TraceRice

Machine Learning-Driven Classification and Quality Assessment of Rice Varieties Using NIR Spectroscopy

TRACING RICE AND VALORIZING SIDE STREAMS ALONG MEDITERRANEAN BLOCKCHAIN

28.10.2024

TRACE-RICE with Grant nº 1934, (call 2019, section 1 Agrofood) is part of the PRIMA Programme supported under Horizon 2020, the European Union's Framework Programme for Research and Innovation

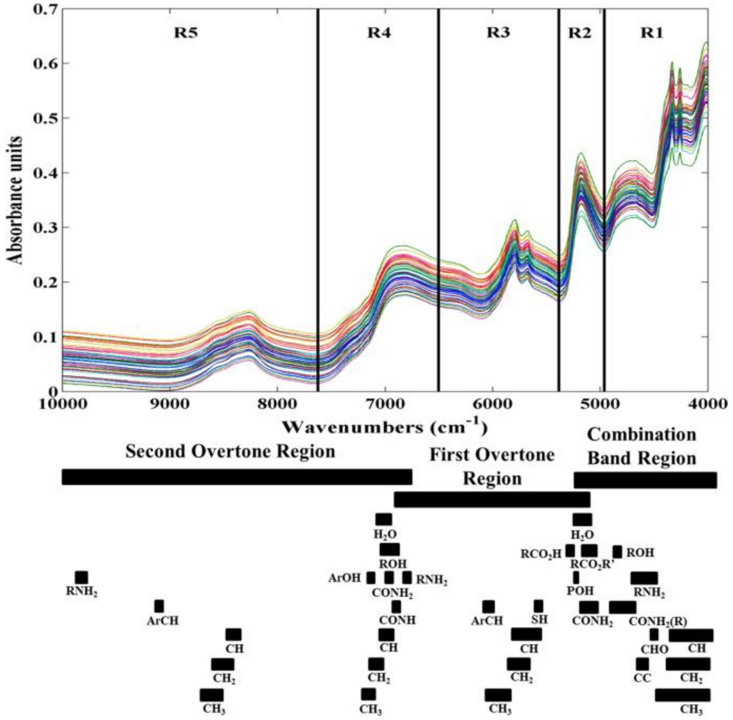
PRIMA PARTNERSHIP FOR RESEARCH AND INNOVATION IN THE MEDITERRANEAN AREA

iniav Instituto Nacional de Investigação Agrária e Veterinária, I.P.

Introduction

Near Infrared Spectroscopy

- NIR covers the spectral region defined by wavenumber: **14,000–4000 cm^{-1}.**
- The absorption bands come from **overtones**, **combinations of overtones** and/or combinations of fundamental **vibrational motions**.
- NIR can be used for **polymorphism characterization**.



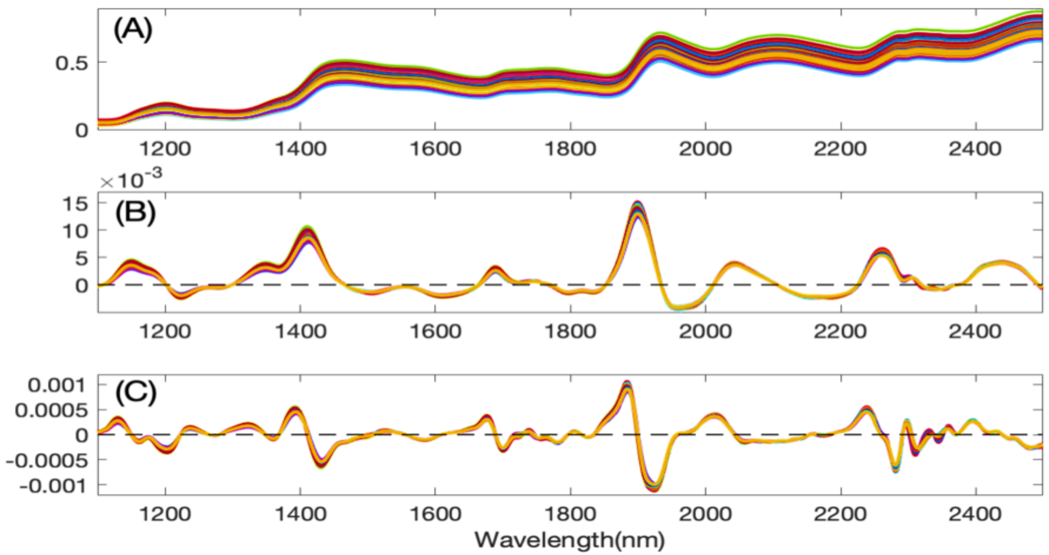
Introduction

Near Infrared Spectroscopy

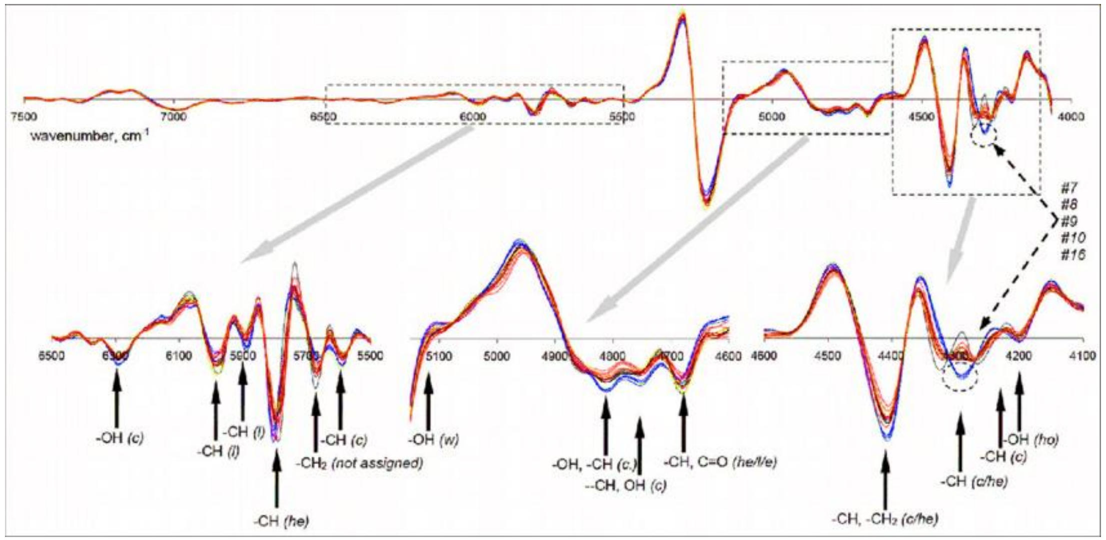
NIR spectra are pre-processed using **specific algorithms** such as the multiplicative scatter correction, derivatives (1st and 2nd).

The derivatives method allows to **remove of unimportant signals** from samples.

The smaller variations remaining are due to the **chemical differences** between samples.



NIR spectra (A); First derivative of spectra (B); Second derivative (C)



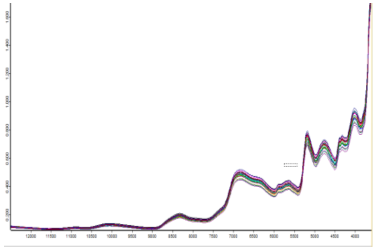
Introduction

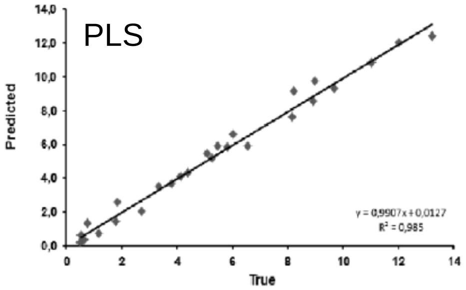
Analytical methods in food industry are **highly expensive** and **time-consuming**.

