SISSG
OLI e GRASSI ALIMENTARI
INNOVAZIONE e SOSTENIBILITÀ
NELLA PRODUZIONE E NEL CONTROLLO

Giugno 15 - 17 **2022**
PERUGIA Complesso monumentale di San Pietro Borgo XX Giugno n°74
Dipartimento di Scienze Agrarie Alimentari ed Ambientali (DSA3), Università di Perugia

FlavourSpec® & Machine learning the new frontiers in instrumental support for the sensory analysis of virgin olive oils

Tullia Gallina Toschi(1), Alessandra Bendini(1), Ilaria Grigoletto(1), Enrico Valli(1), Cesare Rossini(2), Andrea Giomo.

(1) Dipartimento di Scienze e Tecnologie Agro-Alimentari - Università di Bologna

(2) Lab Service Analytica S.r.l. Anzola dell'Emilia (BO)

Agenda

01 Introduction

02 Flavour Spec®

03 Positioning & Experimental workflow

04 From Raw Data to Descriptive & Predictive Models

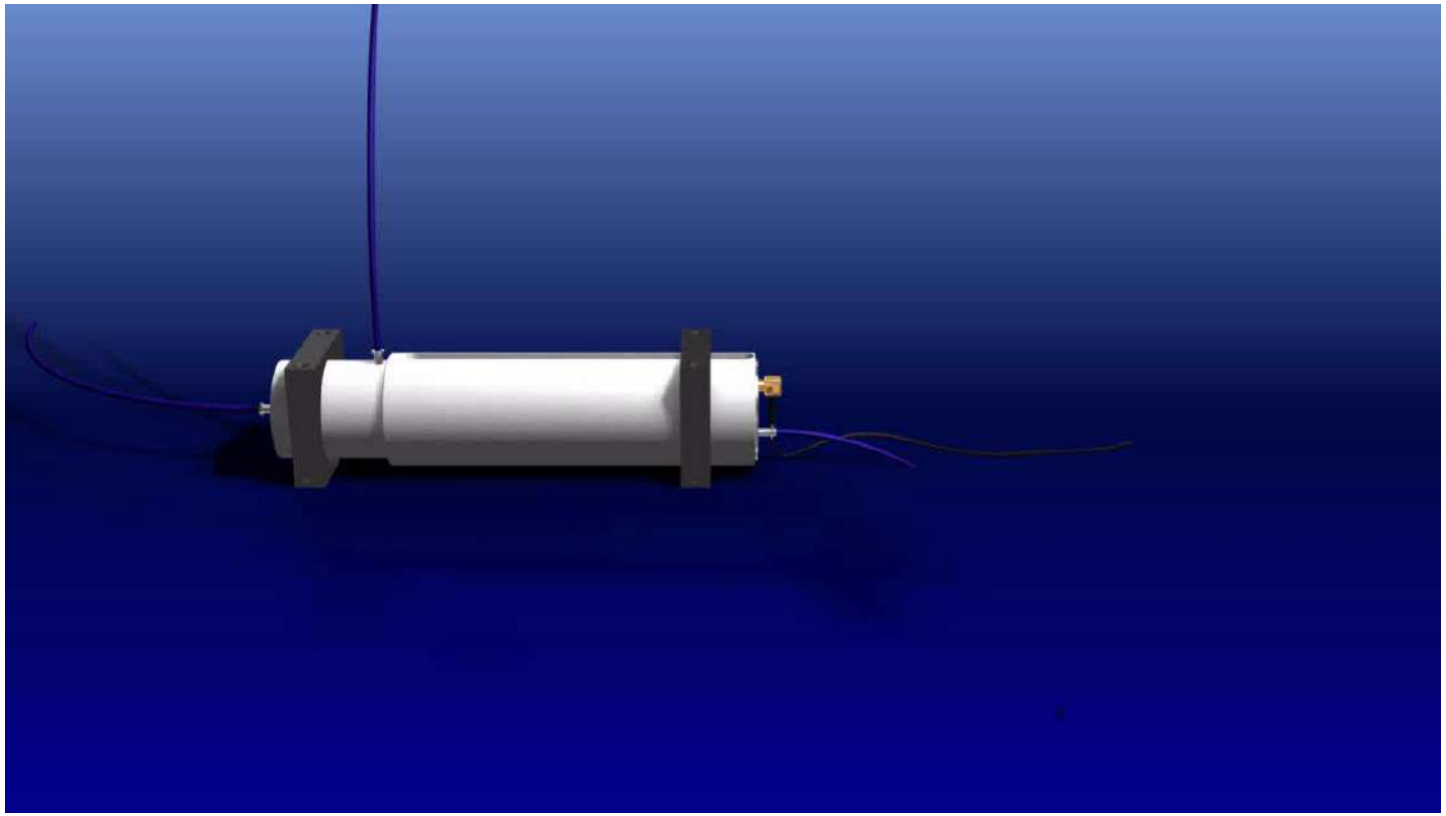
05 Conclusions

Introduction

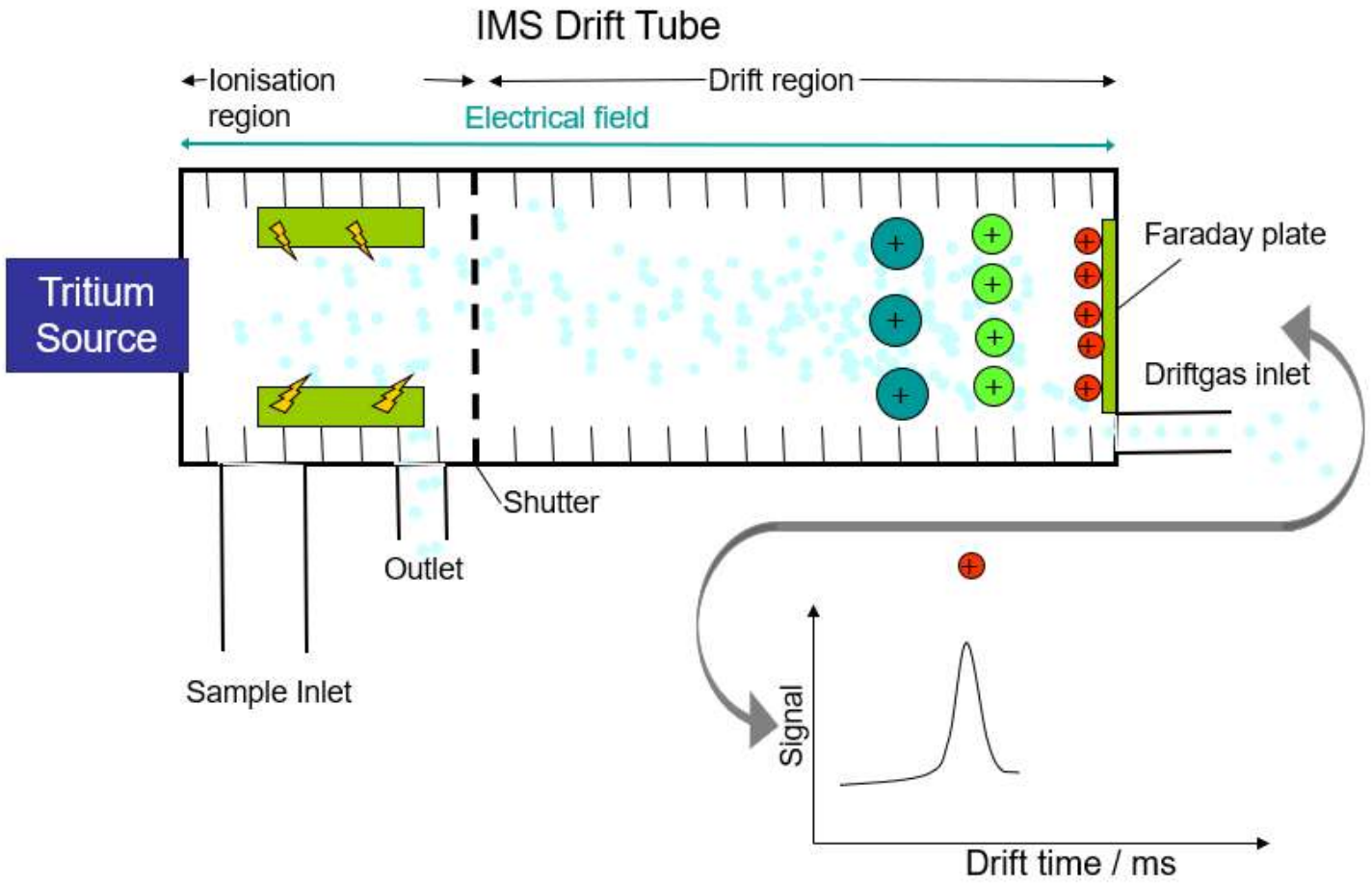
1. The FlavourSpec® (G.A.S. Dortmund, Germany) is an analytical instrument based on gas chromatography coupled with ion mobility spectrometry (HS-GC-IMS).
2. Lab Service Analytica is distributing G.A.S. Dortmund instruments since 2017 .
3. Lab Service Analytica and the University of Bologna started collaborating in 2018 in the frame of OLEUM project .
4. The goal of this collaboration was to develop an HS-GC-IMS Method for the Quality Classification of Virgin Olive Oils to support the Panel Test process.
5. In 2020 , the University of Bologna published an article .

Flavour Spec ®

1. The instrument is a GC with a special detector: **Ion Mobility**



Flavour Spec ®

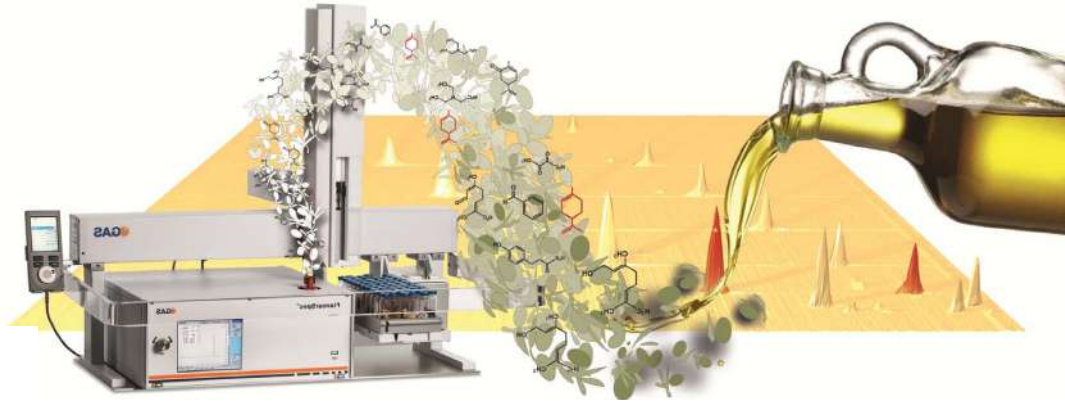


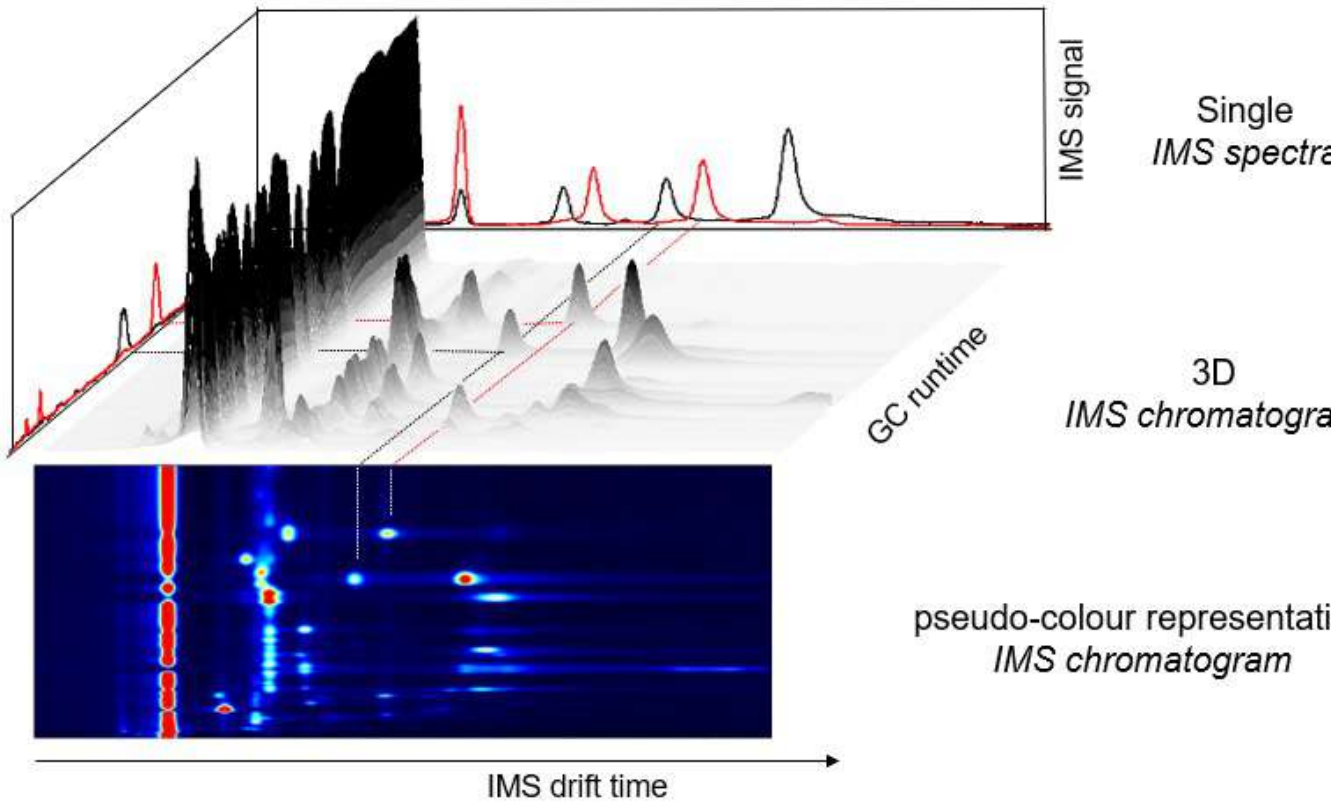
Drift velocity: $v_d = K\,E$ Mobility: $K = \frac{d}{t_d E}$

$$K = \frac{3}{16}\frac{e}{N}\sqrt{\frac{1}{m}+\frac{1}{M}}\sqrt{\frac{2}{k}\frac{\pi}{T}}\frac{1+\Delta}{\pi\, r^2\,\Omega}$$

