|  |  |  |  |
| --- | --- | --- | --- |
| Name | Category | Brief Description | Botanical Example |
| Discriminant Analysis | Supervised Learning | Assumes each group follows a multivariate normal distribution and uses this to calculate functions that separate between the groups. | The original method was described by Fisher using a botanical dataset (*Iris* dataset) (1936).  Blasco et al. (2009) separated pomegranate aril images for quality control purposes |
| K-Nearest Neighbour | Supervised Learning | Measures the distance of an unknown sample to each known sample and classifies it to the closest group. | None successful to our knowledge. |
| CART: Classification and Regression Trees | Supervised Learning | Functions in a manner similar to a dichotomous key, where features are used to separate samples at each node. | Hladik and Alber (2014) successfully separate marsh vegetation types. |
| Random Forest | Supervised Learning | In essence a repetition of multiple CARTs, with the membership for unknown samples selected through a majority vote. | Benito Garzón et al. (2008) used Random forests to model climate change impact on the distribution of forest species. |
| Support Vector Machine | Supervised Learning | Support Vector Machines separate groups by establishing a boundary plane between them. | Lu et al. (2012) separated mature *Camellia* L. species using fully mature leaves. |
| Artificial Neural Networks – Multilayer Perceptron | Supervised Learning | Inspired by biological neural networks, they are a series of interconnected elements organised in different layers - most common setup including an input layer, a hidden layer, and output layer. | First used for plant identification by Clark and Warwick (1998) to identify species within the genus *Lithops* N.E.Br.. |
| Naïve Bayes Classification | Supervised Learning | A probabilistic approach, based on a straightforward implementations of Bayes’ theorem – during training conditional probabilities are established which are then used to classify unknown samples. | Ornella et al. (2010) successfully used Naïve Bayes for separation of maize cultivars using molecular data. |
| Cluster Analysis  (incl. k-means) | Unsupervised Learning | A large collection of techniques based on different algorithms. The overarching aim is to divide data into similar groupings. | Extensively used in exploratory data analysis stages. Gardner et al. (2012) used agglomerative clustering to explore the differences in morphology between *Actaea racemosa* L. populations. |
| Artificial Neural Networks – Self-Organizing Map (SOM) | Unsupervised Learning | Similar to the supervised version of artificial neural networks, distributing samples in multidimensional space. | The performance of self-organising maps was demonstrated by Clark et al. (2009) and Clark et al. (2017) using *Tilia* L. species. |