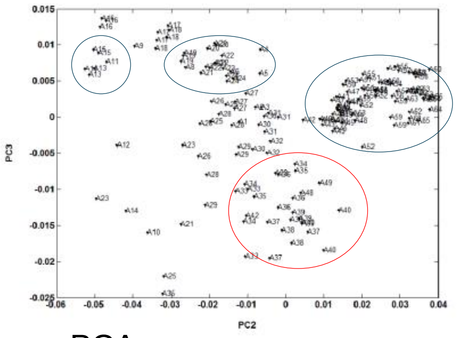
Near-Infrared Spectroscopy (NIRS)

- **Non-destructive** method - the samples can be directly analyzed, replacing the wet chemistry analysis.
- **Predict accurately the biochemical** parameters: fat, protein, moisture, etc.
- **The classification** procedure is based on the biochemical, pasting parameters or NIR spectra.

Objectives







PCA

CHARACTERIZATION AND QUALITY EVALUATION

BIOCHEMICAL PARAMETERS

- Amylose
- Ash
- Fat
- Protein

PASTING PROFILES

- Breakdown
- Final viscosity
- Peak viscosity
- Trough
- Setback

Near-infrared spectroscopy (NIRS) and **machine learning** methodologies were used to develop models to predict some biochemical/pasting/ related to rice quality parameters.

Trace-rice Collection

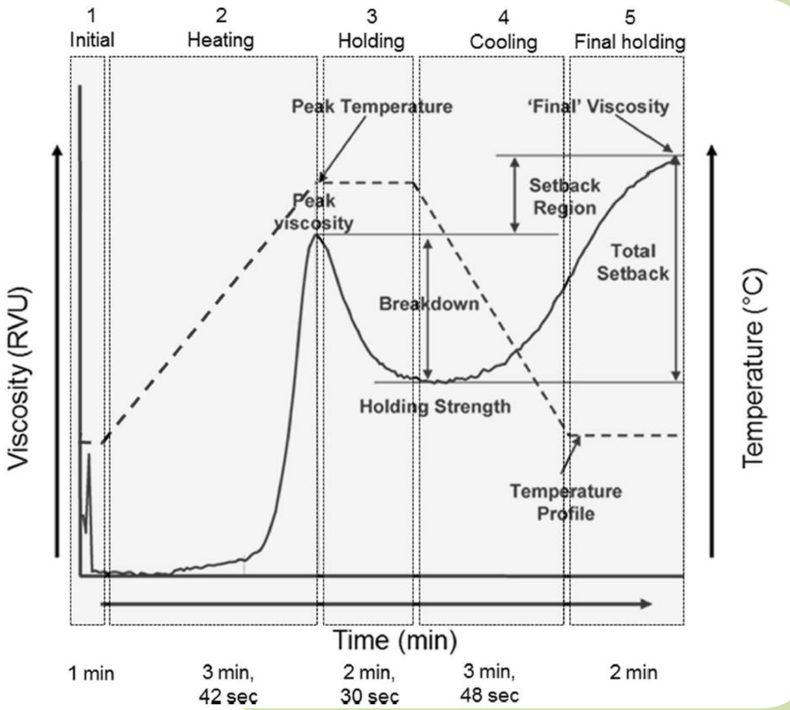
- **Twenty varieties** were cultivated in the Mediterranean Region from Europe and Egypt.
- **Two varieties**, including Basmati, were imported from outside Europe.

Samples are categorized into the following commercial types:

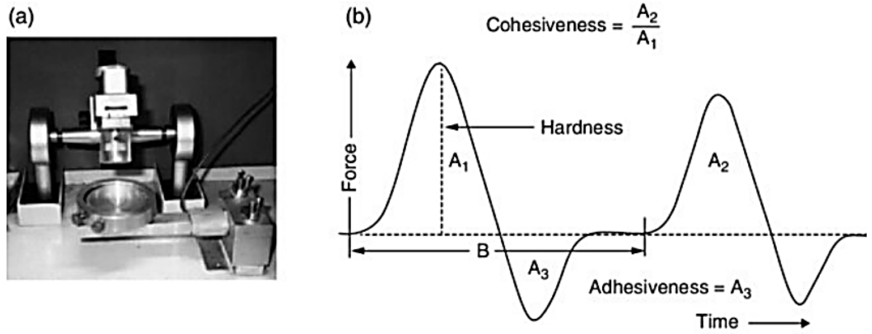
- Long A
- Long B
- Medium grain
- Short grain
- European aromatic
- Basmati

Physicochemical Characterization

- **Biometric Parameters & Grain Appearance**: Length, Width, Chalkiness, Total whiteness, Vitreous whiteness.
- **Biochemical parameters:** starch, protein, fat, fibre, ash, etc.
- **Viscosity Profiles:** Final viscosity; Setback; Breakdown; Trough viscosity; Peak viscosity; Gelatinization Temperature.
- **Cooking Parameters:** Water uptake; Volumetric expansion ratio; Solids leached;
- **Texture Profile Analysis:** Cohesiveness; Chewiness; Adhesiveness; Hardness; Gumminess; Springiness.



Rapid Visco Analyser (RVA) pasting profile (Balet et al. 2018)



Texture profile analysis (TPA) using a Texturometer.

Prediction and classification models

PARTIAL LEAST SQUARES (PLS)

- Algorithm that **estimates** and **quantifies** the components in a particular sample based on regression.

PRINCIPAL COMPONENT ANALYSIS (PCA)

- Machine learning method used to **reduce the dimensionality** and simplify a large data set into a smaller set while maintaining significant patterns and trends.

PARTIAL LEAST SQUARES-DISCRIMINANT ANALYSIS (PLS-DA)

- Classification tool that allows calculating **predictive models** based on a partial least squares regression algorithm that searches for latent variables with maximum covariance.

K-NEAREST NEIGHBOURS (K-NN)

