Deliverable 1.6

Predictive models of rice authenticity

November 2024

http://trace-rice.eu

TECHNICAL REFERENCES

Project Acronym	TRACE-RICE
Project Coordinator	Carla Moita Brites
	carla.brites@iniav.pt
Project Duration	September 2020 –October 2024 (50 months)

Deliverable No.	D1.6
Dissemination level*	CONFIDENTIAL
Work Package	WP 1 - TEST SOLUTIONS FOR AUTHENTICITY, ORIGIN AND TRACEABILITY
Task	TASK 1.4 BUILD PREDICTIVE GENERATION MODELS
Lead beneficiary	INIAV
Contributing beneficiary/ies	
Due date of deliverable	31th October 2024
Actual submission date	17 December 2024

	Date	Beneficiary	Change
V1	31/10/2024	INIAV	Version sent to coordinator
V2	18/12/2024	INIAV	Final version approved by project coordinator

Written by: Pedro Sampaio & Carla Brites

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

EXECUTIVE SUMMARY

The TRACE-RICE project has developed advanced predictive models to assess rice quality traits, such as amylose content and viscosity profiles, and classify rice types. These models leverage Near-Infrared (NIR) spectroscopy and are hosted on the DMPortal, the national Dataverse instance managed by Biodata.pt. DMPortal serves as a long-term centralized repository for genomic, phenotypic, and associated metadata generated under the TRACE-RICE project. The models can be accessed at:

Predictive Models for Rice Quality Traits Using MATLAB - TRACE-RICE Dataverse

These models were developed using MATLAB (R2023a) and integrate cutting-edge techniques, including:
- **Partial Least Squares (PLS)**
- **Partial Least Squares Discriminant Analysis (PLS-DA)**
- **Support Vector Machines (SVM)**
- **Machine Learning**
- **Artificial Neural Networks (ANN)**

The PLS, iPLS, siPLS, and mwPLS models were implemented with MATLAB and the **iToolbox** (available at http://www.models.life.ku.dk/itoolbox), while classification procedures utilized the **Classification Toolbox (v2.0)** from Milano Chemometrics and QSAR Research Group (http://michem.disat.unimib.it/chm).

Methodology Overview
The workflow began with the collection of raw NIR spectra from rice samples using a transflection setup. These spectra were preprocessed with industry-standard techniques, such as:
- **Standard Normal Variate (SNV) Transformation**
- **Multiplicative Scatter Correction (MSC)**
- **Smoothing Derivative Algorithms**

After preprocessing, predictive models were developed using supervised and unsupervised techniques, including:
- **Principal Component Analysis (PCA)**
- **Partial Least Squares Discriminant Analysis (PLS-DA)**
- **Support Vector Machines (SVM)**
- **k-Nearest Neighbors (KNN)**

These models enable precise predictions of rice authenticity and key quality traits, facilitating advanced classification and quality control measures.

Dataset Components
The dataset hosted on the Dataverse repository includes the following models:
1. **AMYLOSE_MODEL_FOOD_CHEMISTRY**
   - A MATLAB-based model for predicting amylose content using NIR spectroscopy and PLS chemometric algorithms.

