and m_{try} for origin discrimination.



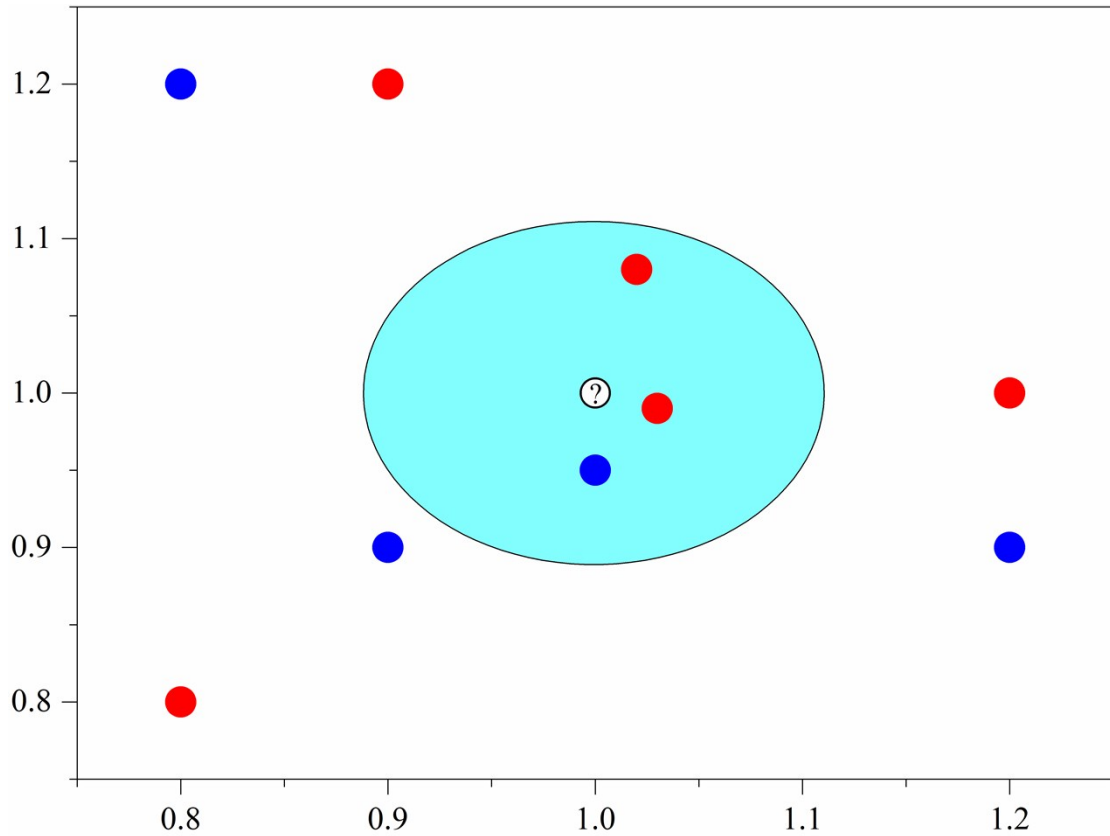
SFig. 1

(Note: This figure shows the UV absorption of samples based on the extraction of different solvents. Comparatively, chloroform can present the best peak shape, which can better reflect the chemical information of mushrooms. As the UV cutoff affected the quality of spectra, the data from 190 nm to 240 nm were deleted in data analyze process. Meanwhile, the blank absorption (240-400 nm) was deducted before the data analyze.)