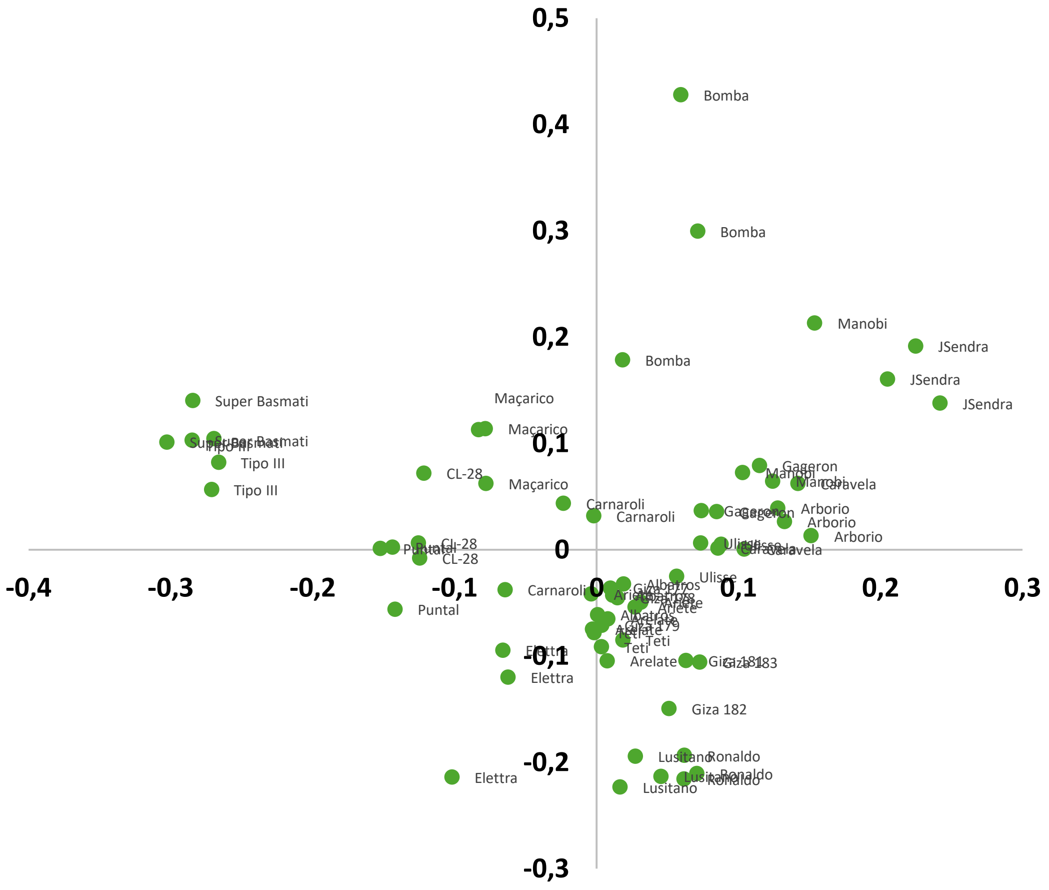
- Chemometric technique where **classification rules are based on the neighborhood** of training set objects based on Euclidian distances or correlation coefficients between the unknown object and the training set objects.

RESULTS

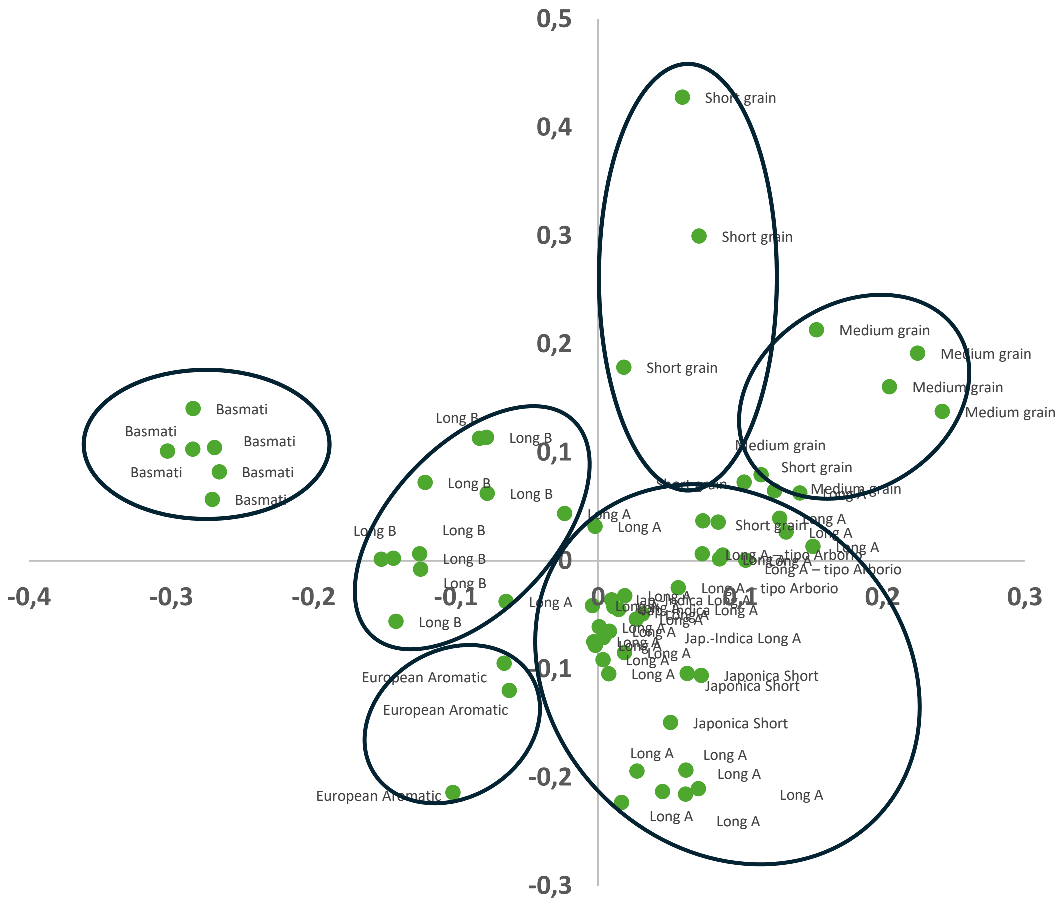
PLS models for different physicochemical parameters

Parameter	R^2	RMSECV	Spectral region (cm^{-1})	Spectral Pre-processing
Whiteness level	0.94	1.730	9827 - 8038 6263 – 4481	Straight Line Subtraction
Width average	0.94	0.111	10715 - 9820 8933 - 7151 6263 – 4481	Multiplicative Scattering Correction
Total whiteness	0.95	2.490	10715 - 8925 6263 – 3594	Multiplicative Scattering Correction + 1 st Derivative
Vitrea whiteness	0.91	1.18	9403 – 4597	Vector Normalization
Setback	0.77	365	9403 - 8447 4427 – 4242	Vector Normalization + 1st Derivative
Gelatinization temperature	0.99	0.52	9827 - 6256 5376 – 4481	No spectral data processing
Peak viscosity	0.94	183	9403 – 4242	Vector Normalization + 1st Derivative
Hardness	0.89	223	8933 – 4481	Multiplicative Scattering Correction
Adhseviness	0.81	33.6	10,715 - 9820 7158 - 6256 5376 – 4481	Constant offset elimination
Resilience	0.96	0.545	9403 - 7498 5777 - 5446 4605-4420	2nd Derivative
Springiness	0.98	1.65	9827-8038	Straight line subtraction
Gumminess	0.91	76.5	9403-4242	Mulpitplicative scattering correction
Chewinness	0.81	92.2	10715-9820 7158-4481	Constant offset elimination
Cohesion	0.89	0.013	9404-7498; 6102-5446; 4605-4420	2nd Derivative
Solid leach	0.99	0.091	12498-11594; 8933-8038	Min-Max normalization
Water up ratio	0.87	0.099	8933-7151;6264-4482	Straight line subtraction

