The identification method of tree species using UV-VIS-IR technology and deep learning methods. Case study - Indepedenta Forest.

Gabriel Murariu
European Center of Excellence for the Environment
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
Gabriel.Murariu@ugal.ro

Adrian Murariu
Department of Physics, Imperial College London, Kensington, London SW7 2AZ, UK
adrianmurariu@yahoo.com

Mihai Daniel Dragu
European Center of Excellence for the Environment
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
dragumihaidaniel@gmail.com

Dan Munteanu
Department of Computer and Information Technology
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
dan.munteanu@ugal.ro

Lucian Dinca
the National Institute of Research and Development in Forestry "Marin Drăcea" – Brasov, 13, Closca Street,,
dinka.lucian@gmail.com

Ciprian Vlad
Department of Computer and Information Technology
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
ciprian.vlad@ugal.ro

Alexandru Dorosencu
Danube Delta National Institute for Research and Development, Tulcea ,
alexandru.dorosencu@ddni.ro

Marian Tudor
Danube Delta National Institute for Research and Development, Tulcea ,
mtudor1971@gmail.com

Silvius Stanciu
Department of Technology Transfer and Training
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
Silvius.Stanciu@ugal.ro

***Abstract*— In this paper we present a comparative analysis between statically methods based on PCA analysis and an expert system built to identify tree species based on their GC-FTIR and UV-VIS-IR spectra. We also present a spectroscopic analysis to evaluate the relevance of each type of sample set on which recognition of an unknown sample is based.**

Keywords*— *statically methods, PCA analysis, deep learning methods

I. INTRODUCTION

The Environmental monitoring and especially prevention of degradation is currently a major challenge. Soil degradation, reduction of forested areas, global warming, etc., all generate effects difficult to assess in the future. In this respect, any effort to identify changes in environmental monitoring and ground is welcome Lately, a great emphasis has been placed on inventorying and monitoring the dynamics of vegetation. in this paper we briefly present the identification methods used to differentiate and identify tree species.

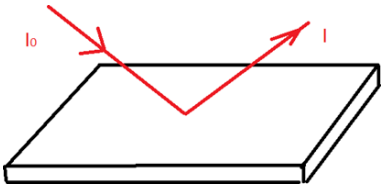

Fig. 1. The reflectance measurement

II. THEORETICAL ASPECTS

The identification of tree species by optical methods is at present a desideratum. In fact, it is intended to be able to identify plant and tree species in real time based on aerial images taken from UAVs or planes.

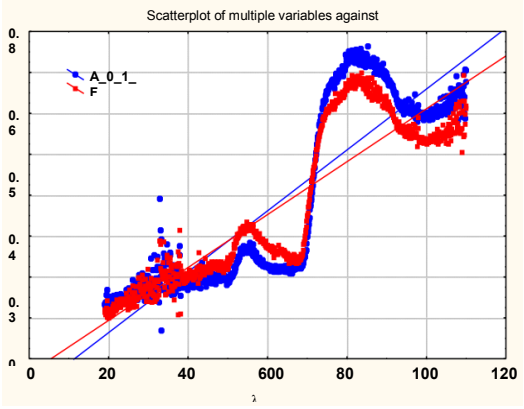

Fig. 2. The reflectance spectra for maple leaf sides

Nowadays, procedures based on data analysis on the computer have become the best way to facilitate the interpretation of spectral data. Among the different methods, infrared spectrometry and UV-VIS spectrometry are popular and versatile tools for the identification and structural analysis of complex compounds [1-2]. Since spectral, many species are difficult to distinguish in the visible field, it results that the strongest applied techniques are needed to identify them. The utility of UV-VIS-IR spectroscopy is as follows.

This reflection coefficient generally depends on the wavelength. In figure 2 is presented the reflectance for a maple leaf for each side. As can be seen from Figure 2, in the visible field, i.e. between 400 nm and 700 nm, the spectra cannot be easily distinguished.