E	electric field strength	d	drift path length
t_d	drift time	e	unit charge
m	ion mass (analyte)	M	molecular mass (drift gas)
N	number density (drift gas)	k	Boltzmann-constant
T	temperature	r	minimum in the potential curve
Ω	collision integral 1. order	Δ	correction term for approximations

Flavour Spec ®





Single
IMS spectra

3D
IMS chromatogra

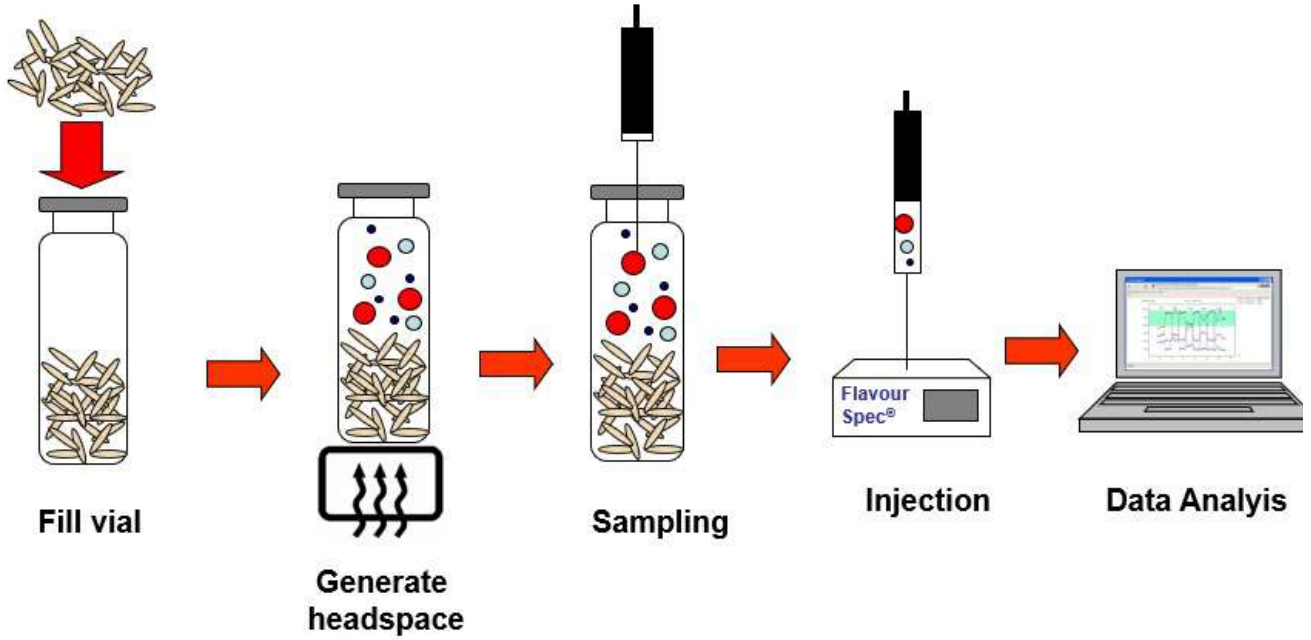
pseudo-colour representati
IMS chromatogram

Protone Affinities		
Aromatic Amines	930.0 KJ/mol	Pyridine
Amines	899.0 KJ/mol	Methyl Amine
Phosphorous Compounds	890.6 KJ/mol	Trimethylphosphate
Sulfoxides	884.4 KJ/mol	Dimethyl Sulfoxide
	853.6 KJ/mol	Ammonia
Ketones	832.7 KJ/mol	2-Pentanone
Esters	821.6 KJ/mol	Methyl Acetate
Alkenes	805.2 KJ/mol	1-Hexene
Alcohols	789.2 KJ/mol	Butanol
Aromatics	750.4 KJ/mol	Benzene
	691.0 KJ/mol	**Water**
Alkanes	543.5 KJ/mol	Methane

Source: Gary Eiceman & Zeev Karpas, *Ion Mobility Spectrometry*, CRC Press, 2005, ISBN 0-8493-2247-2

Protone affinities of various VOCs can be found at the NIST chemistry webbook http://webbook.nist.gov/chemistry/

Positioning & Experimental workflow



foods

MDPI

Article

An HS-GC-IMS Method for the Quality Classification of Virgin Olive Oils as Screening Support for the Panel Test

Enrico Valli [1,2], Filippo Panni [1], Enrico Casadei [1,*], Sara Barbieri [3], Chiara Cevoli [1,2], Alessandra Bendini [1,2], Diego L. García-González [4] and Tullia Gallina Toschi [1,2]

[1] Department of Agricultural and Food Science, Alma Mater Studiorum—Università di Bologna, 47521 Cesena, Italy; enrico.valli4@unibo.it (E.V.); filippo.panni2@unibo.it (F.P.); chiara.cevoli3@unibo.it (C.C.); alessandra.bendini@unibo.it (A.B.); tullia.gallinatoschi@unibo.it (T.G.T.)
[2] Interdepartmental Center for Industrial Agrofood Research, Alma Mater Studiorum—Università di Bologna, 47521 Cesena, Italy
[3] Department of Pharmacy and Biotechnology, Alma Mater Studiorum—Università di Bologna, 47521 Bologna, Italy; sara.barbieri@unibo.it
[4] Instituto de la Grasa (CSIC), 41013 Sevilla, Spain; dlgarcia@ig.csic.es
[*] Correspondence: enrico.casadei15@unibo.it; Tel.: +39-054-7338-121

Received: 17 April 2020; Accepted: 17 May 2020; Published: 20 May 2020

Positioning & Experimental workflow

1. Decisive criteria
2. Individual variance
3. Difficult in alignment

1. Simple to use :

   NO SAMPLE preparation

2. Solid physical chemistry
3. Excellent for untargeted approach (fingerprinting)

1. “Golden Standard”
2. Complete data base of compounds.
3. Excellent for quantitation of single compounds
4. Need of an intermediate step (SPME)

Synopsis

Data analysis

- **Targeted** (focus on known voc)
- **Untargeted** (whole voc exploration)

Learning models

- **Supervised** (known labels)
- **Semi-supervised**
- **Unsupervised** (unknown labels)

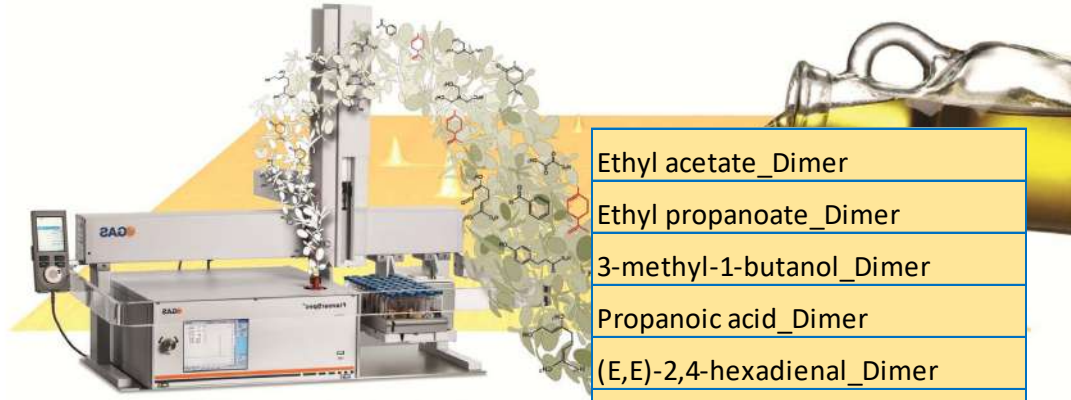
Synopsis

Learning models (classic)

- Quadratic Discriminant Analysis
- Tree Net Classifier
- Support Vector Machine Classifier
- Kernel-SVM Classifier (polynomial)
- Naive Bayes Classifier (gaussian)
- Adaptive Boosting Classifier (adaboost)
- Gaussian Process Classifier
- K-Nearest Neighbor (KNN)
- Logistic regression
- Random forest (criterion: entropy/Gini)

TARGETED

Targeted (visualization)



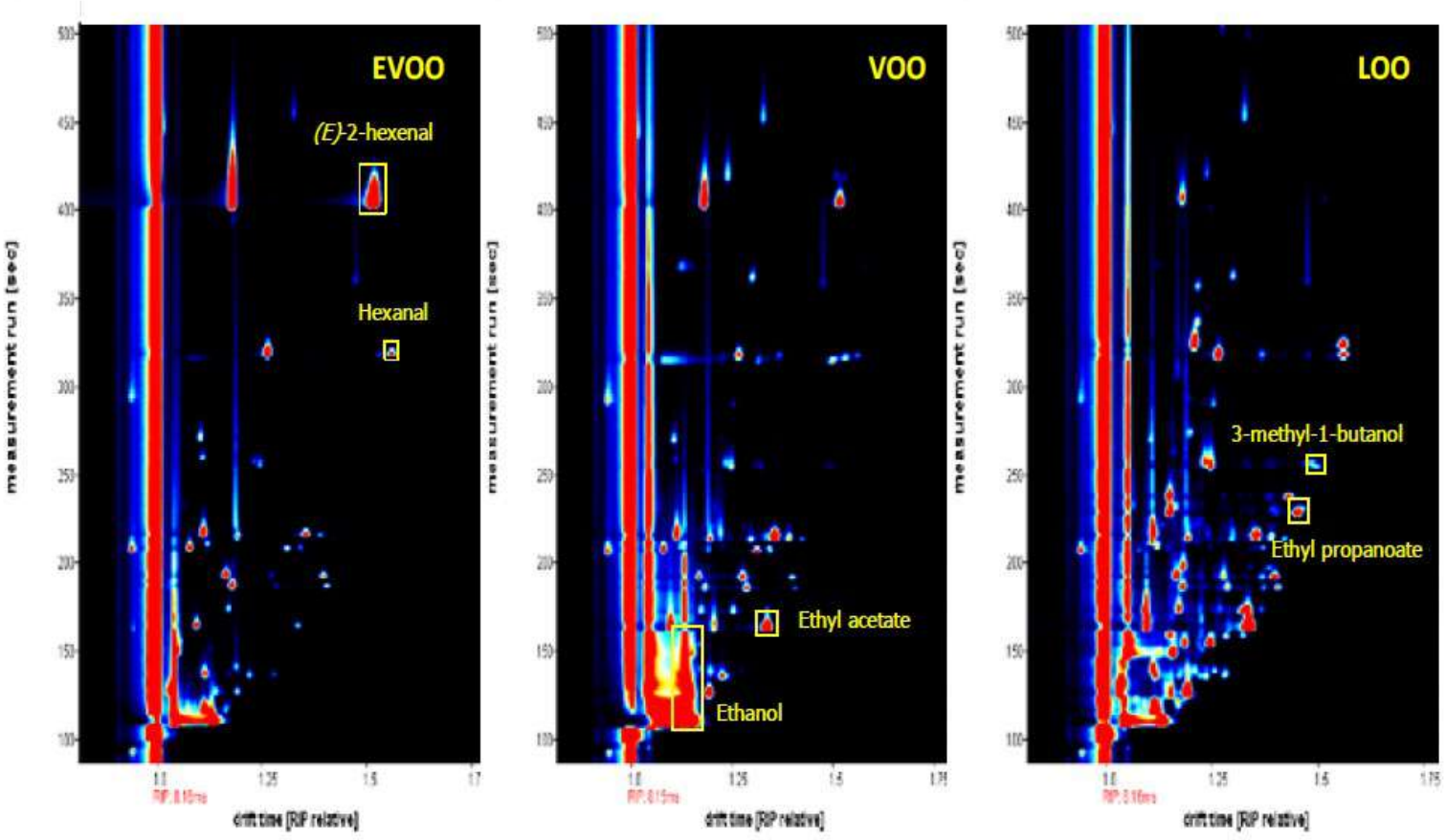
foods

MDPI

Article

An HS-GC-IMS Method for the Quality Classification of Virgin Olive Oils as Screening Support for the Panel Test