PCA-Physicochemical Characterization

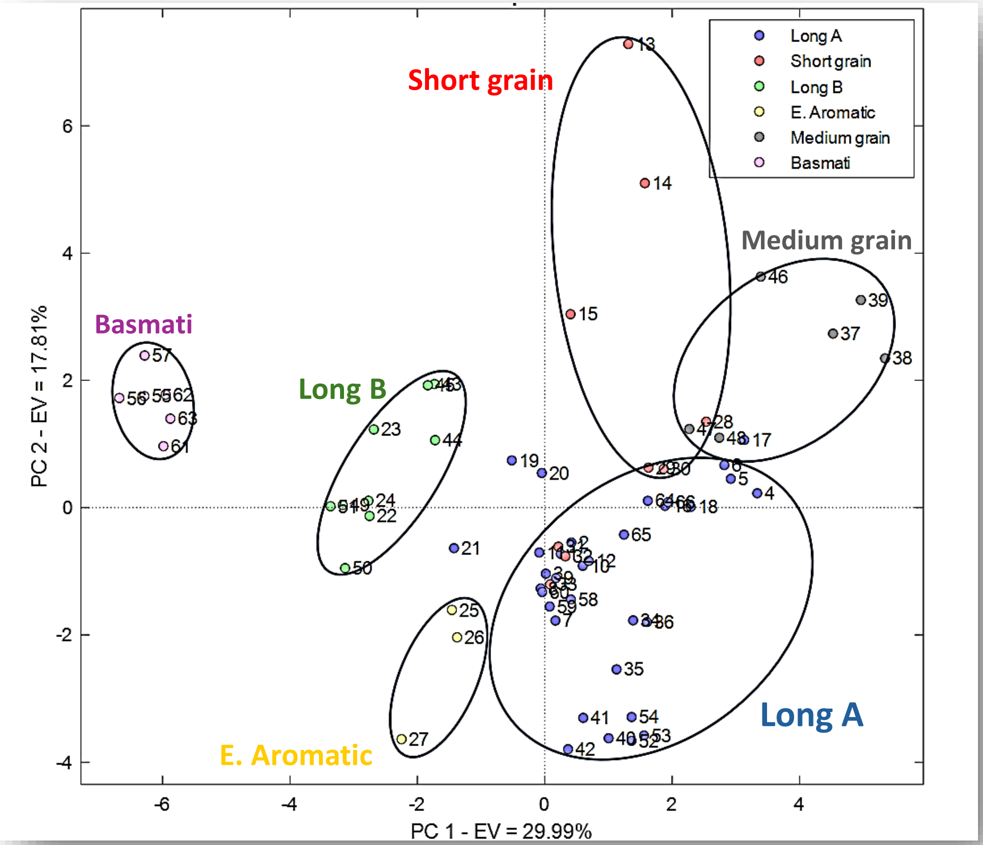


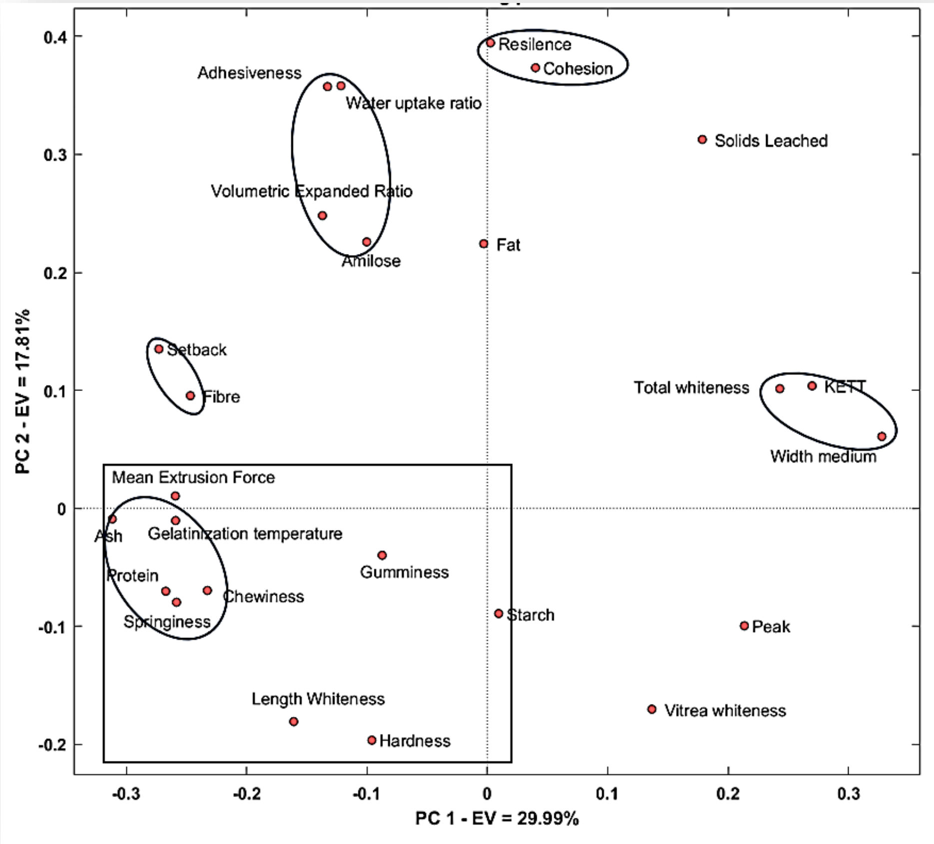
PCA – Analysis and representation of rice in terms of code name.



PCA – Analysis and representation of rice in terms of commercial type.

PCA-Physicochemical Characterization





- PCA analysis:
- **Six different clusters** were developed based on the biochemical, pasting, biometrics, and cooking parameters.
- **Loadings** indicate the contribution of each parameter to the principal components.

Physicochemical Characterization of Commercial Rice Types

- The cluster defined by **Long A** variety is characterized by starch content, vitreous whiteness, peak viscosity, and gumminess.

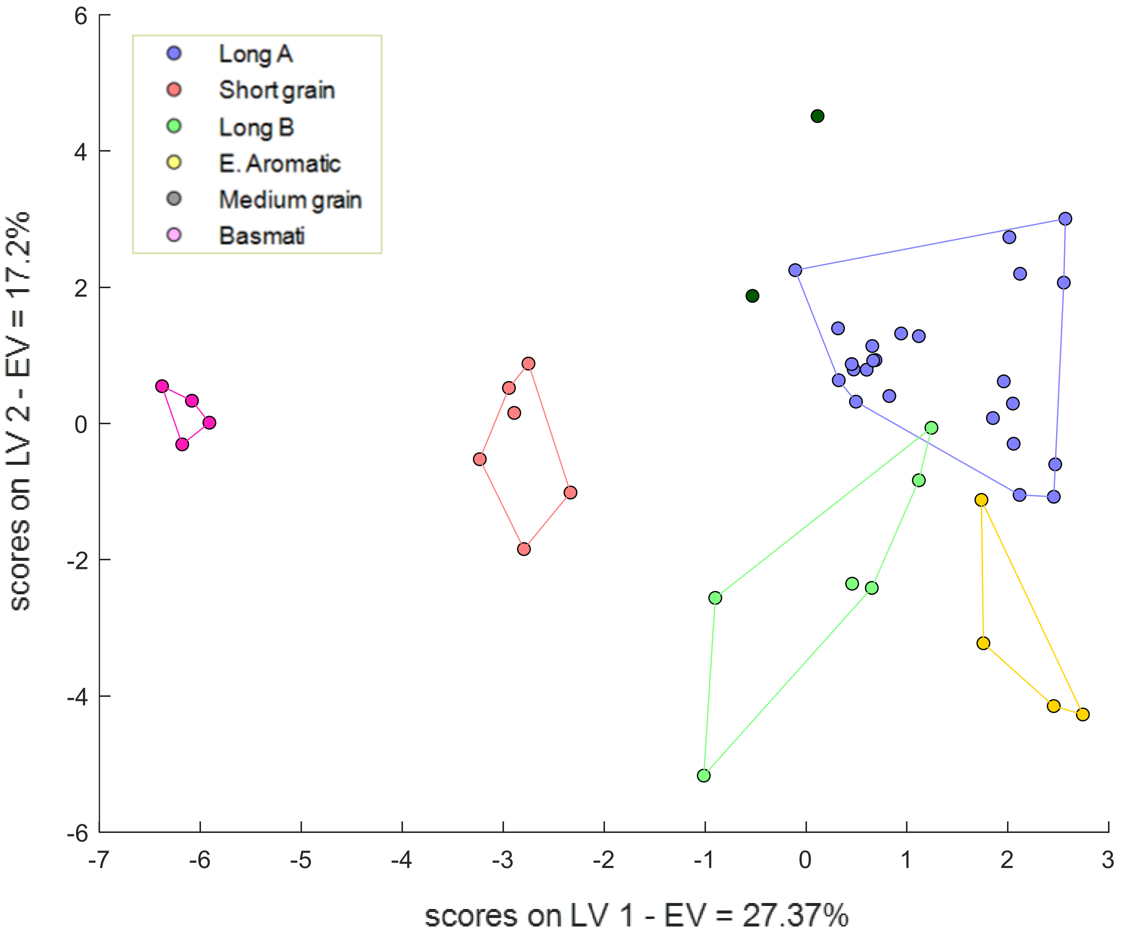
- The cluster constituted by **Short** variety is characterized by **resilience, cohesion, solids leach and fibers.**