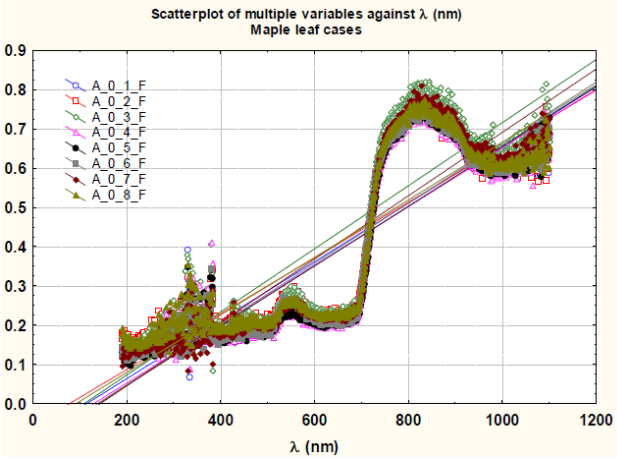

Fig. 3. The reflectance spectra for a set of maple leaf sides

If we wish to keep a data base with these UV-VIS-IR spectra, the specimens of the tree leaves belonging to the same species form a family of curves that show certain differences (Fig. 3). As can be seen from Figure 3, in the visible field the differences between spectra cannot be easily distinguished.

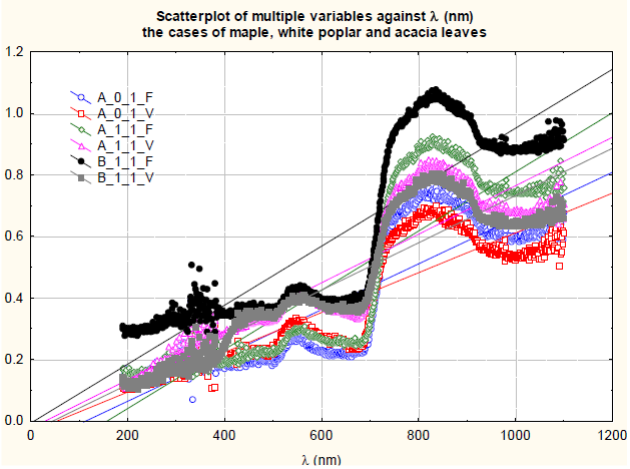

Fig. 4. The reflectance spectra for a maple (sample A_0), white poplar (sample A_1) and acacia (sample B_1) leaf sides

The algorithm considered [3-4]. was applied recursively starting with the first three spectral images. After investigating and obtaining the clusters following the process described below, a set of conventional colours is chosen and the map is coloured using these conventional hues.

Preliminary processing [4-6]:

a) Rows are observations (individuals) and columns are variables

b) Any missing value in the data must be removed or estimated.

c) The data must be standardized (i.e., scaled) to make variables comparable. Recall that, standardization consists of transforming the variables such that they have mean zero and standard deviation one

d) The classification of observations into groups requires some methods of statistical analysis as Pearson correlation and PCA analysis. In this paper, we present, by comparison, two classical methods: statistical approach vs deep learning technique.

SVM (Support Vector Machines) is a technique in machine Learning and Data Mining used for supervised learning. Supervised Learning means obtaining the parameters of a function (called classifier) that maps the input to an output based on training data that contains many examples of input and output. The set of examples is called training data.

The idea of SVM is to find a hyperplane which maximizes the margin, separating the positive from negative examples from input data. The margin between two classes represents the longest distance between closest data examples from those classes. [1] The generalization of the classifier is better (hence the classification error is lower) when the margin between is larger. Support vectors are the nearest points from the two classes considered. The optimal hyperplane is obtained based only on support vectors and not by other points.

Hyperplane term means that the dimensionality of the plane is by one less than the dimensionality of the entire space R^n. A point is a hyperplane in R; a line is a hyperplane in R^2; a plane is a hyperplane in R^3, and so on [2].

Kernel trick, with the name taken from the use of kernel functions which maps the attribute space from a lower dimensional to a higher dimensional space. The coordinates in the higher dimensional space are not computed for the data, the inner product of the point in lower dimension gives the distance between point in the higher dimension [7-8].

