



**HAL**  
open science

## **Analysis of imbalanced data from porosity measurements on high-pressure hydrogen vessel (Type IV)**

Lina Achour, Salvador Madrigal Castillo, Wenlong Chen, Sébastien Destercke, Zoheir Aboura, Kamel Khellil

### ► **To cite this version:**

Lina Achour, Salvador Madrigal Castillo, Wenlong Chen, Sébastien Destercke, Zoheir Aboura, et al.. Analysis of imbalanced data from porosity measurements on high-pressure hydrogen vessel (Type IV). ASTM INTERNATIONAL – Conference on Hydrogen in Materials, Jun 2025, La Rochelle, France. <hal-05109944>

**HAL Id: hal-05109944**

**<https://hal.science/hal-05109944v1>**

Submitted on 12 Jun 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

# Analysis of imbalanced data from porosity measurements on high-pressure hydrogen vessel (Type IV)

Lina ACHOUR<sup>1,2,3</sup>, Salvador Madrigal Castillo<sup>3</sup>, Wenlong Chen<sup>3</sup>, Sébastien DESTERCKE<sup>2</sup>, Zoheir ABOURA<sup>1</sup>, Kamel KHELLIL<sup>1</sup>

<sup>1</sup> Université de technologie de Compiègne, Roberval, Compiègne, France

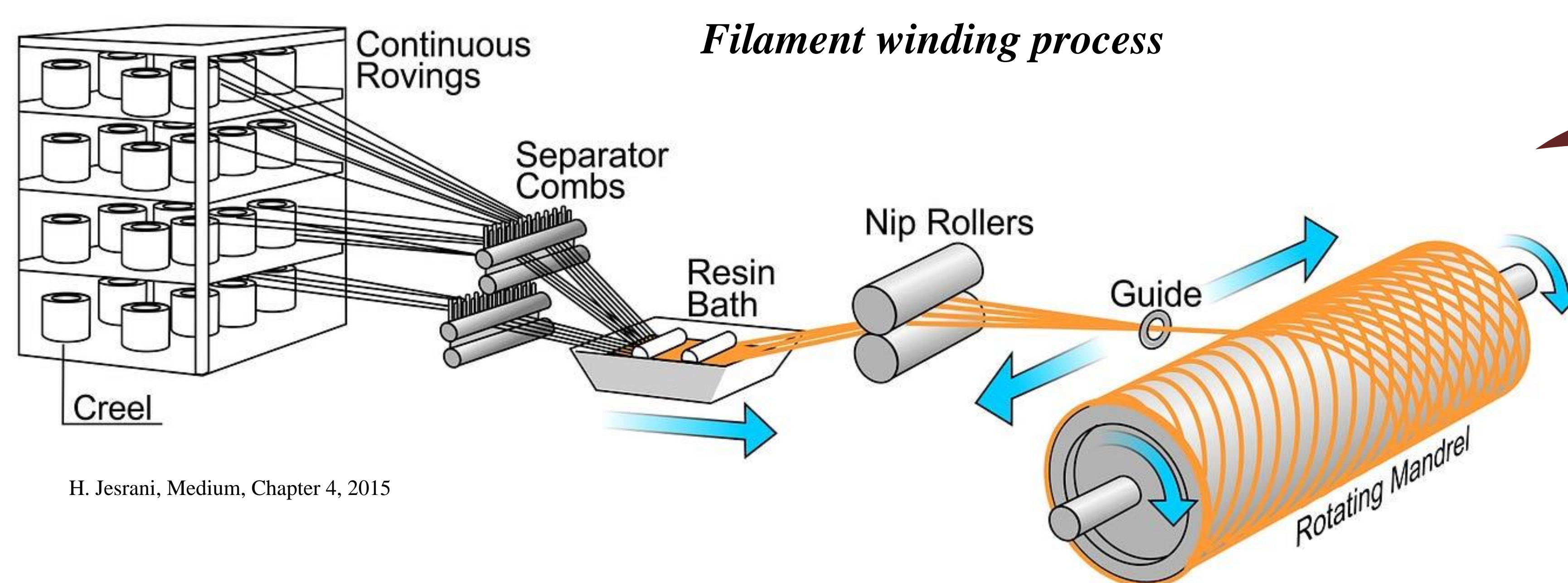
<sup>2</sup> Université de technologie de Compiègne, CNRS, Heudiasyc, Compiègne, France