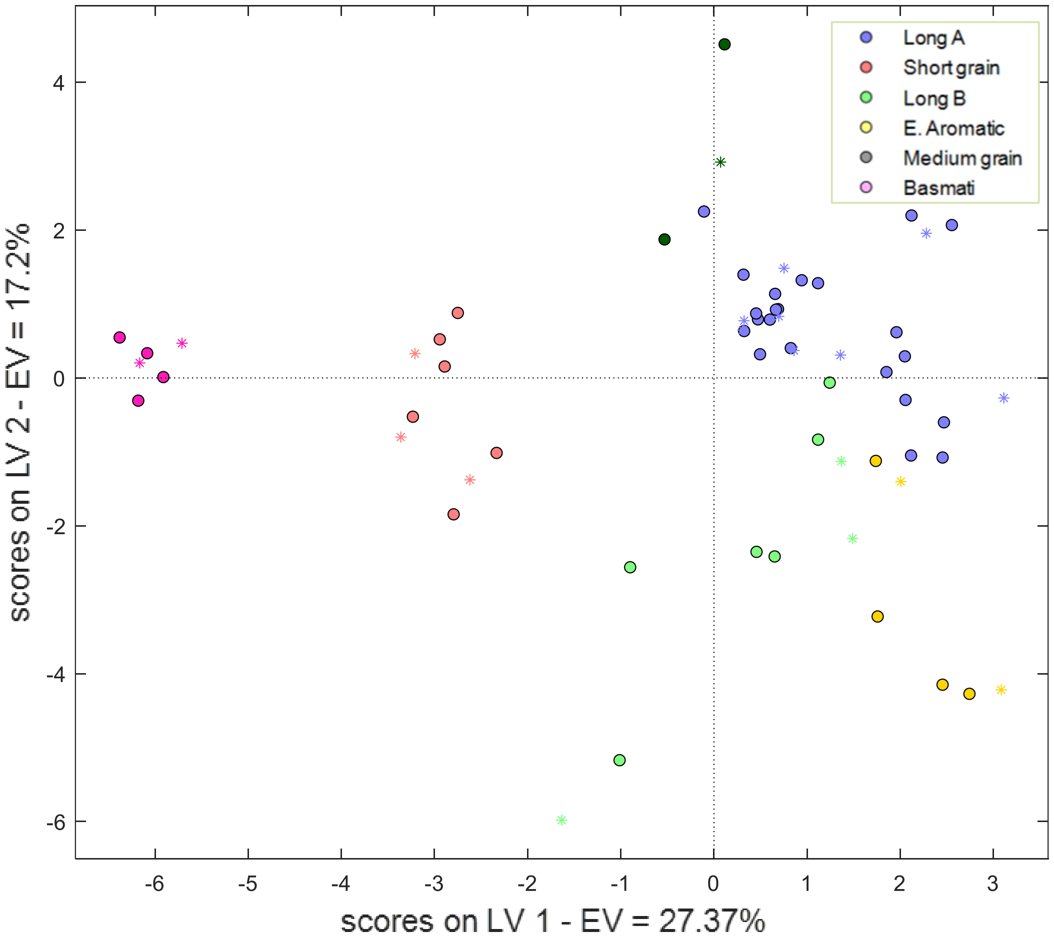
- The grain length, hardness, and gumminess properties characterize the **Long B variety**.

- Aromatic rice varieties (**Basmati and Super Basmati**) are defined by ash, fiber, protein, gelatinization temperature, setback, springiness, gumminess, and mean extrusion force.

PLS-DA

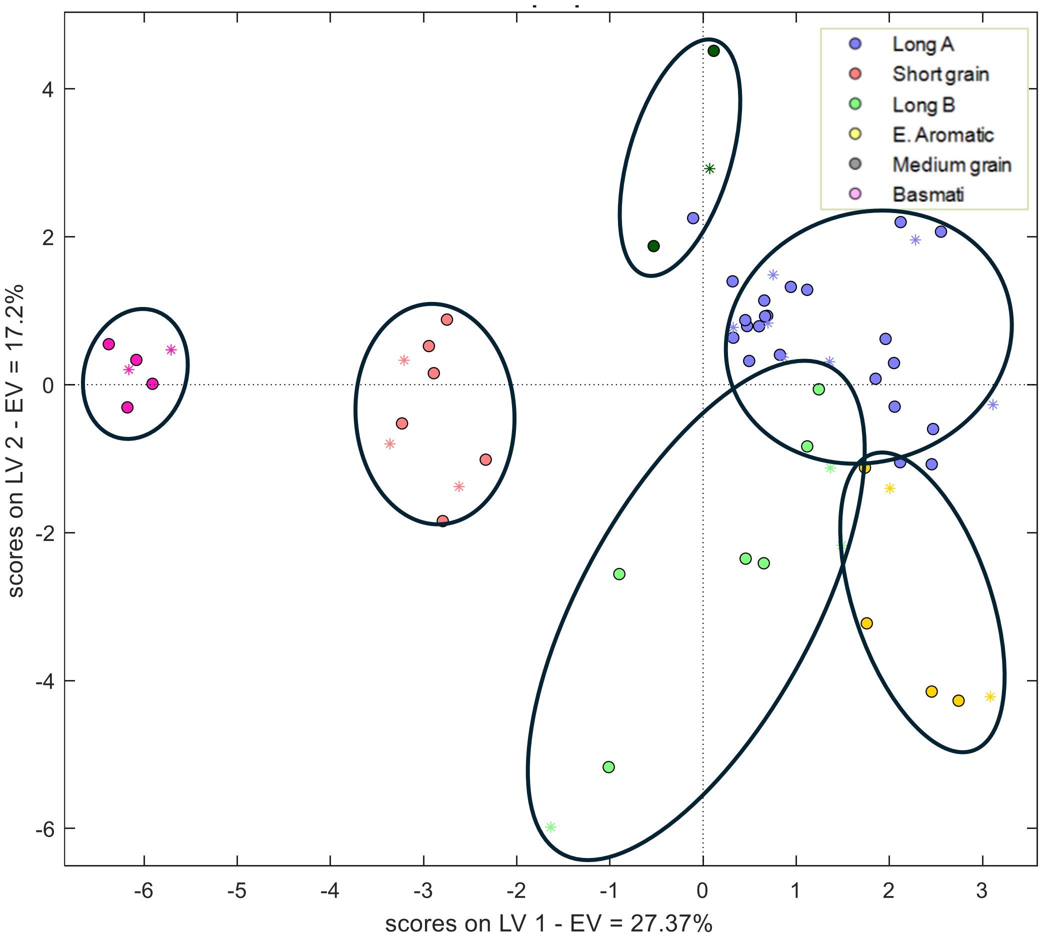
Classification of rice types based on the 25 parameters.