SVM was designed for binary classification. Multiclass classification problem is divided into multiple binary classification problems [8, 9]. There are multiple methods for obtaining binary classification (one versus all, one versus one, Directed acyclic graph SVM, Error-correcting output). The approach used in this article is one versus one in which multiple binary classifiers are used to distinguish between every pair of classes. Advantage of SVM is that it can be used with both linear and nonlinear data (by using kernel method). The disadvantage is that the model is not self-

explanatory (but it can be explained by the use of decision trees [10].

III. Experimental Rig

In this work, a series of 24 leaf set of each tree species from the independent Forest were considered. The UV- VIS-IR spectrometer was of Elmer Perkins lambda type with good performances between 190 nm (UV) to 1190 nm (IR). Each sample was scanned twice, for each side, and the scanner results were recorded in a special database with the spectral record.

IV. Experimental Results

At first we considered the methods of statistical analysis, taking the first 3 studied species of trees: maple, white poplar and acacia.

Considering preliminary statistical approaches, for example, for the first 9 samples of maple and white poplar.

Correlations (A_0_1-9_A_1_1-10_B_1_1-12.sta)
Marked correlations are significant at p < .05000
N=911 (Casewise deletion of missing data)

Variable	A_1_1_F	A_1_2_F	A_1_3_F	A_1_4_F	A_1_5_F	A_1_6_F	A_1_7_F	A_1_8_F	A_1_9_F	A_1_10_F
A_0_1_F	.9959	.9963	.9965	.9959	.9963	.9959	.9962	.9963	.9966	.9964
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_2_F	.9967	.9960	.9968	.9958	.9966	.9954	.9954	.9963	.9962	.9965
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_3_F	.9952	.9949	.9962	.9948	.9957	.9955	.9958	.9963	.9956	.9959
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_4_F	.9946	.9948	.9961	.9948	.9953	.9958	.9957	.9964	.9953	.9962
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_5_F	.9960	.9954	.9964	.9954	.9964	.9963	.9961	.9967	.9959	.9963
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_6_F	.9962	.9961	.9964	.9961	.9962	.9963	.9960	.9967	.9964	.9967
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_7_F	.9973	.9969	.9969	.9964	.9973	.9965	.9966	.9968	.9971	.9969
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_8_F	.9961	.9961	.9968	.9954	.9968	.9961	.9960	.9970	.9962	.9967
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_9_F	.9951	.9952	.9955	.9954	.9951	.9957	.9954	.9963	.9957	.9958
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00

Fig. 5. The Pearson correlation coefficients for maple and white poplar leaf samples

By comparing, for example, the reflection spectra between the leaf faces of the same species, the correlation coefficients sometimes have lower values than in the previous case (Fig. 6).

This clearly leads to the fact that classical methods based on correlation matrices - such as the PCA method - can no longer be used successfully (Fig. 7).

Correlations (A_0_1-9_A_1_1-10_B_1_1-12.sta)
Marked correlations are significant at p < .05000
N=911 (Casewise deletion of missing data)

Variable	A_0_1_V	A_0_2_V	A_0_3_V	A_0_4_V	A_0_5_V	A_0_6_V	A_0_7_V	A_0_8_V	A_0_9_V
A_0_1_V	1.0000	.9942	.9953	.9932	.9955	.9940	.9929	.9942	.9925
	p= ---	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_2_V	.9942	1.0000	.9949	.9930	.9939	.9943	.9931	.9941	.9921
	p=0.00	p= ---	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_3_V	.9953	.9949	1.0000	.9936	.9953	.9944	.9925	.9923	.9924
	p=0.00	p=0.00	p= ---	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_4_V	.9932	.9930	.9936	1.0000	.9923	.9941	.9882	.9919	.9929
	p=0.00	p=0.00	p=0.00	p= ---	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00
A_0_5_V	.9955	.9939	.9953	.9923	1.0000	.9943	.9912	.9914	.9900
	p=0.00	p=0.00	p=0.00	p=0.00	p= ---	p=0.00	p=0.00	p=0.00	p=0.00
A_0_6_V	.9940	.9943	.9944	.9941	.9943	1.0000	.9893	.9915	.9910
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p= ---	p=0.00	p=0.00	p=0.00
A_0_7_V	.9929	.9931	.9925	.9882	.9912	.9893	1.0000	.9939	.9902
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p= ---	p=0.00	p=0.00
A_0_8_V	.9942	.9941	.9923	.9919	.9914	.9915	.9939	1.0000	.9930
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p= ---	p=0.00
A_0_9_V	.9925	.9921	.9924	.9929	.9900	.9910	.9902	.9930	1.0000
	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p=0.00	p= ---