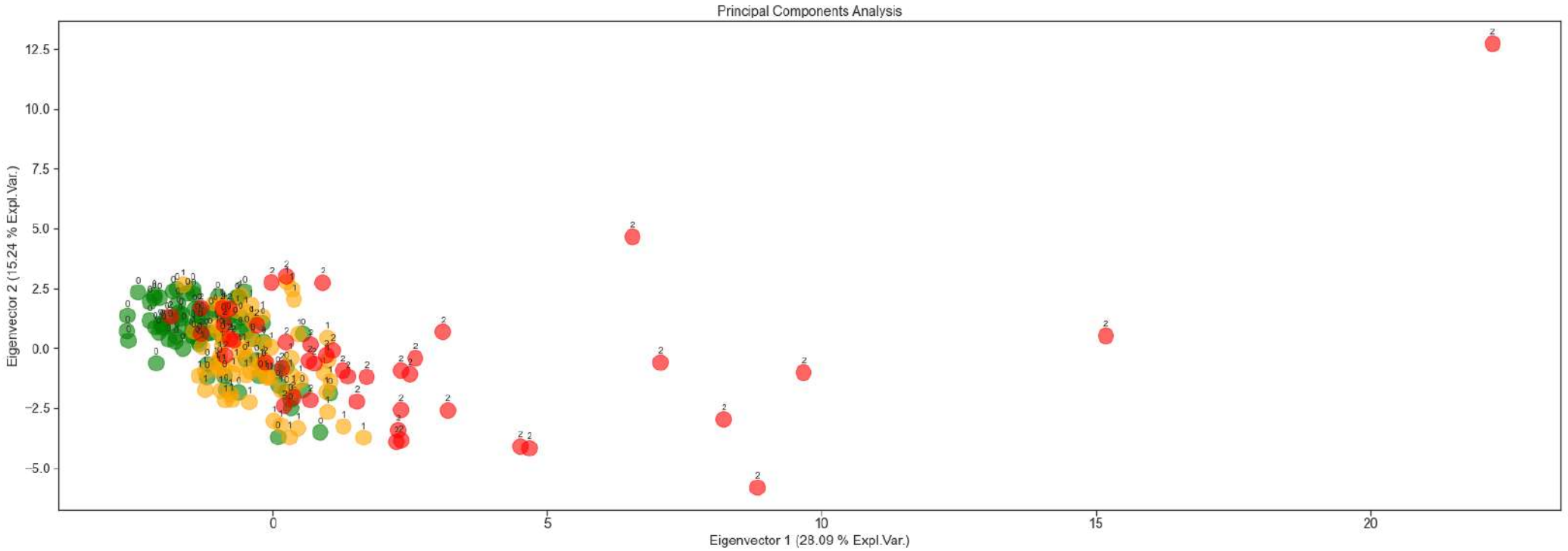
Enrico Valli [1,2], Filippo Panni [1], Enrico Casadei [1,*], Sara Barbieri [3], Chiara Cevoli [1,2], Alessandra Bendini [1,2], Diego L. García-González [4] and Tullia Gallina Toschi [1,2]



Ethyl acetate_Dimer
Ethyl propanoate_Dimer
3-methyl-1-butanol_Dimer
Propanoic acid_Dimer
(E,E)-2,4-hexadienal_Dimer
(E)-2-heptenal_Dimer
6-methyl-5-hepten-2-one_Dimer
Ethanol_Dimer
Acetic acid_Dimer
Hexanal_Dimer
(E)-2-hexenal_Dimer
1-hexanol_Dimer
1-octen-3-ol_Dimer
(Z)-3-hexenyl acetate_Dimer
Nonanal_Dimer
(E)-2-heptenal_Monomer
(E,E)-2,4-hexadienal_Monomer
Ethyl acetate_Monomer
Ethyl propanoate_Monomer
3-methyl-1-butanol_Monomer
(Z)-3-hexenyl acetate_Monomer
Hexanal_Monomer
(E)-2-hexenal_Monomer
1-hexanol_Monomer
Acetic acid_Monomer

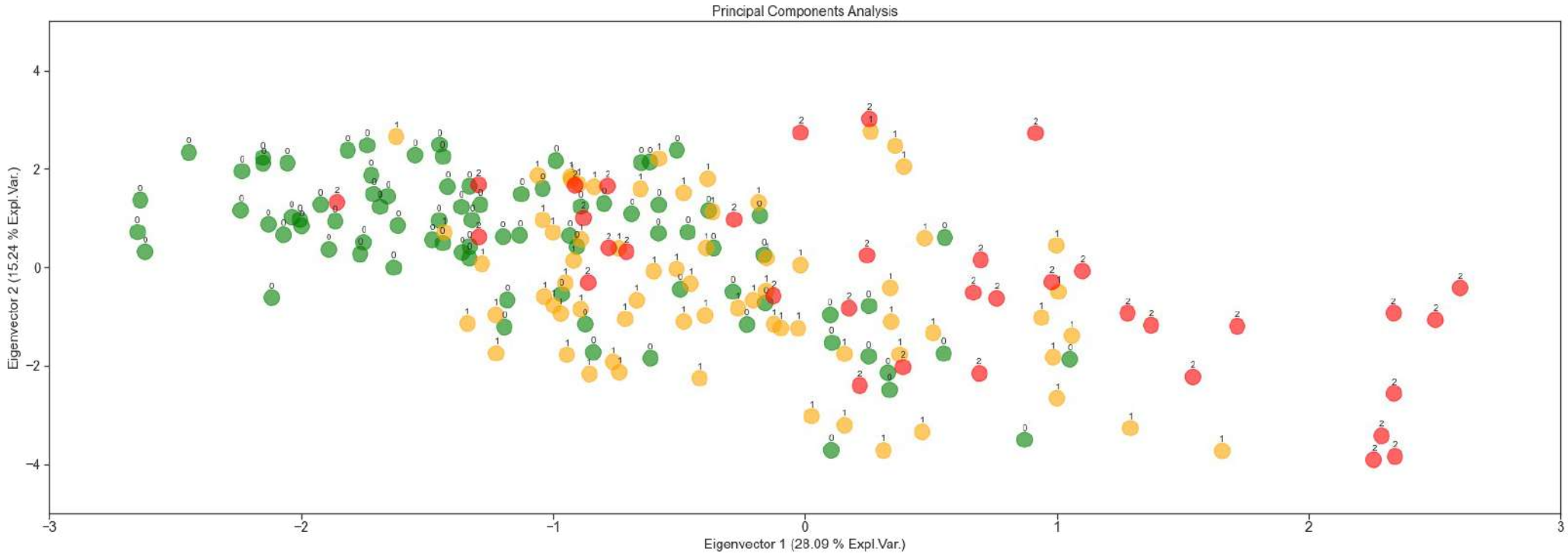
Principal Components Analysis (*full svd*)