2. **PLS-DA_SVM_MODELS**
    - o MATLAB-based models for identifying rice flour types using PLS-DA and SVM methods.
3. **ANN_MODELS_ApSci**
    - o Machine learning-based MATLAB models for predicting rice pasting parameters using NIR spectroscopy.
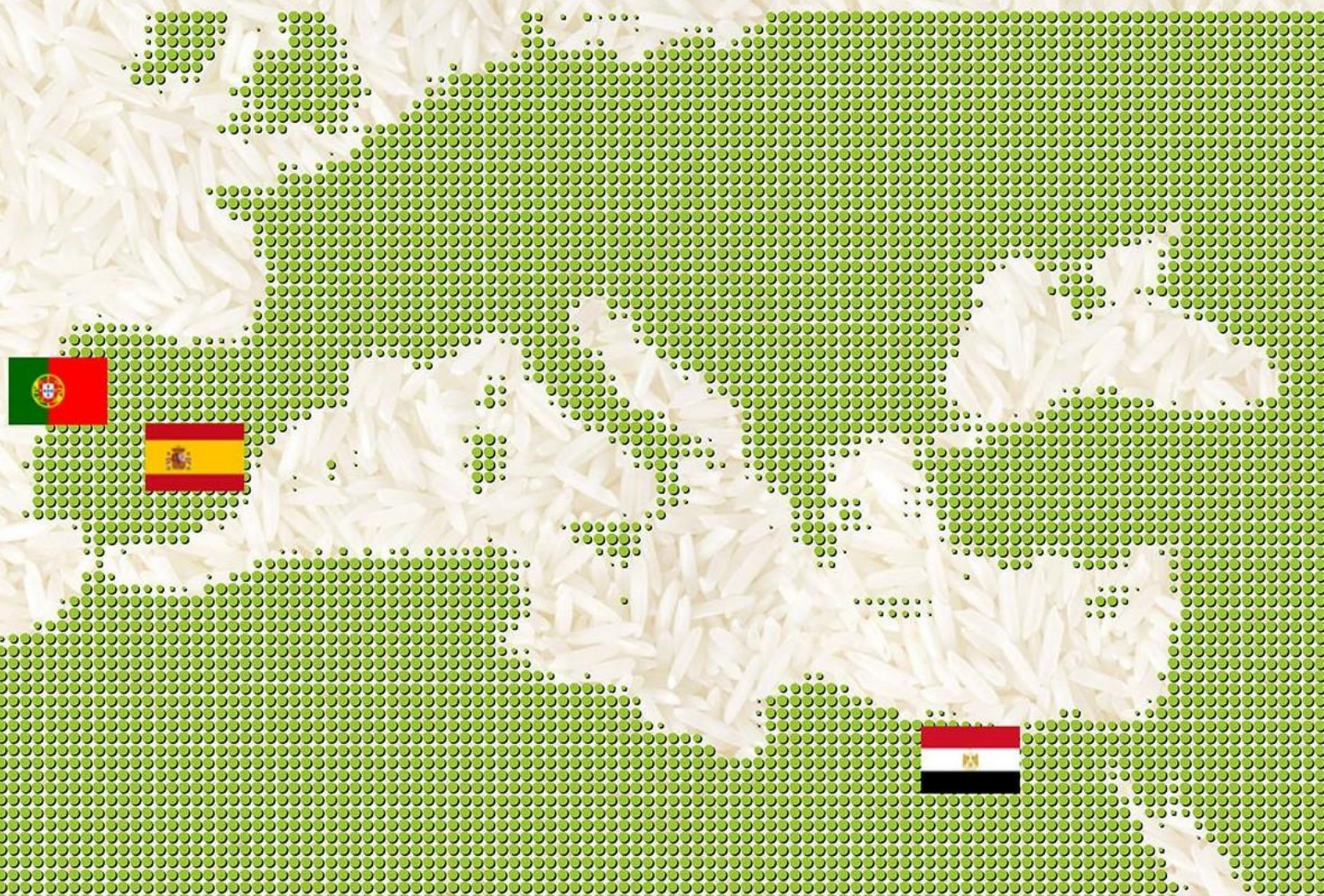4. **ANN_MODELS_PASTING_FOODS**
    - o MATLAB-based ANN models for evaluating rice quality based on grain physical parameters.
    - o

Model Distribution

All models are provided as .mat files in a compressed .rar format. Each model includes performance metrics to ensure reproducibility and facilitate evaluation.

TRACE-RICE Consortium

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

iBET
Instituto de Biologia
Experimental e Tecnológica

iTQB
UNIVERSIDADE
NOVA
DE LISBOA

CASA DO
ARROZ

ernesto
morgad
desde 1920

MATER
DYNAMICS

BGI
www.bgi.pt

Grupo Desarrollo

iata
Instituto de Agroquímica

ALEXANDRIA
UNIVERSITY