Fig. 6. The Pearson correlation coefficients for maple leaf sample

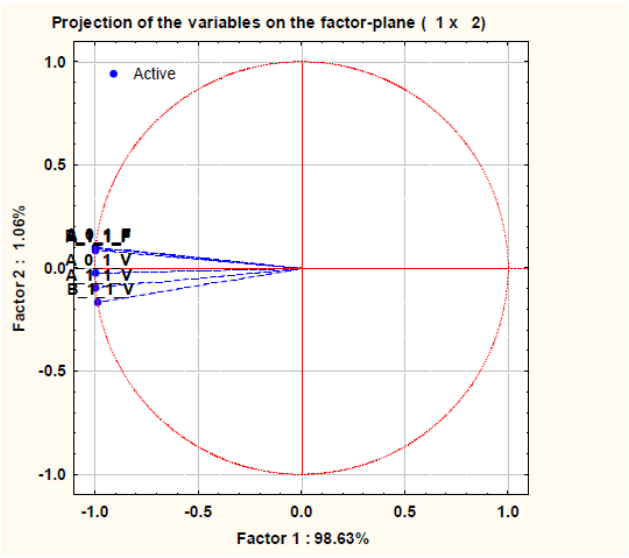

Fig. 7. The PCA statistical analysis for maple and white poplar leaf samples

As it is known, the PCA method follows the grouping of factors that inflate a set of data so as to minimize the data set variation and group by the influence of the data set parameters. In our case, for the spectra of the first three tree species, the PCA method assigns the same facto all 6 data types (Fig. 7).

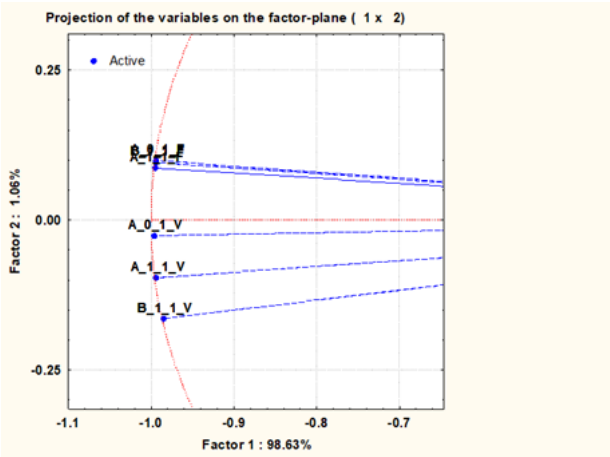

Fig. 8. The PCA statistical analysis or the principal component

A more careful analysis (Fig. 8) highlights the fact that data sets related to the leaves of a species sometimes correlate better with data sets of reflection spectra on the backs of other leaf species.

It was necessary to have an expert identification and discrimination method. Many papers in the literature expose different methods of identification and discrimination using ANN neural networks with flexible structure with a small number of layers of internal memory. These methods will be investigated and analysed in future works.

First of all, it was read the MATLAB database and it was able to visualize in scatter plot the distribution of maple and white poplar samples in the representative space (Fig. 9).

All methods of classification learner have been tested: linear discriminant, quadratic discriminant, linear SVM, quadratic SVM, cubic SVM, fine Gaussian, medium Gaussian and coarse Gaussian, and respectively fine KNN, medium KNN and coarse KNN (Fig. 10).