evoo
voo
loo

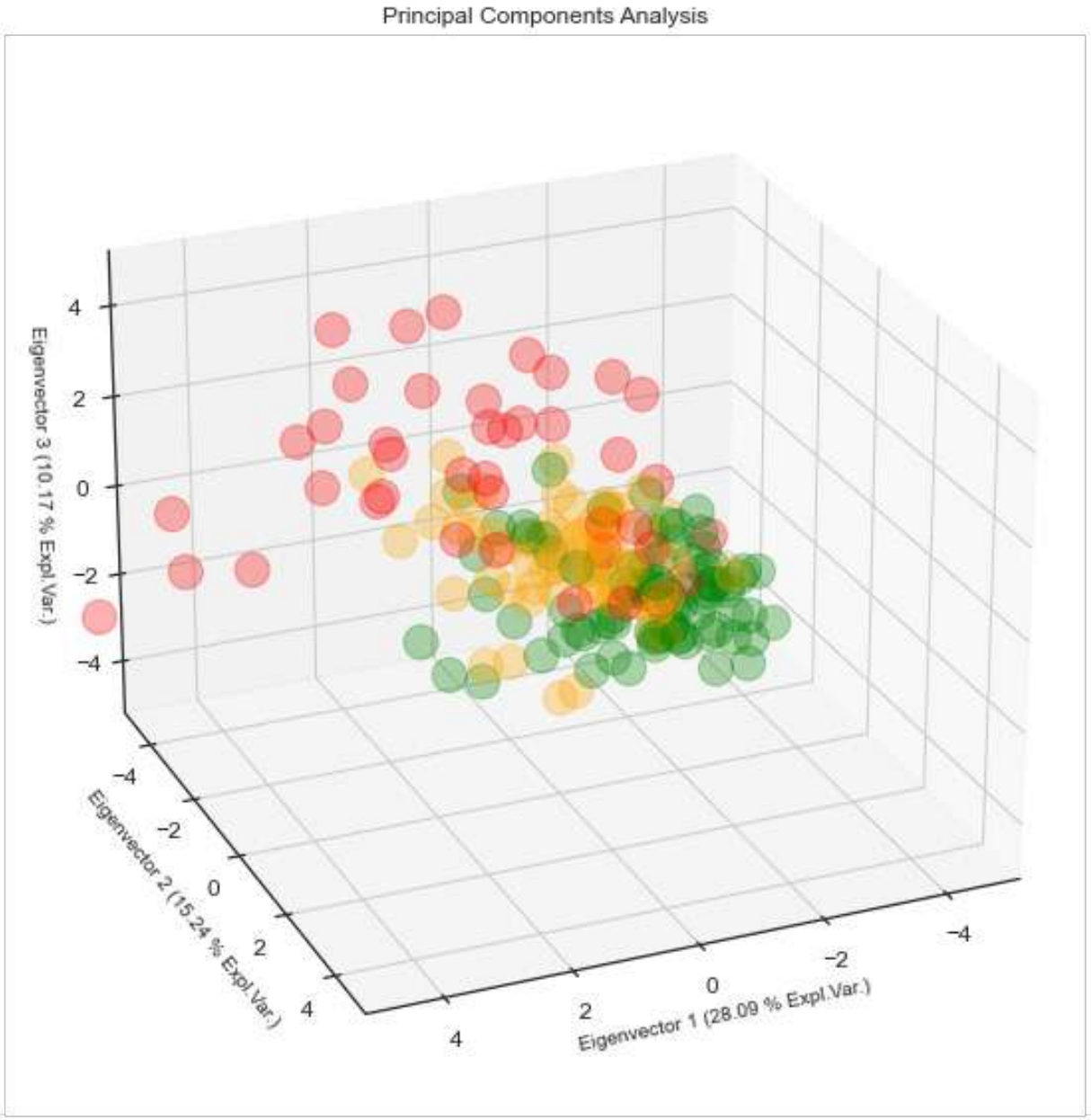


PCA rescaled

evoo
voo
loo



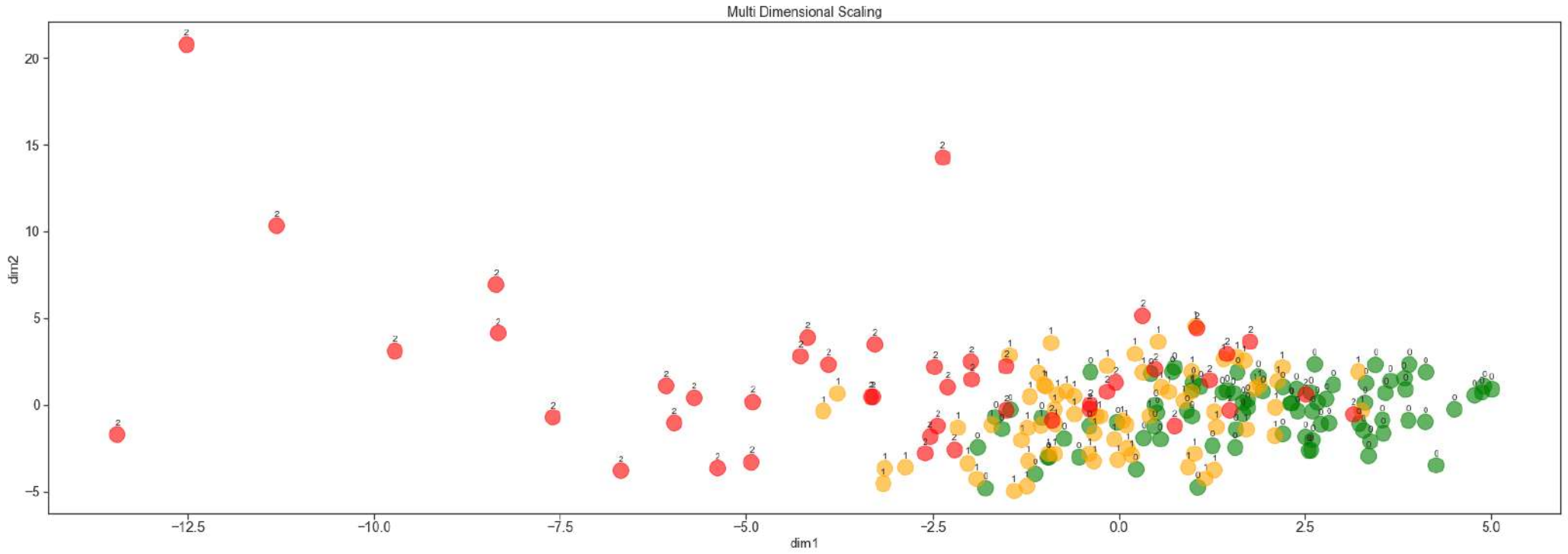
PCA 3D



evoo
voo
loo

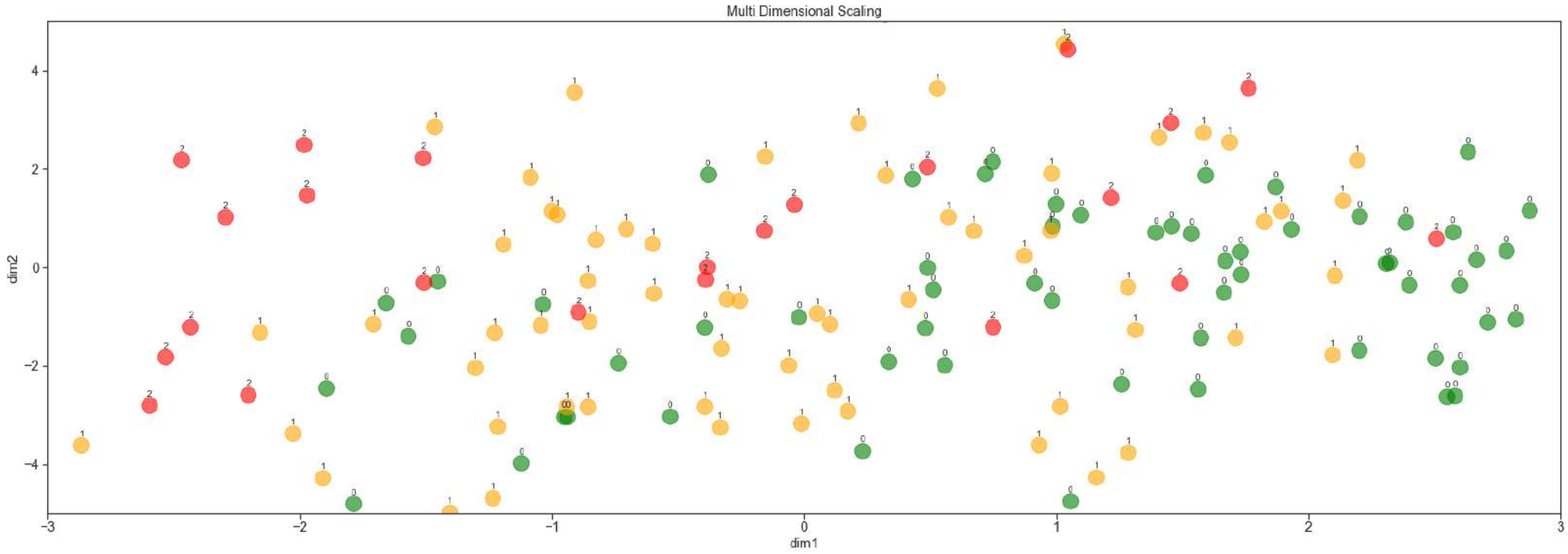
Multidimensional scaling

evoo
voo
loo



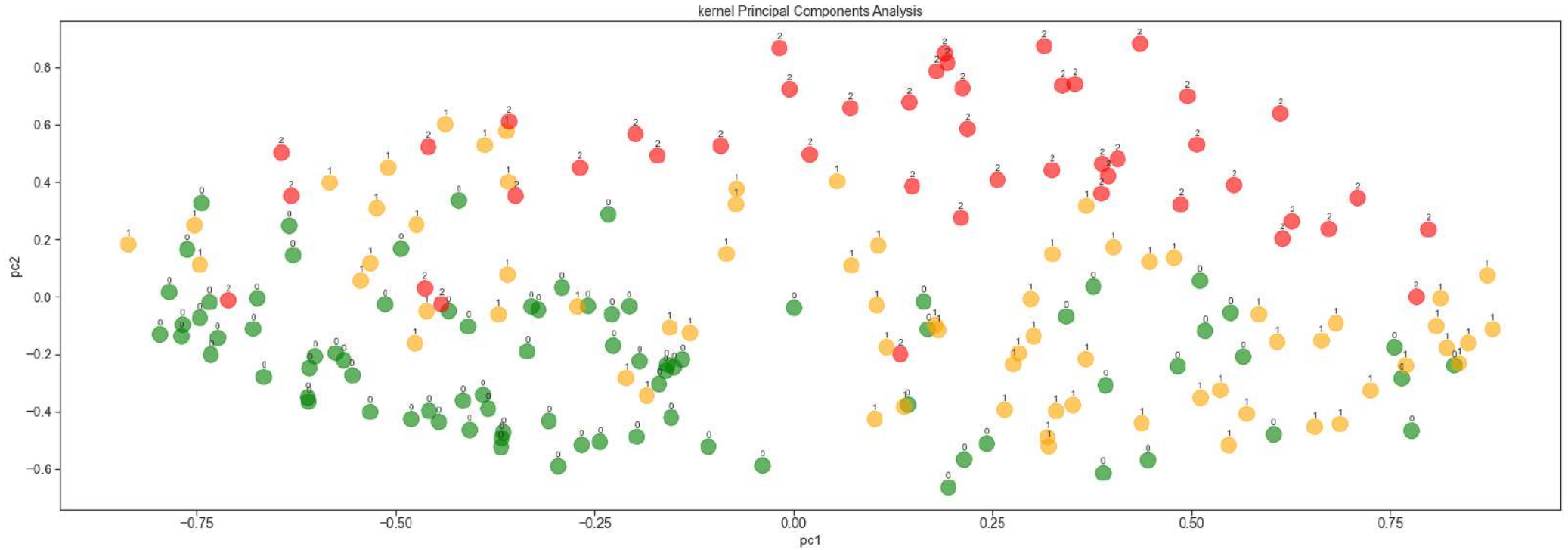
Multidimensional scaling (rescaled)

evoo
voo
loo



Kernel PCA (cosine)

evoo
voo
loo



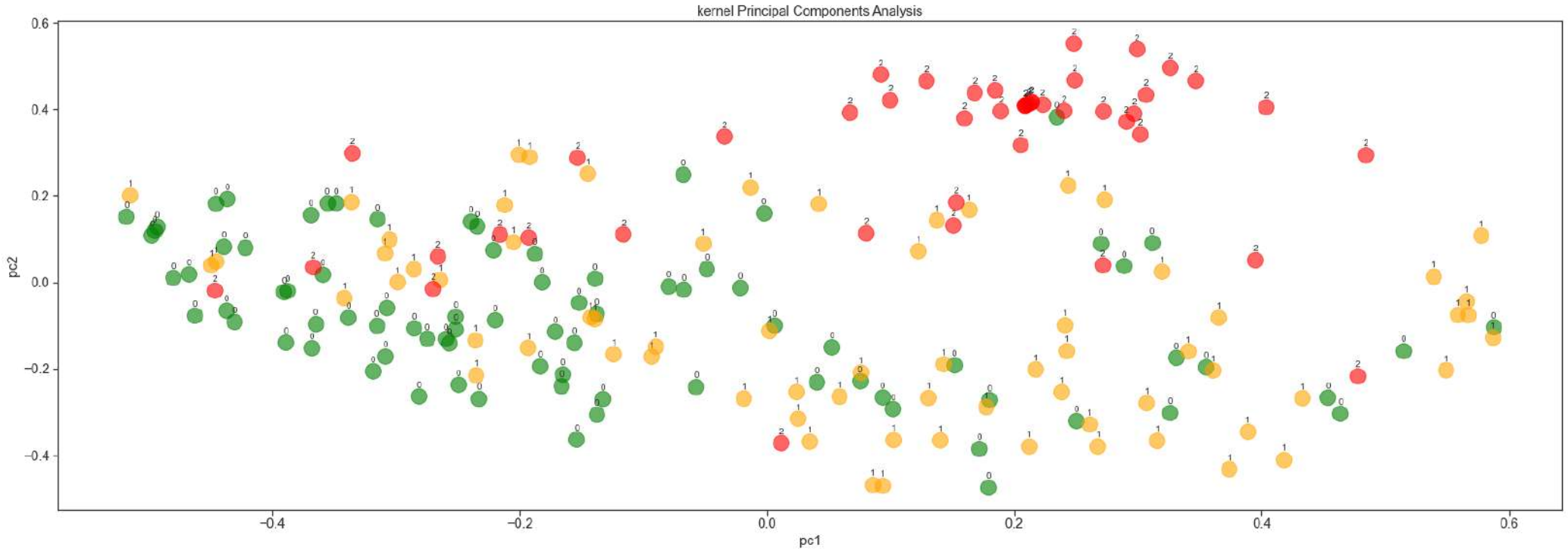
Kernel PCA (rbf)

Radial-basis function kernel (*squared-exponential kernel*)

evoo

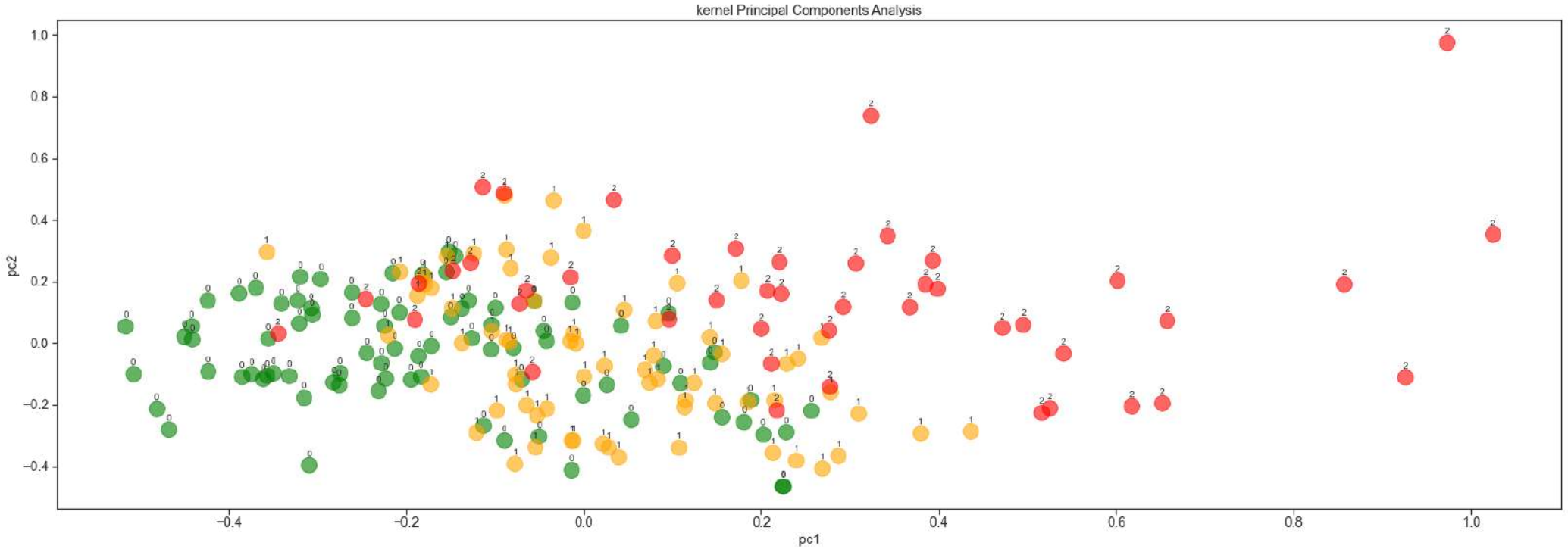
voo

loo

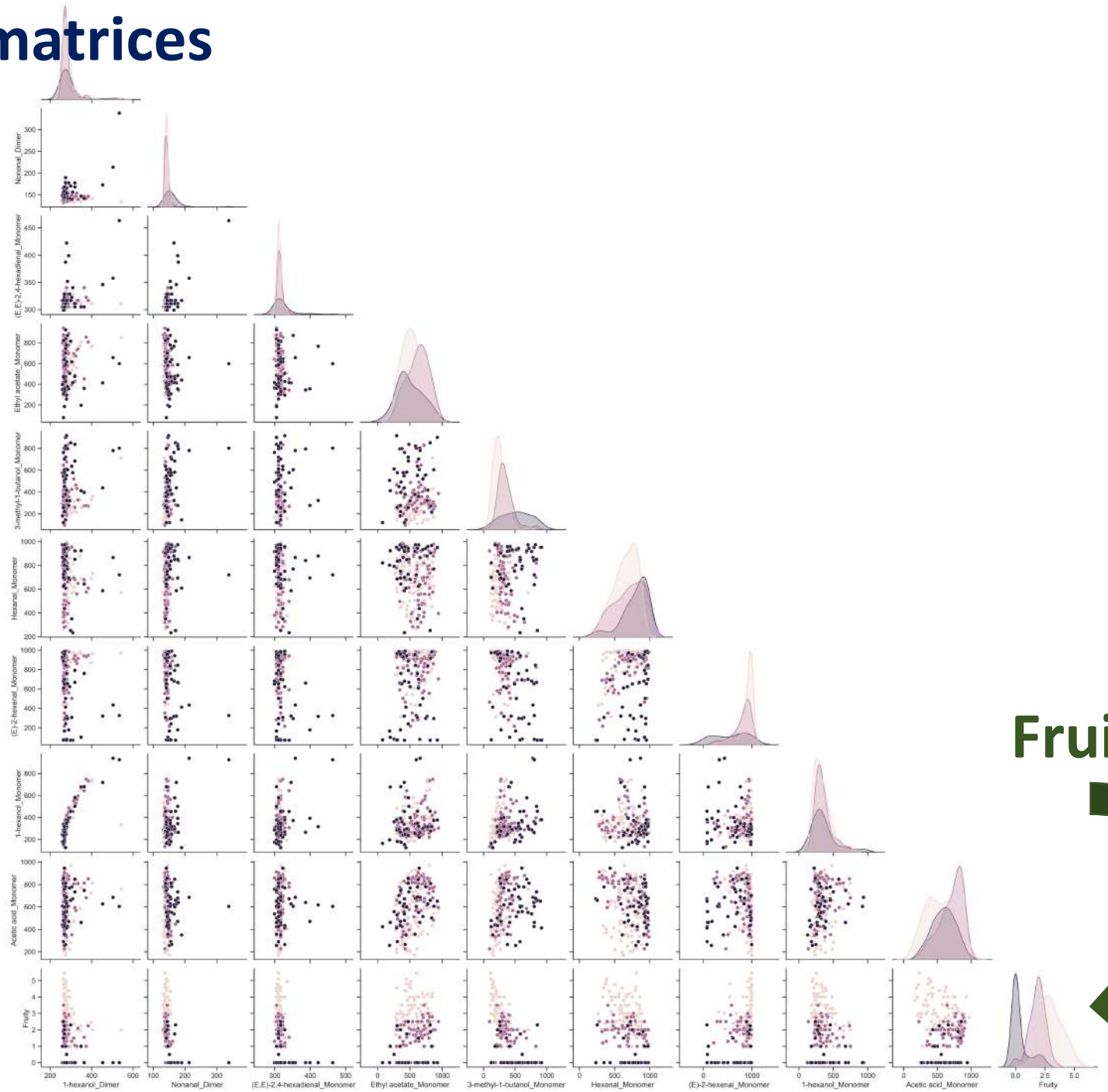


Kernel PCA (sigmoid)