PLS-DA

Classification of rice types based on the 25 parameters.



Model type: PLS-DA
Explained variance: 93%
Error rate: 17%
Error rate CV: 13%

Training
No-error rate: 93%
Accuracy: 100%
Not assigned: 4%

Cross-validation
No-error rate: 87%
Accuracy: 93%
Not assigned: 13%

Prediction on external samples
Accuracy: 100%
Not-assigned: 20%

KNN

Classification of rice types based on the 25 parameters:

Model type: KNN

Error rate: 13%
Error rate CV: 13%

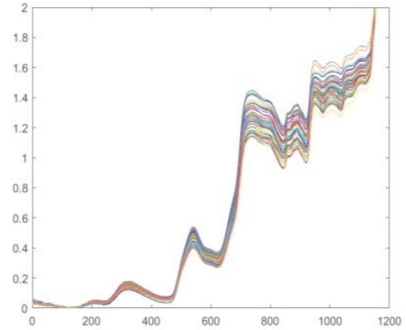
Training
No-error rate: 88%
Accuracy. 96%

Cross-validation
No-error rate: 88%
Accuracy: 96%

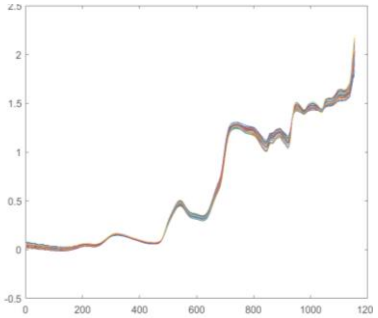
Prediction on external samples
No-error rate: 100%
Accuracy: 100%

NIR spectra of rice grain & rice flour

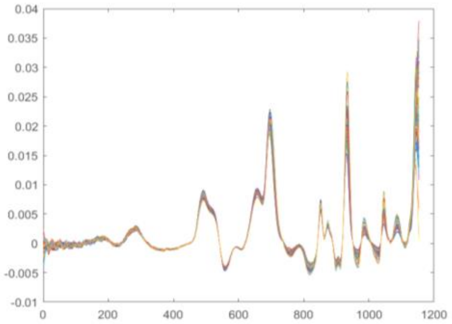
- Spectra were registered using a NIR spectrometer (MPA, Bruker Optics, Germany).
- NIR Spectral were preprocessed by Multiplicative Scatter Correction; 1st Derivative; 2nd Derivative (Matlab *software*).



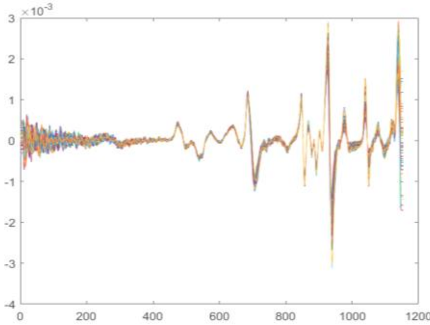
NIR spectra rice grain



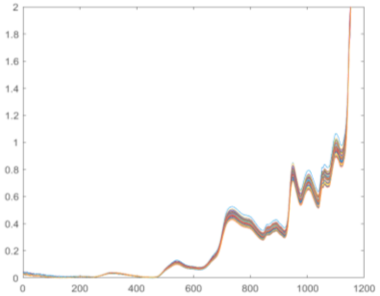
NIR spectra rice grain (MSC)



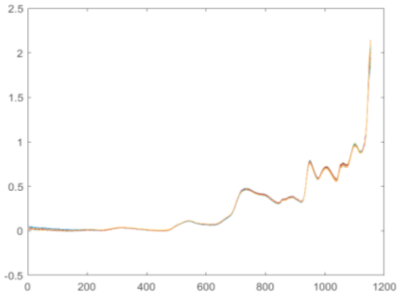
NIR spectra rice grain (1st derivative)



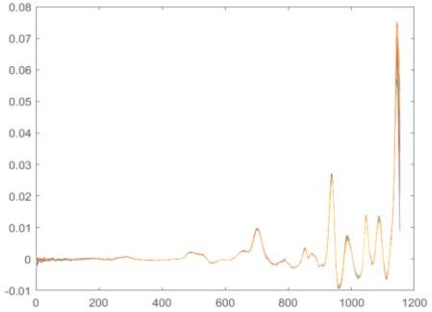
NIR spectra rice grain (2nd derivative)



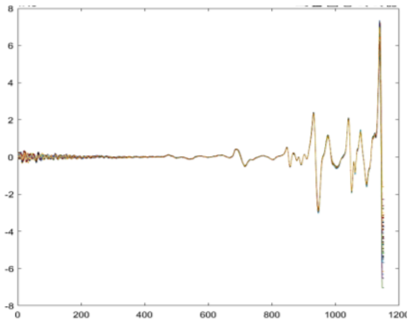
NIR spectra rice flour



NIR spectra rice flour (MSC)

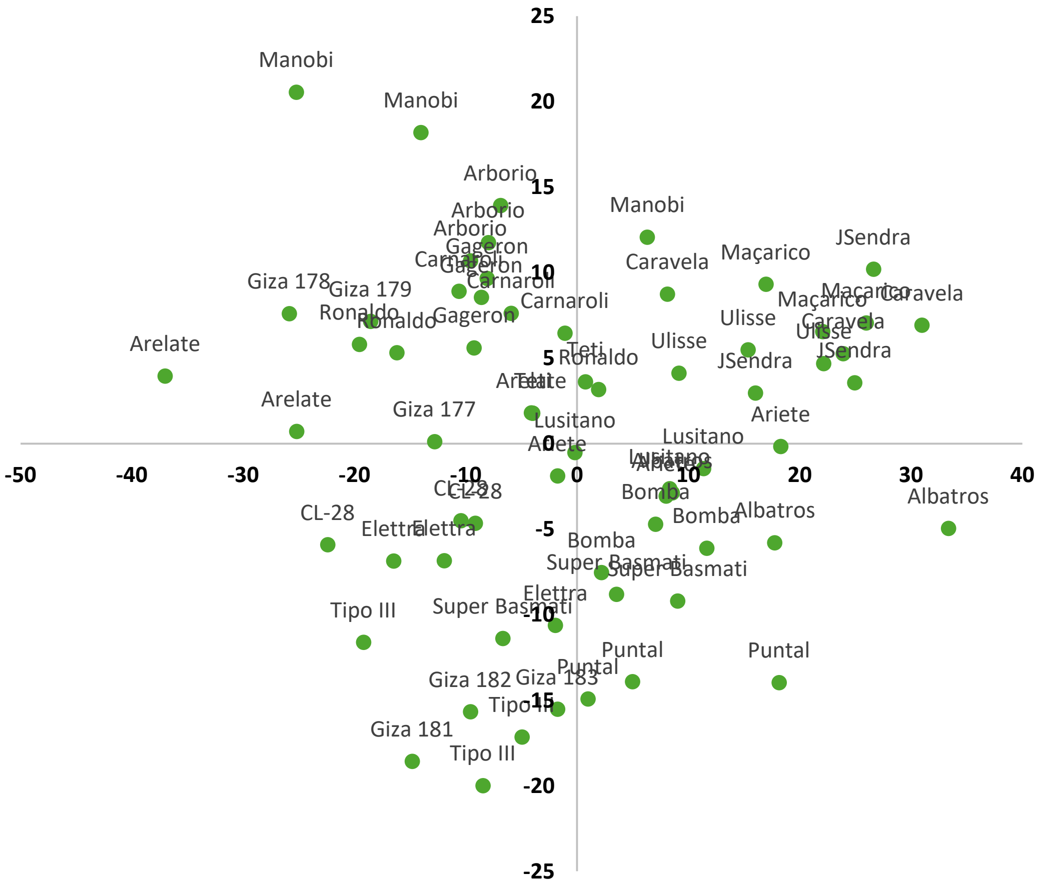


NIR spectra rice flour (1st derivative)