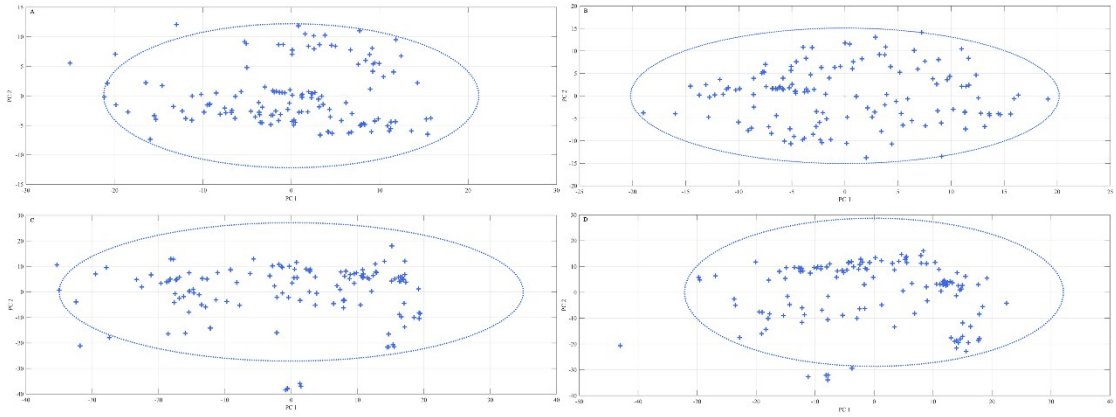
SFig. 2

(Note: This figure shows the data treatment process before mathematical analysis. Four data matrixes defined as UV-Cap, UV-Stipe, FTIR-Cap and FTIR-Stipe were collected in the experiment. Before the final data array was created, four steps named spectra optimization, outlier detection, low-level data fusion for cap and stipe data, and mid-level data fusion for FTIR and UV datasets. Finally, we performed four mathematical algorithms to analyze this obtained data array.)