evoo
voo
loo

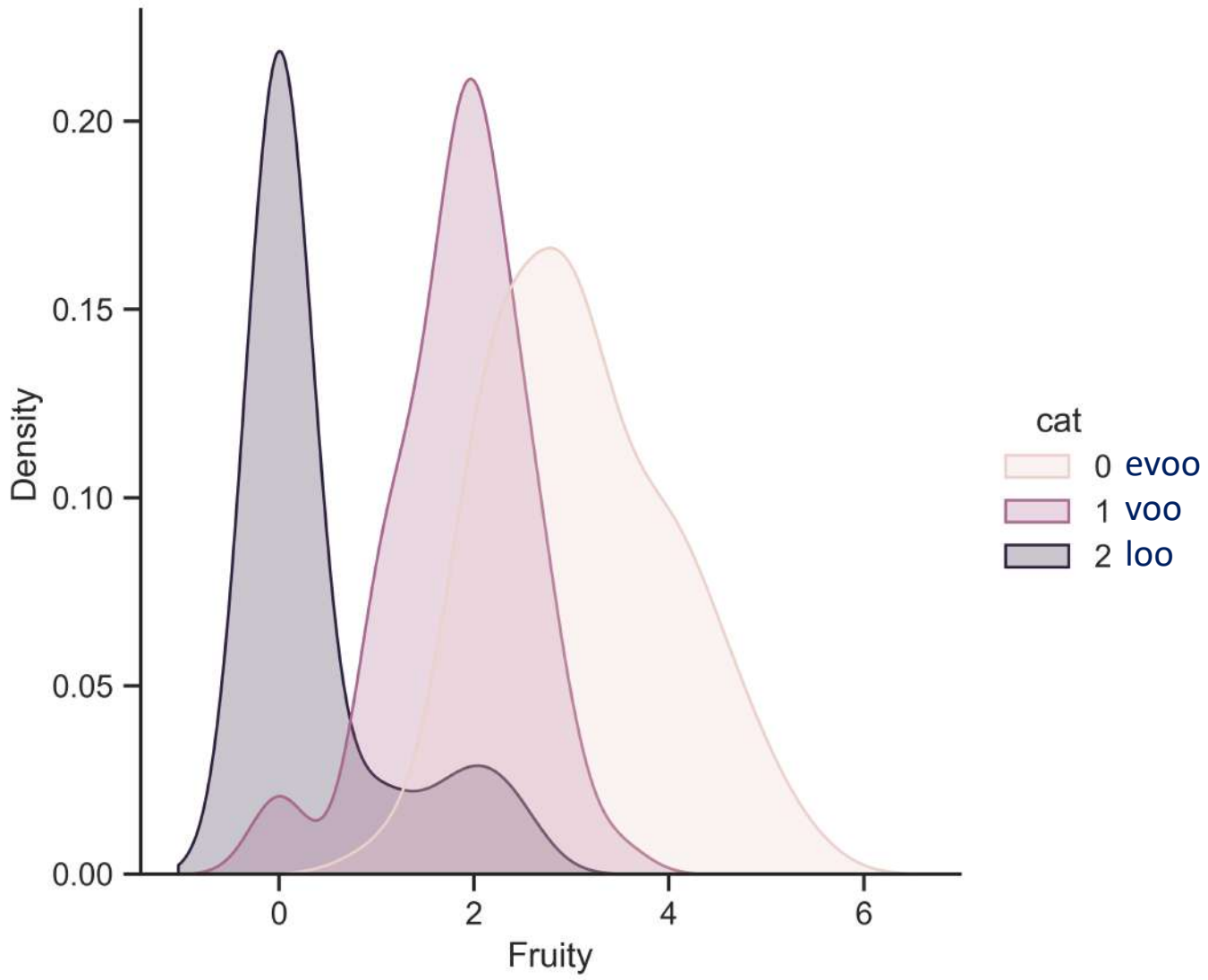


Scatterplot matrices

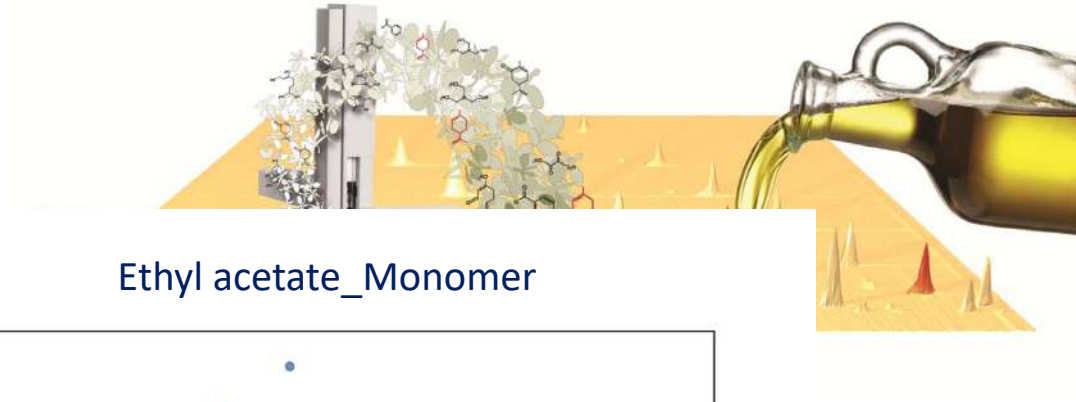


Fruity

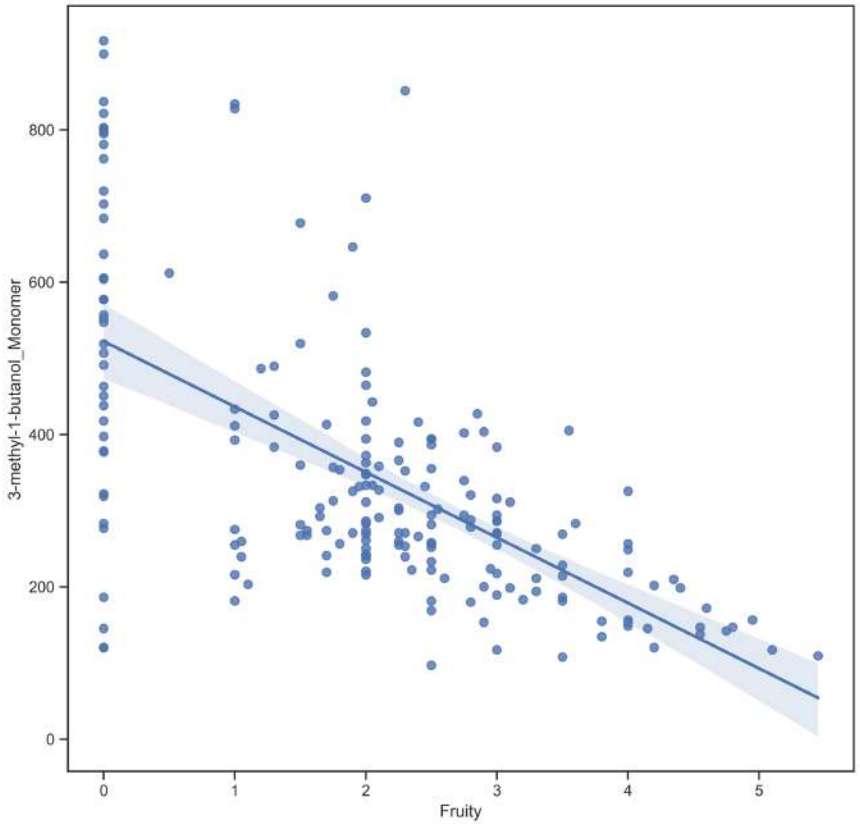
Fruity trend



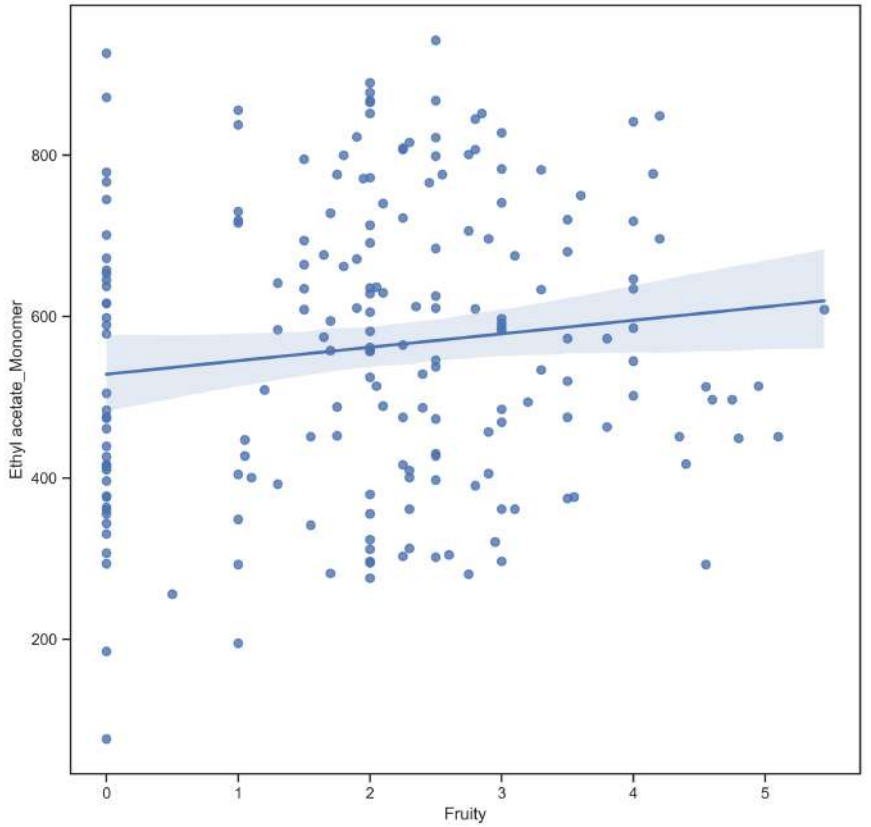
Correlations (?)