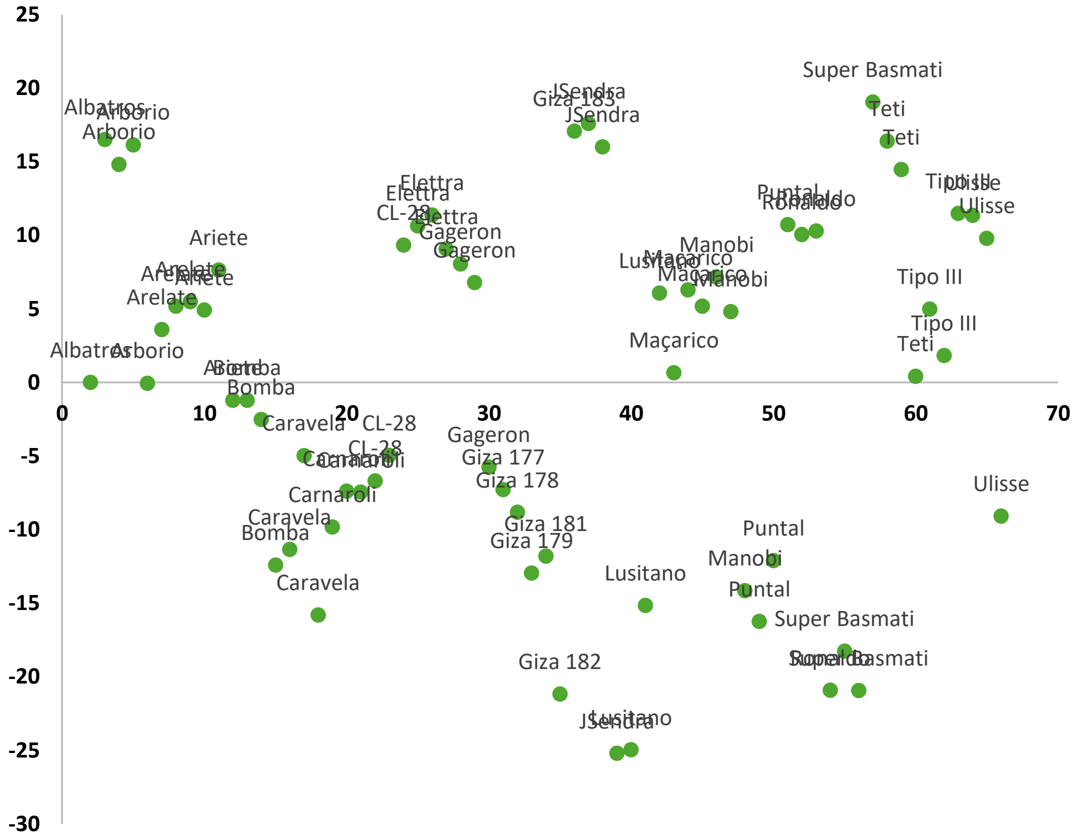


NIR spectra rice flour (2nd derivative)