The better preliminary result is the one obtained for quadratic SVM for which the accuracy is 97.2% and the percentage of false positives is below 5%

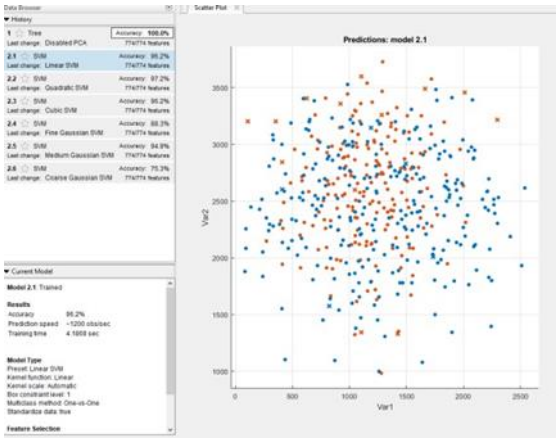

Fig. 9. Preliminary separation procedure between maple and white poplar samples

Fig. 10. Preliminary separation procedure using quadratic SVM method

V. CONCLUSIONS

In this paper we present the results of implementing a machine learning algorithm to identify the species from a UV-VIS-IR spectra database. The results obtained in comparison to reference data are encouraging. The results obtained under laboratory conditions will then be resumed by diffuse spectrum conditions, i.e. obtained in the sunlight. In this way, procedures that prove to be correct and effective at this step will be resumed for the classification of spectra obtained in natural light. Considering the possibilities of improving the algorithm, the research should be continued.

ACKNOWLEDGMENT

This work was supported by Ministry of Research and Innovation project number 4PS, within Sectorial Program: Research to support the development of the capacity to monitor, evaluate and capitalize on natural resources provided by wetlands of international importance in Romania and the Black Sea coastal zone.

REFERENCES

[1]. DF Team, https://data-flair.training/blogs/svm-support-vector-machine-tutorial/

[2]. Alexey Nefedov, *Support Vector Machines: A Simple Tutorial* https://svmtutorial .online/ download. php? file= SVM_tutorial.pdf

[3]. Support Vector Machines for Binary Classification https://www.mathworks.com /help/stats/ support-vector-machines-for-binary-classification.html

[4]. Vapnik, V. and A. Lerner, 1963. *Pattern recognition using generalized portrait method.* Automation and Remote Control. 24, 774-780.

[5]. Aizerman, Mark A.; Braverman, Emmanuel M. & Rozonoer, Lev I. (1964). "*Theoretical foundations of the potential function method in pattern recognition learning*". Automation and Remote Control. 25: 821–837.

[6]. Bernhard E. Boser , Isabelle M. Guyon , Vladimir N. Vapnik, A *Training Algorithm for Optimal Margin Classifiers,* 1992, Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory.

[7] Puscasu, G ; Codres, B Stancu, A Murariu, G, *Nonlinear system identification based on internal recurrent neural networks*, International Journal Of Neural Systems Volume: 19 Issue: 2 Pages: 115-125 Published: APR 2009 [Impact Factor FI (2010) = 2,988] (http:// www.worldscientific.com /doi/abs/ 10.1142/S0129065709001884.

[8] Murariu, G, Dinescu, A.,Murariu, A. G., Condurache-Bota, S., *Comparative discussion on global models For optical properties Of Bi2O3*, Romanian Journal Of Physics, Volume: 61, Issue: 3-4, Pages: 626-637, Published: 2016 [Impact Factor FI (2016) = 1.398] (http://www.nipne.ro/rjp/2016_61_3-4.html)

[9] Mocanu D. G., *The Sociometric analysis of the IInd year students' views – F.E.F.S. Galaţi*, The Annals of the University" Dunarea De Jos" Galati, ISSN-1454-9832 ISSN-L1454-9832, No. 2, Issue 2, 47-58.

[10] Mocanu D. G., *Study about the development of the conditional motor skills at the middle-school level - 6th*

grade boys – by applicative tracks and paths, The Annals of the University" Dunarea De Jos" Galati, ISSN-1454-9832 ISSN-L1454-9832, No. 2, Issue 2, 58-68.