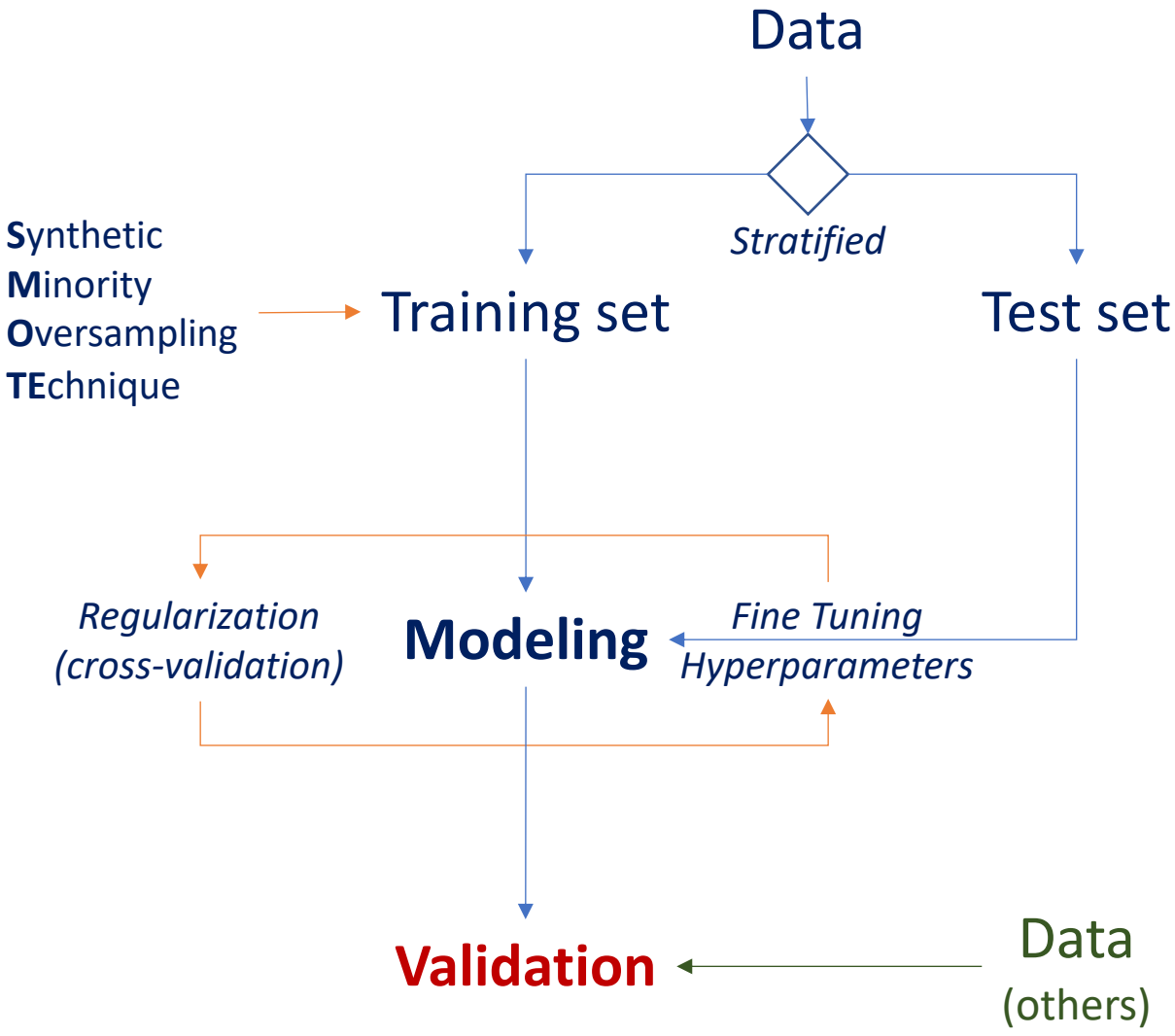
3-methyl-1-butanol_Monomer



Ethyl acetate_Monomer



Machine learning approach



Classification

Technique	Precision	recall	f1-score	Support
E	0.84	0.84	0.84	25
V	0.64	0.76	0.70	21
L	0.90	0.64	0.75	14
Accuracy			0.77	60
Weighted avg	0.78	0.77	0.77	60

TP: true positive
FP: false positive
TN: true negative
FN: flase negative

Precision: accuracy of positive predictions [TP/(TP+FP)]
Recall: the ability of a classifier to find all positive instances [TP/(TP+FN)]
f1 score: percent of positive predictions considered correct [2(Recall*Precision)/(Recall+Precision)]

Classification

Quadratic Discriminant Analysis	Precision	recall	f1-score	Support
E	0.84	0.84	0.84	25
V	0.64	0.76	0.70	21
L	0.90	0.64	0.75	14
Accuracy			0.77	60
Weighted avg	0.78	0.77	0.77	60

SVM Classifier	Precision	recall	f1-score	Support
E	0.85	0.68	0.76	25
V	0.64	0.76	0.70	21
L	0.87	0.93	0.90	14
Accuracy			0.77	60
Weighted avg	0.78	0.77	0.77	60

Tree Net Classifier	Precision	recall	f1-score	Support
E	0.71	0.68	0.69	25
V	0.50	0.52	0.51	21
L	0.71	0.71	0.71	14
Accuracy			0.63	60
Weighted avg	0.64	0.63	0.63	60

kernel-SVM Classifier (polynomial)	Precision	recall	f1-score	Support
E	0.87	0.80	0.83	25
V	0.70	0.67	0.68	21
L	0.71	0.86	0.77	14
Accuracy			0.77	60
Weighted avg	0.77	0.77	0.77	60

Classification

Naive Bayes Classifier (gaussian)	Precision	recall	f1-score	Support
E	0.85	0.92	0.88	25
V	0.72	0.86	0.78	21
L	1.00	0.57	0.73	14
Accuracy	0.86			60
Weighted avg	0.84	0.82	0.81	60

Adaptive Boosting Classifier	Precision	recall	f1-score	Support
E	0.92	0.48	0.63	25
V	0.54	0.95	0.69	21
L	1.00	0.71	0.83	14
Accuracy			0.70	60
Weighted avg	0.81	0.70	0.70	60

XG-Boost	Precision	recall	f1-score	Support
E	0.61	0.56	0.58	25
V	0.53	0.48	0.50	21
L	0.67	0.86	0.75	14
Accuracy			0.60	60
Weighted avg	0.59	0.60	0.59	60

Gaussian Process Classifier	Precision	recall	f1-score	Support
E	0.76	0.52	0.62	25
V	0.52	0.71	0.60	21
L	0.71	0.71	0.71	14
Accuracy			0.63	60
Weighted avg	0.67	0.63	0.63	60

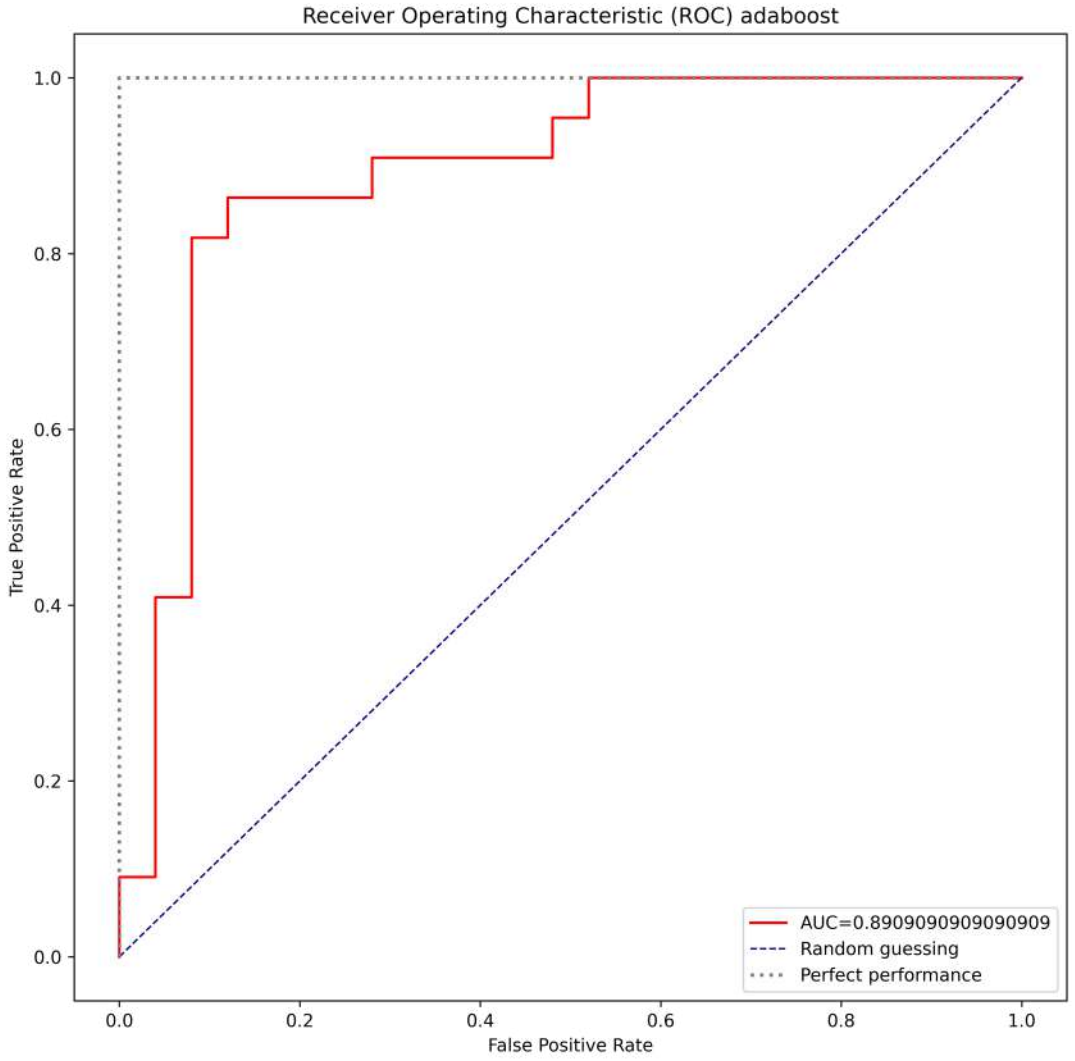
Classification

k-Nearest Neighbor (KNN)	Precision	recall	f1-score	Support
E	0.67	0.64	0.65	25
V	0.52	0.62	0.57	21
L	0.82	0.64	0.72	14
Accuracy			0.63	60
Weighted avg	0.65	0.63	0.64	60

Random forest	Precision	recall	f1-score	Support
E	0.81	0.84	0.82	25
V	0.73	0.76	0.74	21
L	0.92	0.79	0.85	14
Accuracy			0.80	60
Weighted avg	0.80	0.80	0.80	60

Logistic regression	Precision	recall	f1-score	Support
E	0.79	0.60	0.68	25
V	0.45	0.48	0.47	21
L	0.63	0.71	0.71	14
Accuracy			0.58	60
Weighted avg	0.61	0.58	0.59	60

Classification



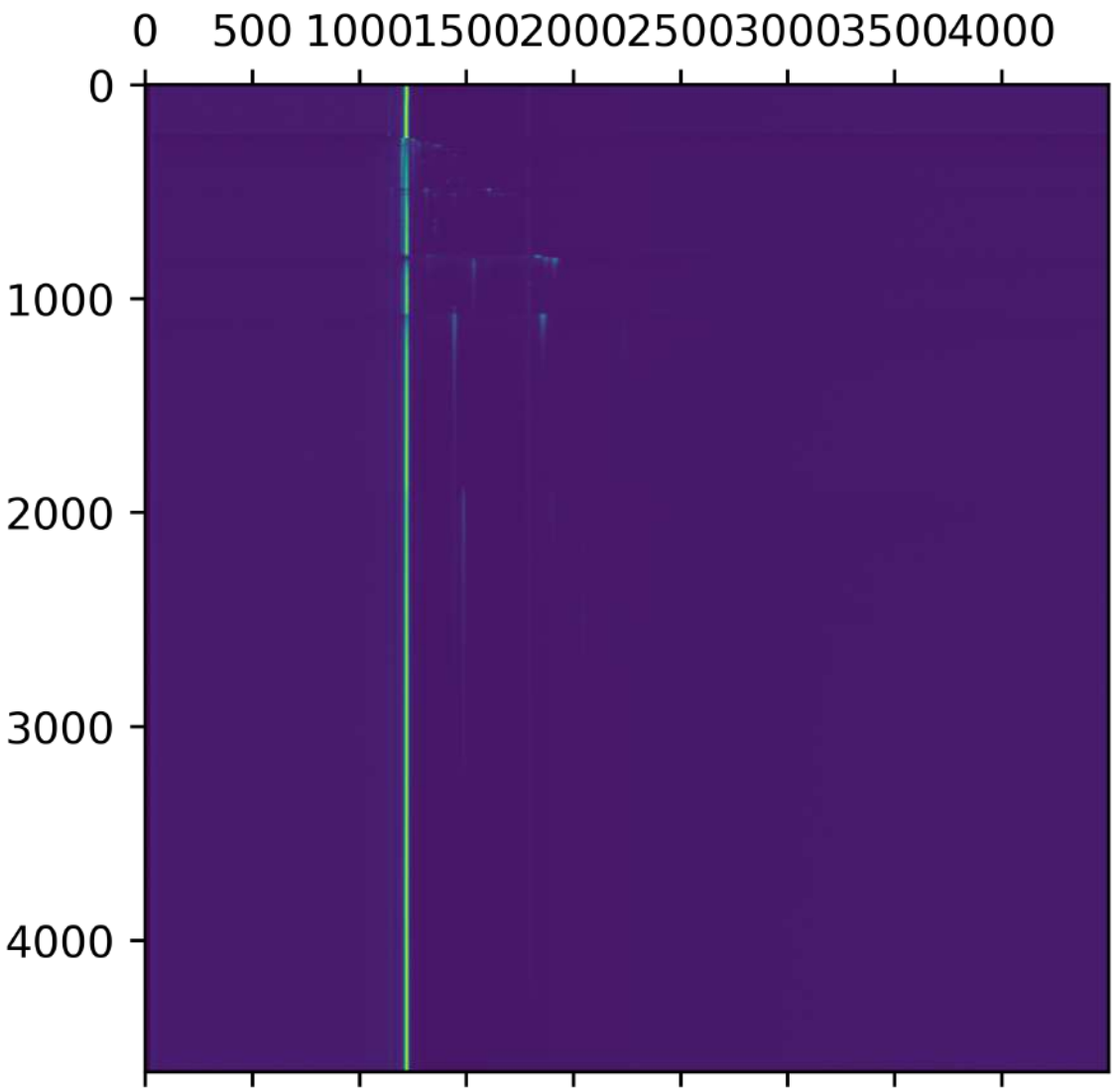
Random forest (binary: evoo/voo)	Precision	recall	f1-score	Support
E	0.95	0.76	0.84	25
V	0.78	0.95	0.84	27
Accuracy			0.85	47
Weighted avg	0.87	0.85	0.8/5	47
Adaptive Boosting Classifier (binary: evoo/voo)	**Precision**	**recall**	**f1-score**	**Support**
E	0.82	0.92	0.87	25
V	0.89	0.77	0.83	22
Accuracy			0.85	47
Weighted avg	0.86	0.85	0.85	47