Principal Component Analysis – rice code

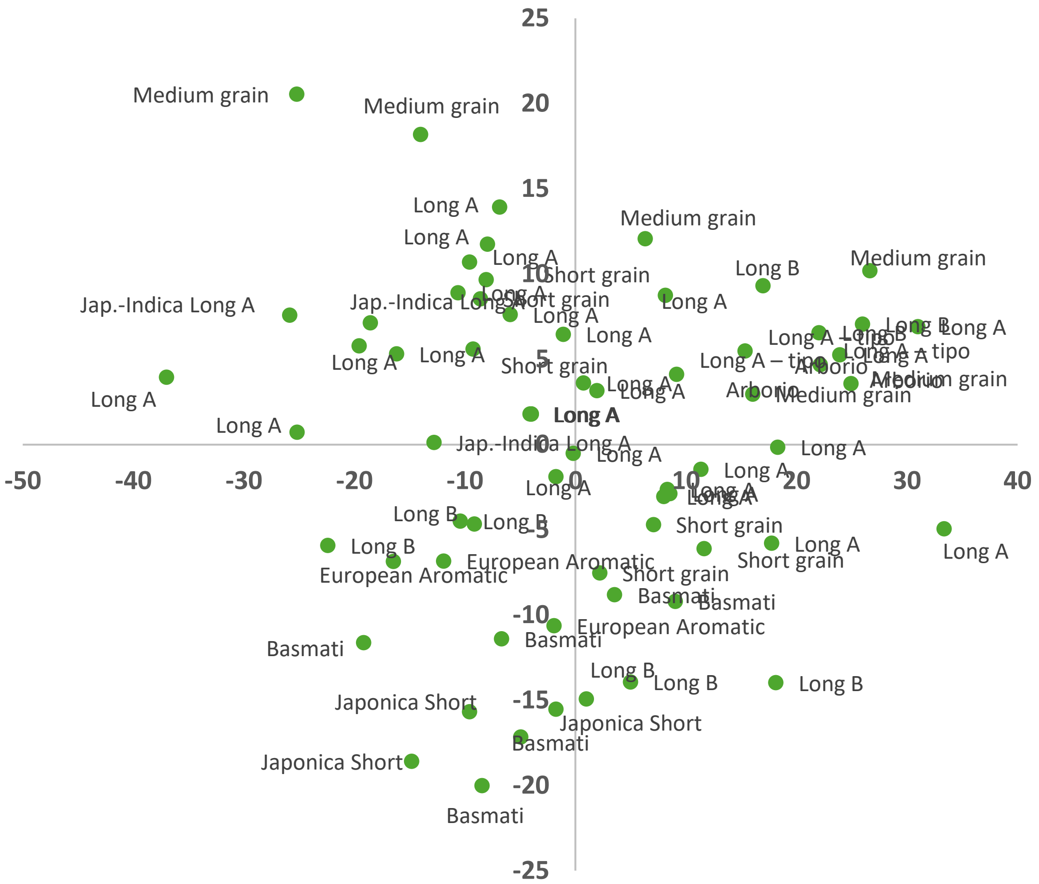


PCA based on NIR spectra obtained from rice (**flour)**

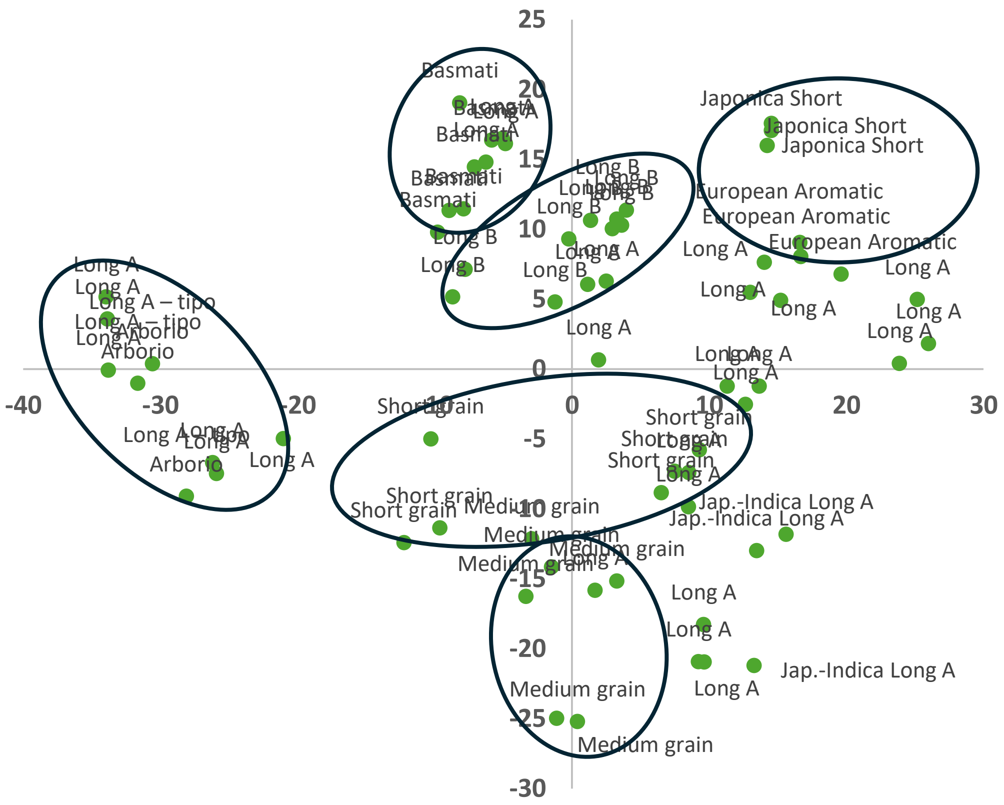


PCA based on NIR spectra obtained from rice (**grain)**

Principal Component Analysis – rice type

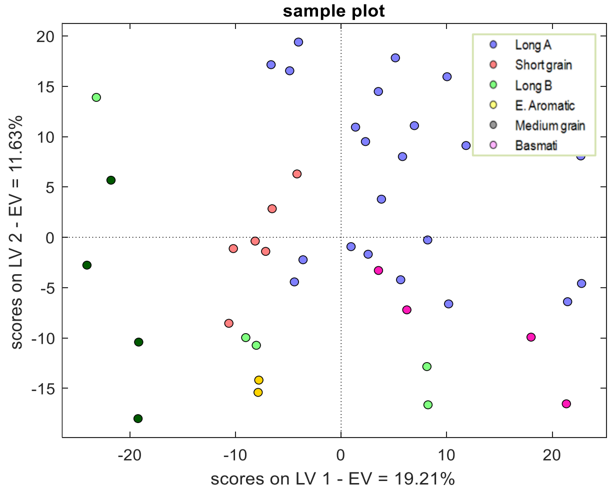


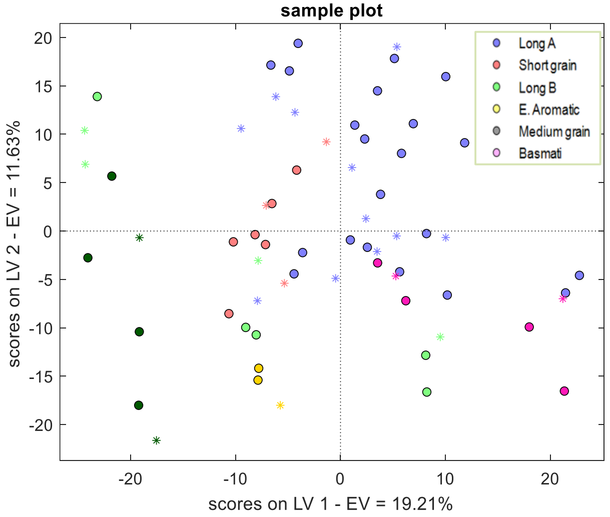
PCA based on NIR spectra obtained from **flour**



PCA based on NIR spectra obtained from **grain**

Classification PLS-DA – Rice flour



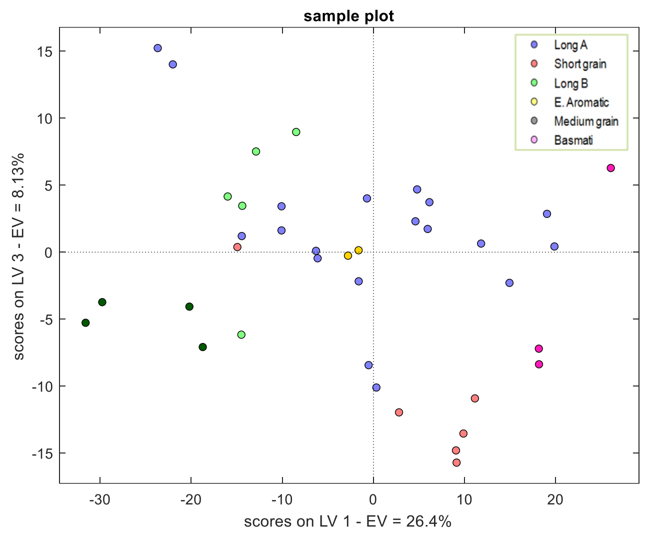


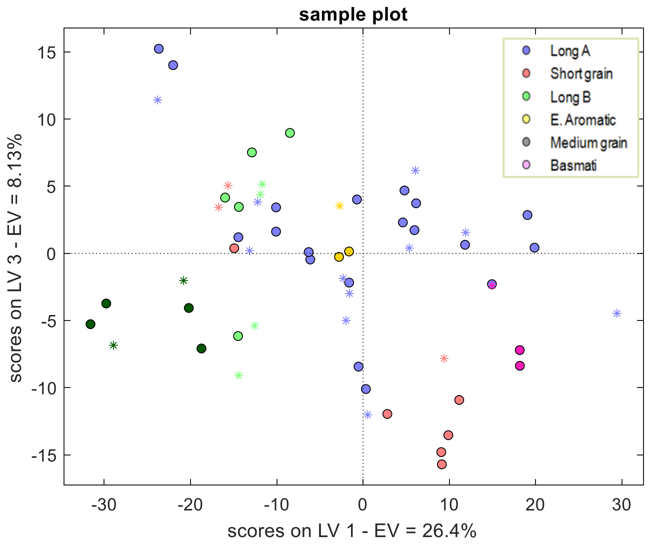
Explained variance: 56%
Error rate: 17%
Test error rate: 38%
Not assigned: 23%

Cross-validation
Error rate CV: 31%
Accuracy: 68%
Not-assigned: 28%

Prediction on external samples
No-error rate: 62%
Accuracy: 65%
Not-assigned: 26%

Classification Methods





PLS-DA – Rice grain

Training
Error rate: 0%
Accuracy 100%

Cross-validation
No-error rate: 43%
Accuracy: 57%

Prediction on external samples
No-error: 100%
Accuracy: 67%

K-NN - Rice grain

Training
Error rate: 40%
Accuracy 60%

Cross-validation
No-error rate: 62%
Accuracy: 63%

Prediction on external samples
No-error: 91%
Accuracy: 61%

Conclusions

► NIR spectroscopy and machine learning are effective tools for developing prediction models related to **pasting** and **biochemical** parameters.

► This strategy represents a promising approach for rice **quality assessment** and **classification based** on PLS-DA based on flour and grain.

IN THE NEXT FUTURE:

Improvement of the models using huge amount of experimental data for prediction and classification procedures.

Pedro Sampaio & Carla Brites

thank you!

INIAV Team Collaboration:
Ana Castanho, Andreia Soares, Bruna Carbas, Carlota Gonçalves, Cristiana Pereira, Inês Sousa, Manuela Lageiro