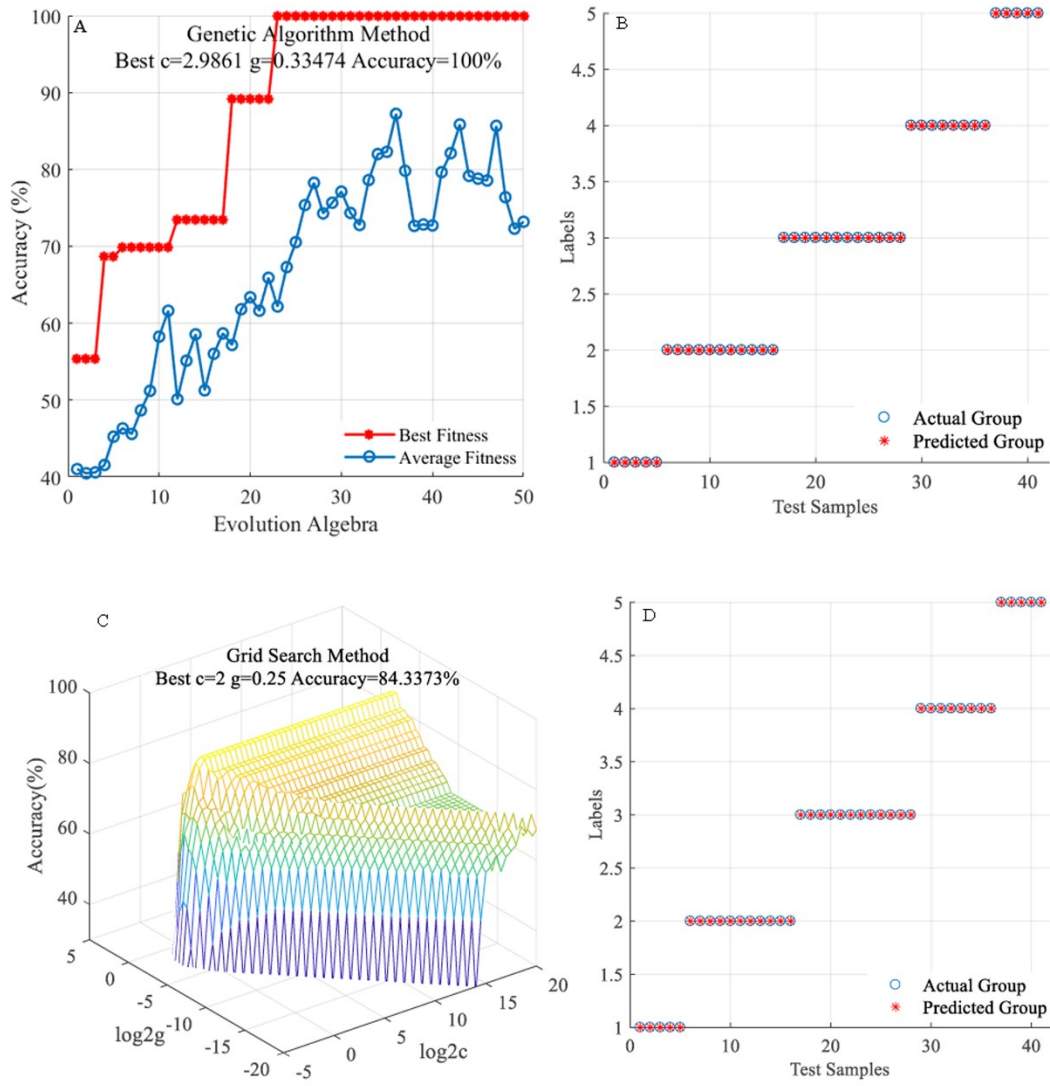
SFig. 3

(Note: This figure shows a simple illustration for explaining the operation principle of k-NN. As seen in this figure, when k was set as 3, Euclidean distance would determine three samples which were adjacent to this unknown samples from calibrated set. This unknown samples should be assigned as red class, because two samples were used to vote in the circle.)



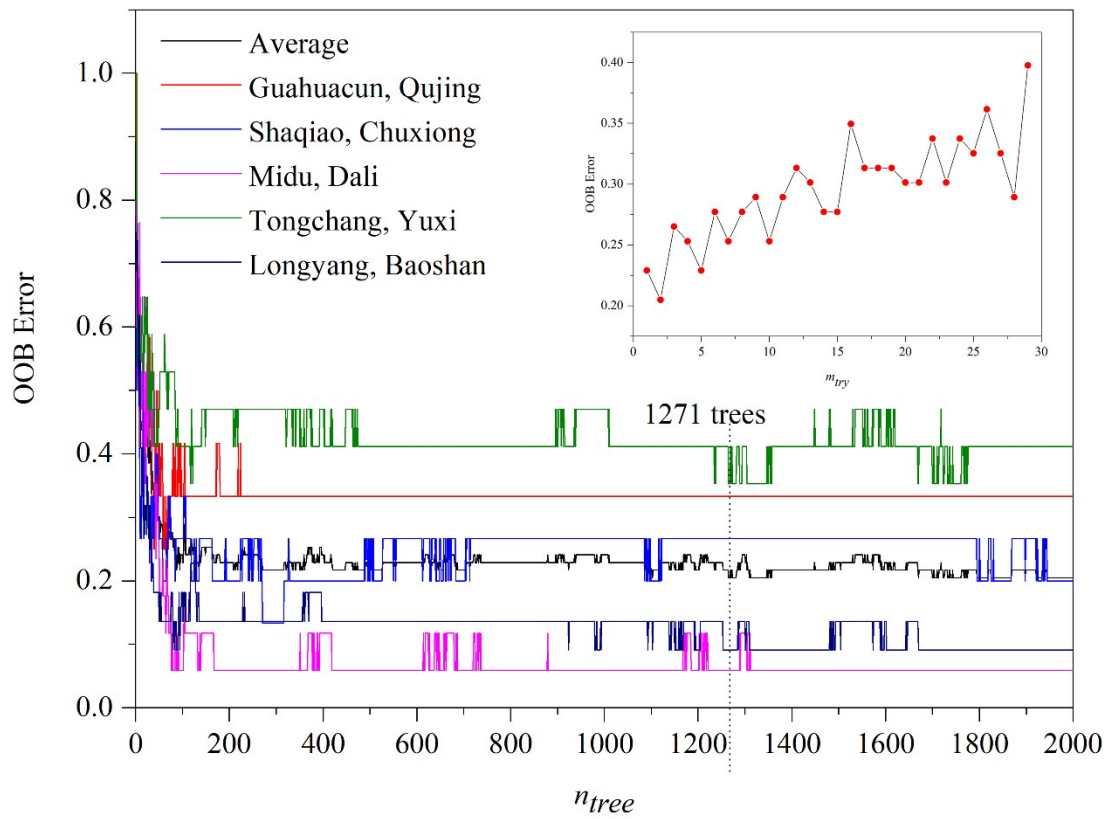
SFig. 4

(Note: According to 95% confidence ellipses of PCA scatter plots, a total of 11 samples were regard as outliers because they are out of this region. These samples were removed in the next analysis.)



SFig. 5

(Note: This figure shows searching process of important parameters and the accuracy of test samples for SVM origin discrimination. The best c and g were screened based on genetic algorithm and grid search method according to the lowest cross-validation accuracy.)



SFig. 6

(Note: This figure shows searching process of n_{tree} and m_{try} for RF origin discrimination.

The best n_{tree} and m_{try} 1271 and 2, respectively, according to the lowest OOB error.)