UNTARGETED

Only for Industry Digital Transformation Applications

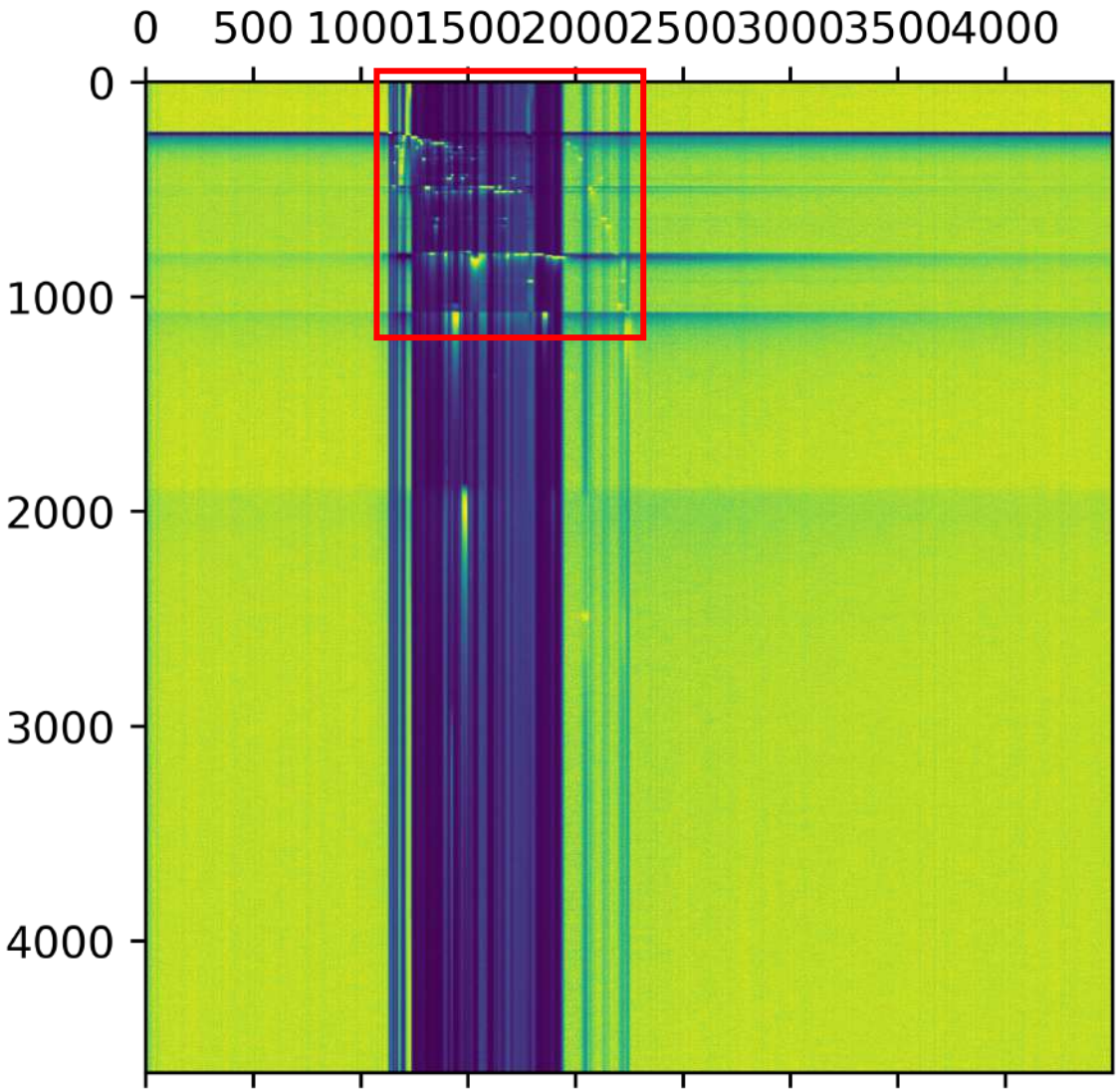
analytica

Data handling: original CSV (comma-separated values)



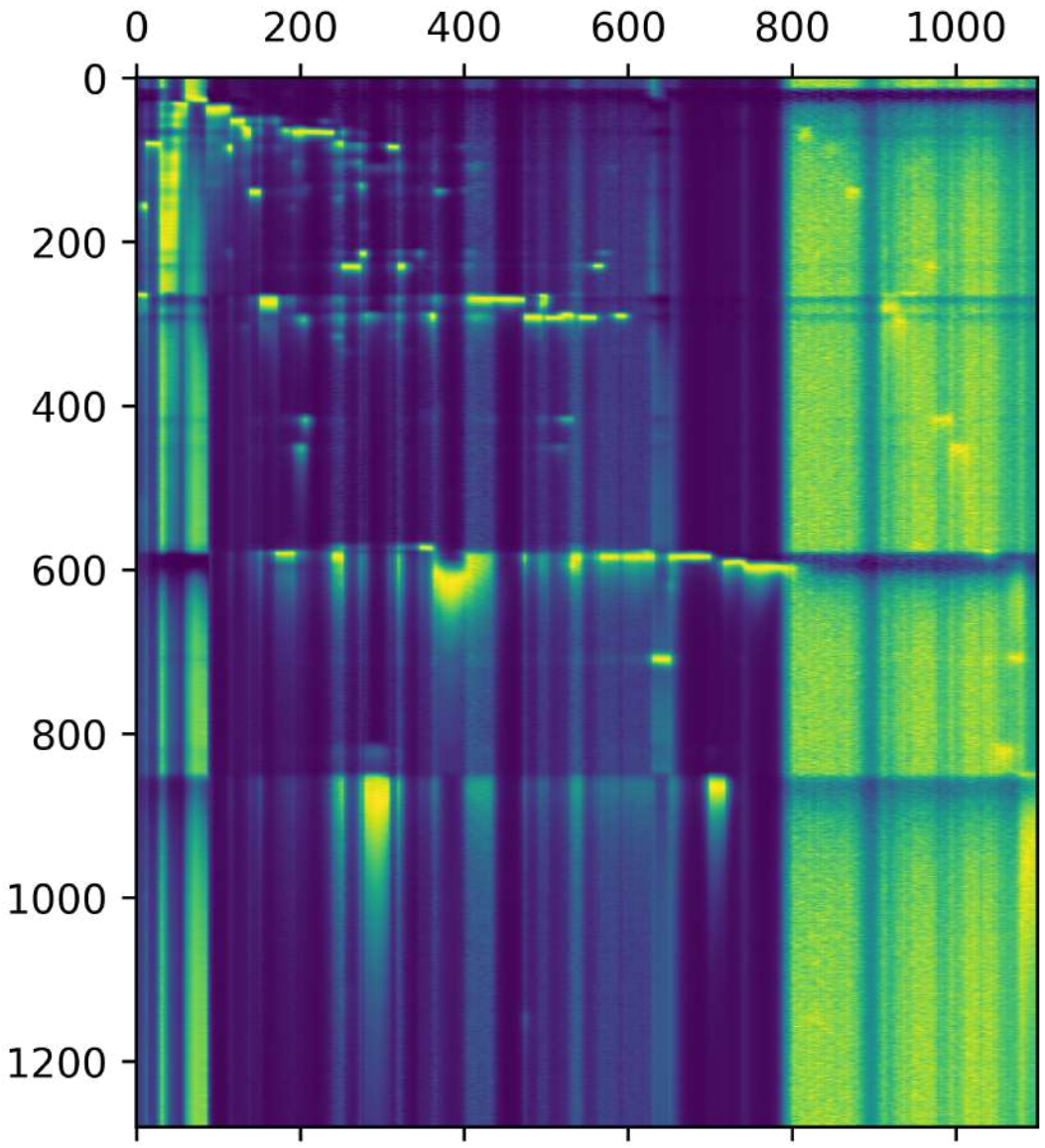
Matrice 4616, 4500 (CSV)

Data handling: Standard Scaler (0,1)



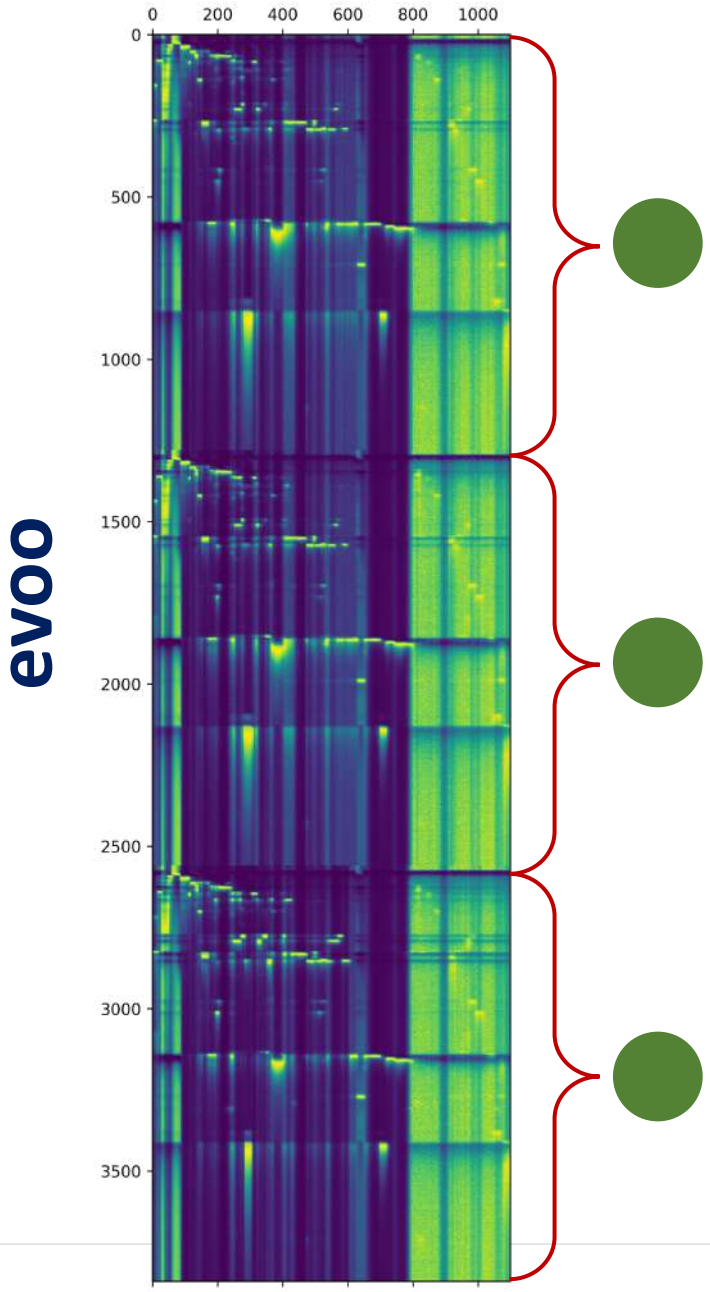
Matrice 4616, 4500 (0,1)

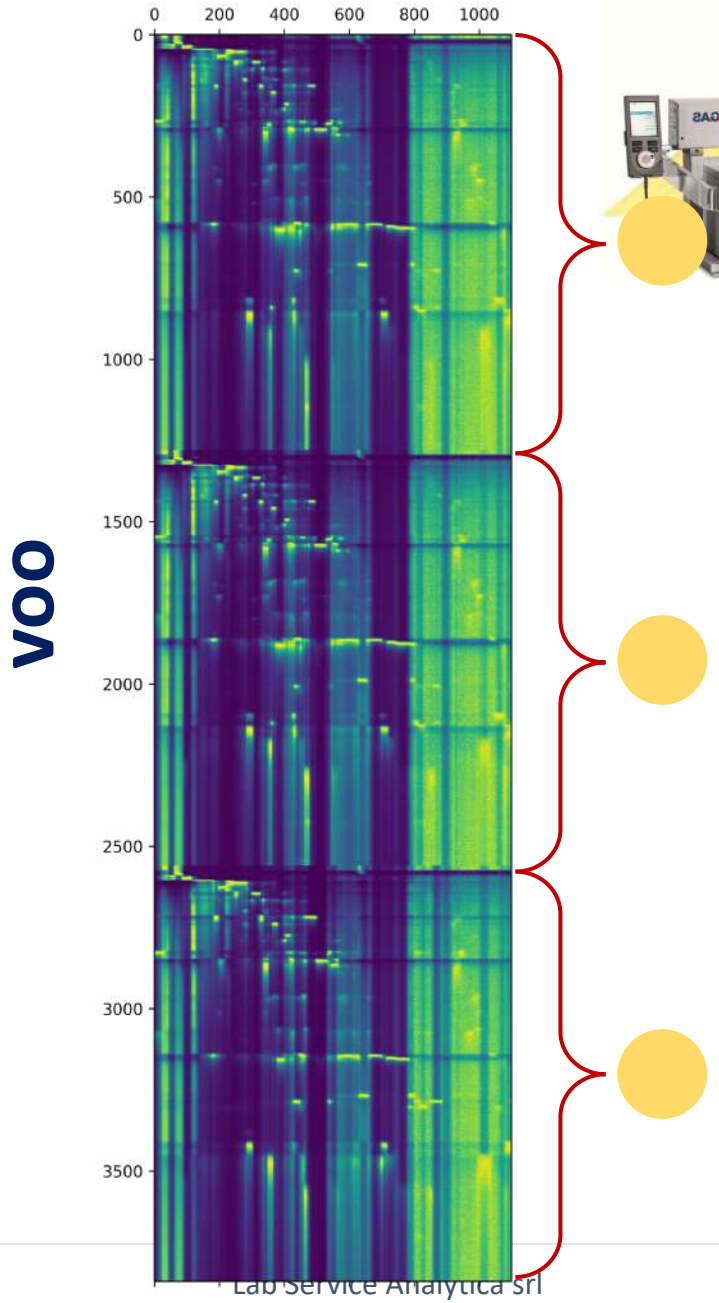
Data handling: pruning



Matrice 1280 x 1100 (0,1)

