TDF\_algorithm classification in R

**Introduction**

TDF\_algorithm classification is a collaborative application of the SC-R: a supervised classification methodology at <https://github.com/SVMendoza/SC-R>.

This R *TDF\_algorithm classification* code contains the functions for land cover classification for the identification of tropical dry forest, applied to a case in northern Colombia, in the area of potential distribution (PDA) of the TDF in the Department of Córdoba. For the use of the code, a shape file containing the control points and a raster file with the bands to be used for the supervised classification is required. In this case we used Sentinel 2 bands B2, B3, B4, NIR and NDVI with 10m resolution.

**Algorithm processing**

Diagrama

Descripción generada automáticamente

The *TDF\_algorithm assembly* code contains 6 scripts “FPrincipal.r”, “Preparedata.r”, “Models.r”, “CombineRF.r”, “ConfuTable.r” and “predictNNMisc.r”.

FPrincipal.r: contains the main code that calls all the necessary functions contained in the other scripts. It also contains the functions to do the assembly of the raster generated by each algorithm based on the pixel mode values. It also contains functions for filtering patches smaller than 1 ha using the 8-neighbor rule and 100 pixel threshold.

Preparedata.r: contains the functions for the normalization and extraction of the raster data (raster\_PDA.tif) according to the control points (ctrl\_point.shp), as well as the preparation of the training and validation data.

Models.r: contains the functions of the four classification algorithms, Random Forest (RF), Support Vector Machine with polynomial kernel (SVM poly), Support Vector Machine with radial kernel (SVM Radial), and neural networks (Nnet). Each algorithm generates its own raster and confusion table with the metrics precision, recall, and F1-score.

ConfuTable.r: contains the functions for the realization of confusion tables and the metrics precision, recall, and F1-score for each raster.

predictNNMisc.r: function for the elaboration of contingency tables and precision estimation, reclall and f1 score.

The randomForest package (Liaw and Wiener, 2022) was used, applying 1,000 decision trees, each evaluated by voting for the final classification. The SVM poly and SVM Radial algorithms were executed with the svm() interface of the e1071 package (Meyer et al., 2023), while the neural networks were classified with the nnet library (Ripley and Venables, 2023).

The code arguments contain the same structure SC-R, OPEN, name.shape, file.img, name.CLASES, n.core, propVal, sel.n, ndt and SAVE.

Dat.todos corresponds to the data extracted from the image, based on the shape of control points. From these the training and validation data are obtained.

mRF is the classification model based on Random Forest, mPoly is based on Support Vector Machine (SVM) with polynomial kernel, mRad is based on SVM with radial kernel and mnet is based on neural network.

**References**

Liaw, A., Wiener, M., 2022. randomForest: Breiman and Cutler’s Random Forests for Classification and Regression. https://doi.org/10.32614/CRAN.package.randomForest

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., Lin, C.-C., 2023. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. https://doi.org/10.32614/CRAN.package.e1071

Ripley, B., Venables, W., 2023. nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models. https://doi.org/10.32614/CRAN.package.nnet