Data set: images





Untargeted Classification

Tree Net Classifier	Precision	recall	f1-score	Support
E	0.86	0.75	0.80	8
V	0.82	0.90	0.86	10
Accuracy			0.83	18
Weighted avg	0.84	0.83	0.83	18

XG-Boost	Precision	recall	f1-score	Support
E	0.71	0.62	0.67	8
V	0.73	0.80	0.76	10
Accuracy			0.72	18
Weighted avg	0.72	0.72	0.72	18

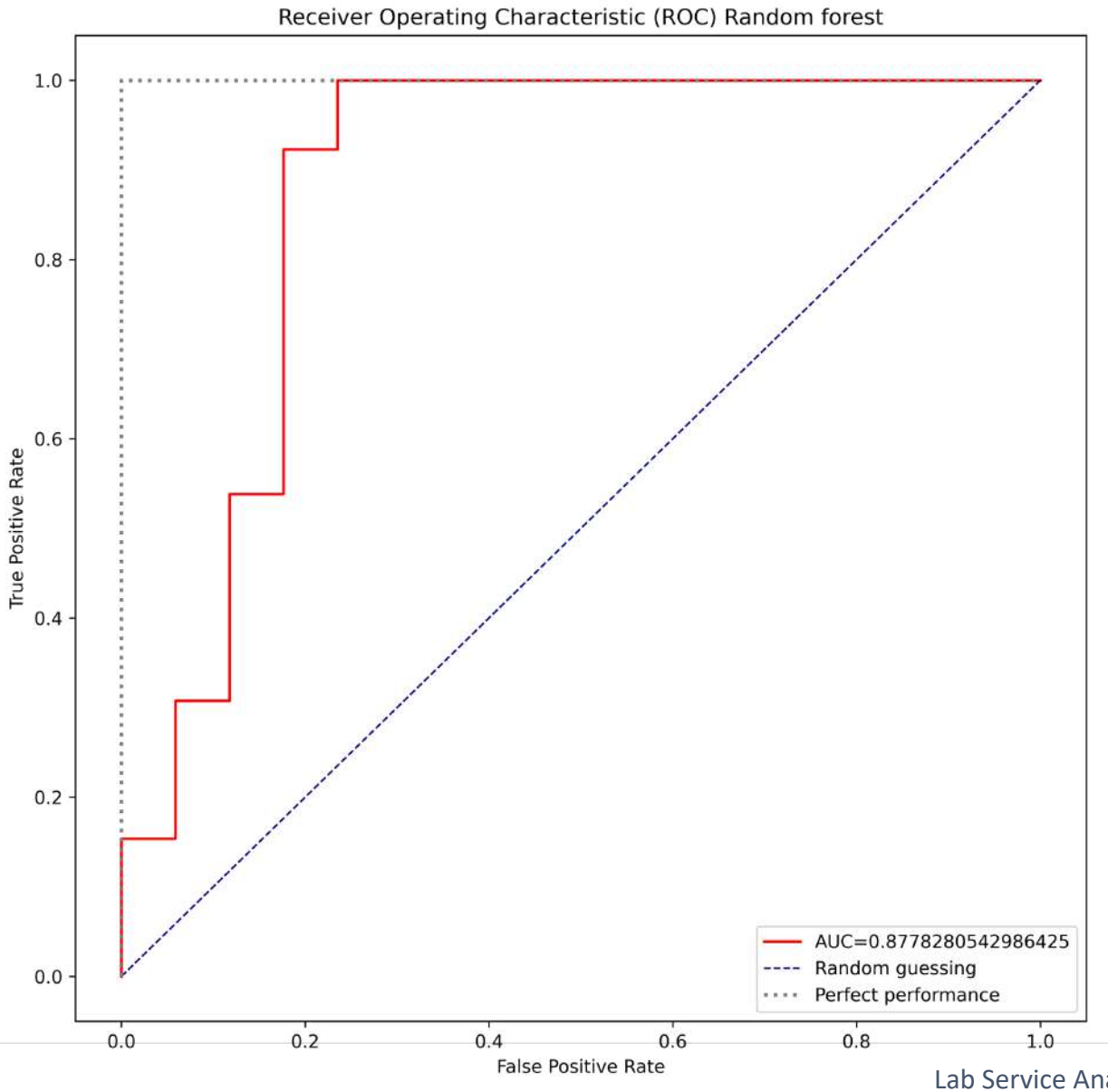
SVM Classifier	Precision	recall	f1-score	Support
E	0.78	0.88	0.82	8
V	0.89	0.80	0.84	10
Accuracy			0.83	18
Weighted avg	0.84	0.83	0.83	18

Gaussian Process Classifier	Precision	recall	f1-score	Support
E	0.44	1.00	0.62	8
V	0.00	0.00	0.00	10
Accuracy			0.44	18
Weighted avg	0.20	0.44	0.27	18

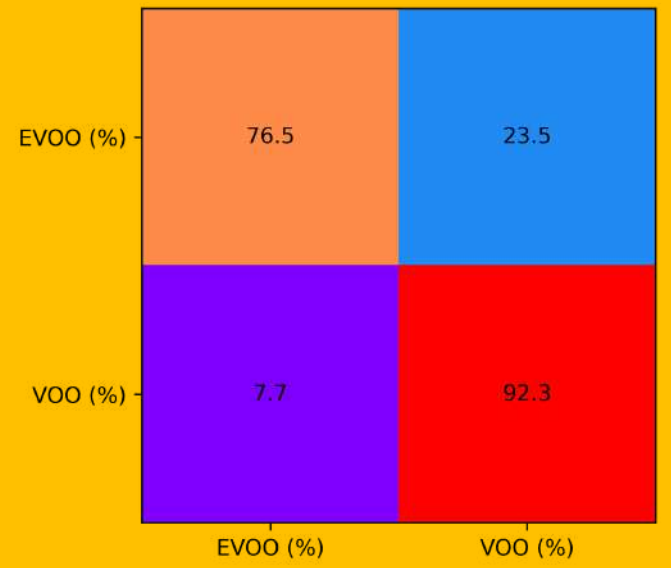
Untargeted Classification

k-Nearest Neighbor (KNN)	Precision	recall	f1-score	Support
E	0.83	0.62	0.71	8
V	0.75	0.90	0.82	10
Accuracy			0.78	18
Weighted avg	0.79	0.78	0.77	18
Logistic regression	**Precision**	**recall**	**f1-score**	**Support**
E	0.71	0.62	0.67	8
V	0.73	0.80	0.76	10
Accuracy			0.72	18
Weighted avg	0.72	0.72	0.72	18

Untargeted Classification: Random forest



Actual vs Predicted labels



Random forest (criterion: entropy)	Precision	recall	f1-score	Support
E	1.00	0.75	0.86	8
V	0.83	1.00	0.91	10
Accuracy			0.89	18
Weighted avg	0.91	0.89	0.89	18

Artificial Neural Network (ANN)

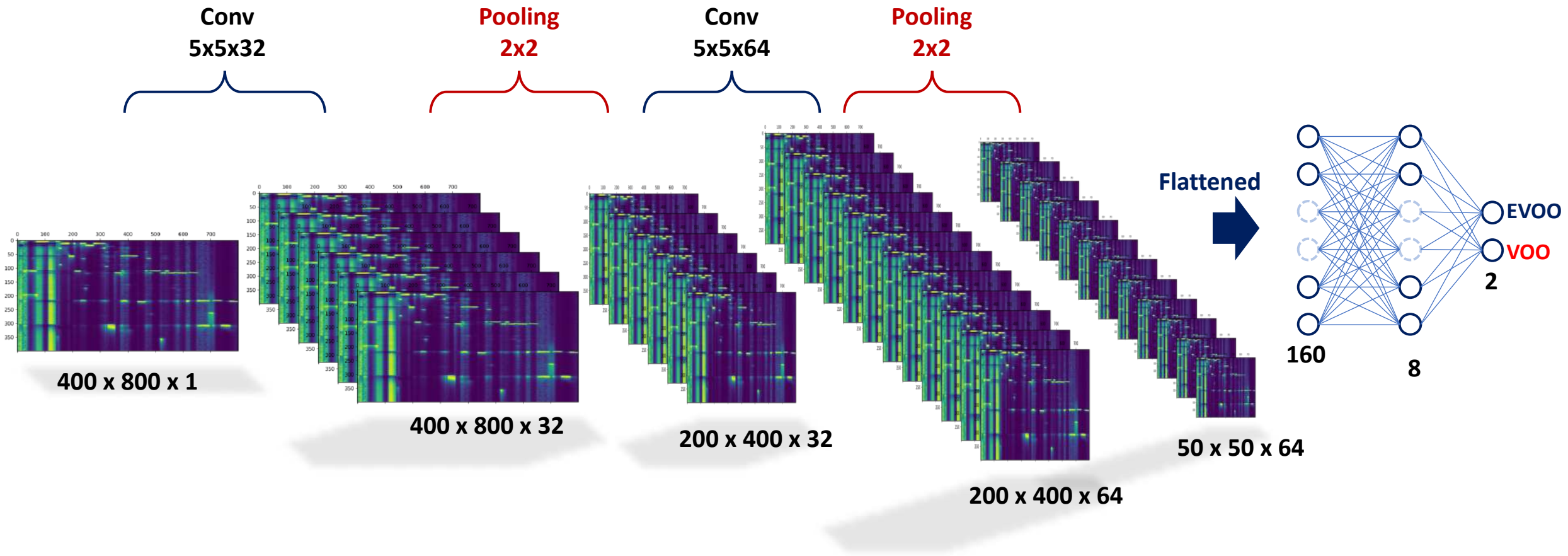
Geoffrey Hinton
(RBM's)

Yann LeCun
(CNN)

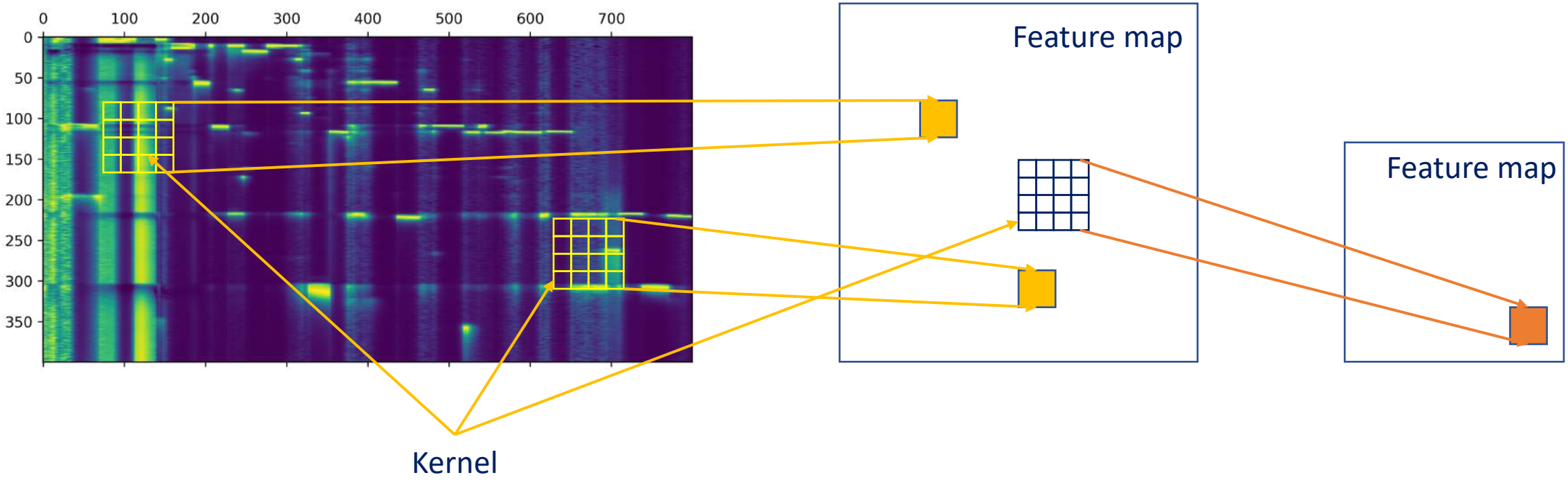
Yoshua Bengio
(DNN's)

Ian Goodfellow
(GAN's)

Classification: Convolutional Neural Networks



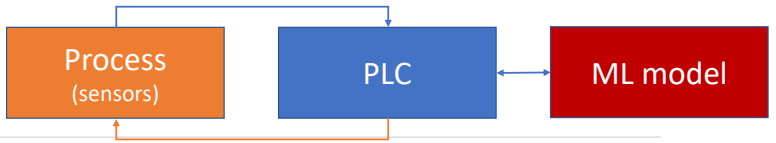
Classification: Convolutional Neural Networks



Conclusions

more whole informations
is better than single ones..

1. Good performances
2. Less time consumption
4. Less efforts
5. More precision
6. Modeling automatic process control



Thanks for Your Attention!

Scansiona il codice per visitare il nostro sito web

Strumenti

Standard

Applicazioni

Lab Service Analytica srl
Via Emilia 51/c 40011
Anzola dell'Emilia (BO)
www.labservice.it
info@labservice.it