<sup>3</sup> Opmobility, ZAC du Bois de Plaisance, 60280 Venette, France

## Context & Objectif

Real-time monitoring of the impregnation quality of carbon fiber with epoxy resin is crucial for the safe and efficient manufacturing of Type IV high-pressure hydrogen vessels (HPV). These vessels must withstand pressures up to 700 bar while maintaining their mechanical properties.

The objective of this study is to **detect and classify porosity levels** in the HPV composite structure, by leveraging **artificial intelligence (AI) algorithms** and using key **filament winding process parameters** such as fiber tension, resin volume, winding angle, and mandrel speed as input variables. The study is primarily focused on addressing the significant **imbalance** between porosity classes.

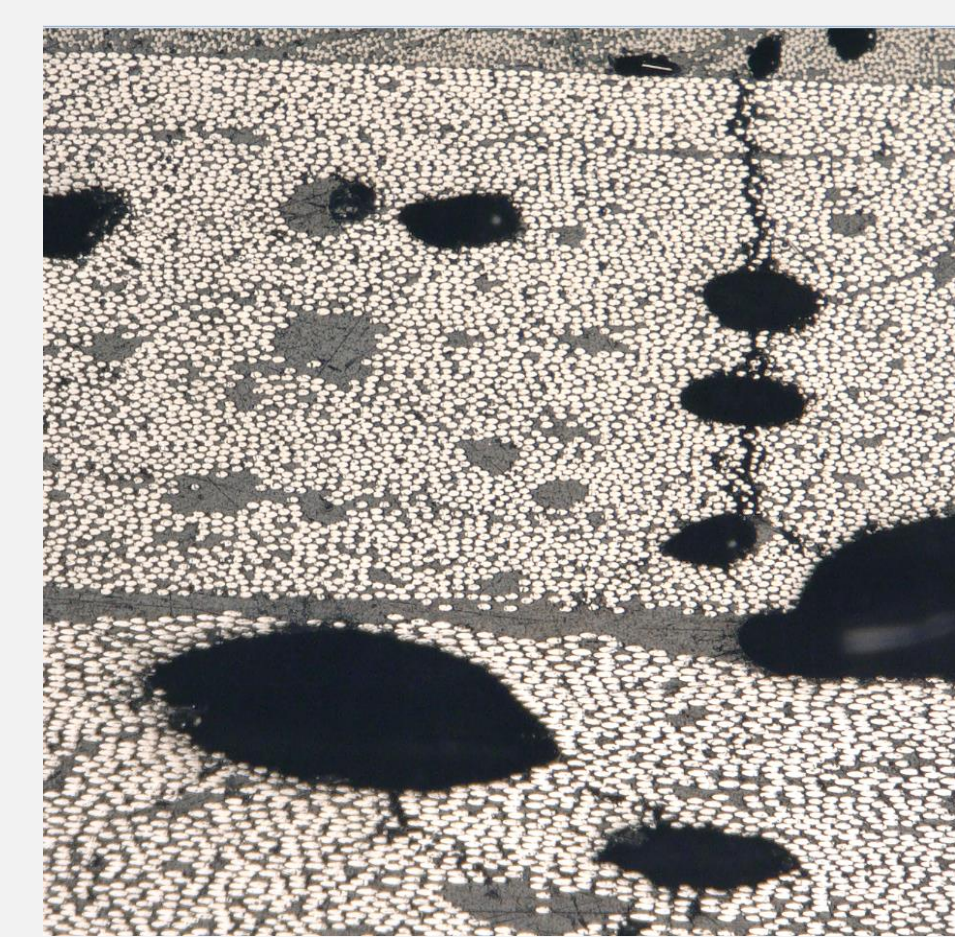


## Manufacturing Process & Impregnation Risks

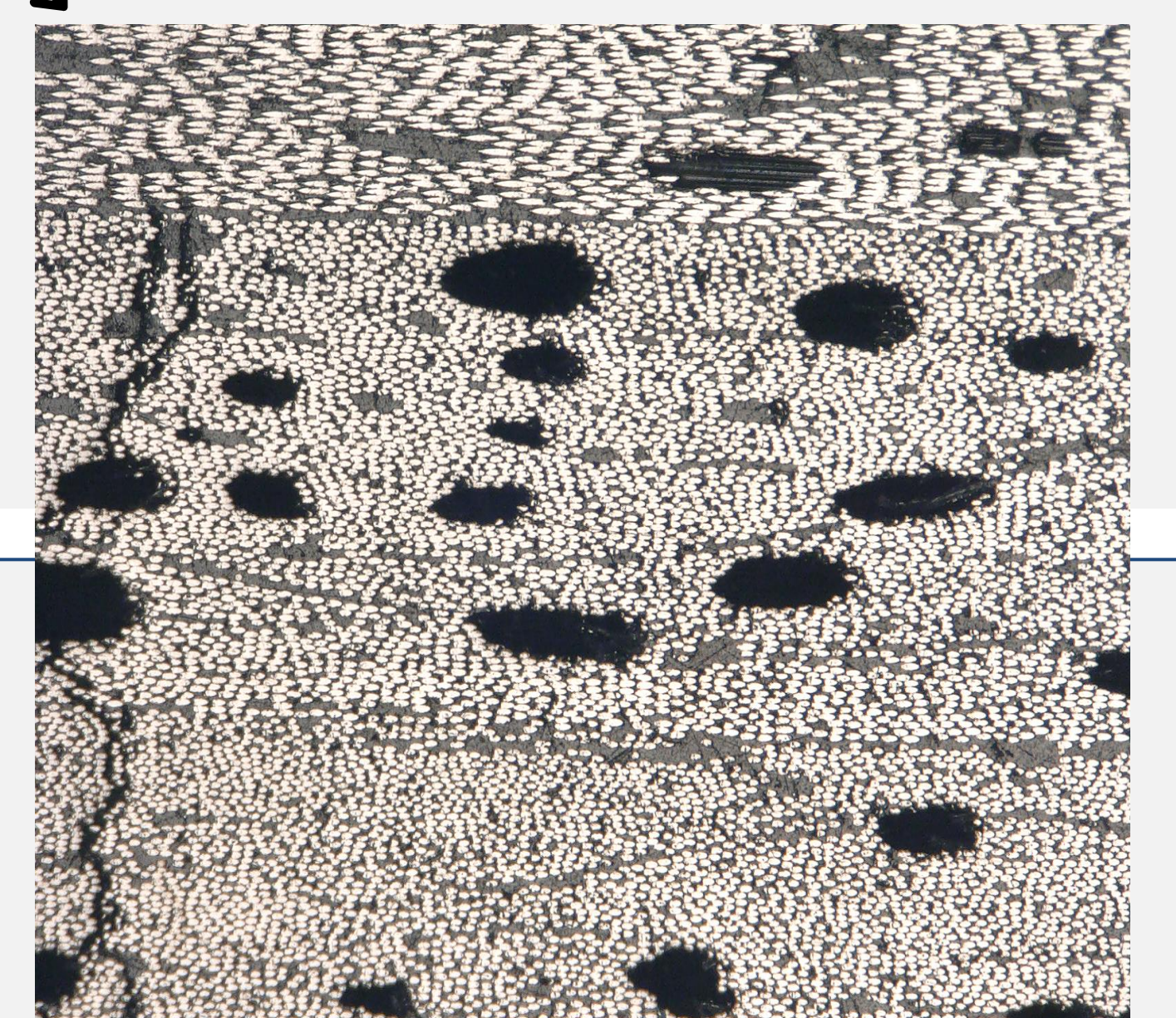
**The process is complex and multiparametric**

**The dangers of inadequate impregnation:**

- Reduced mechanical behavior,
- Fatigue damage,
- Increased porosity,
- Non-uniform resin distribution,
- Poor interlaminar adhesion,
- Failure of the vessel in the compliance tests required by the standards,
- ...



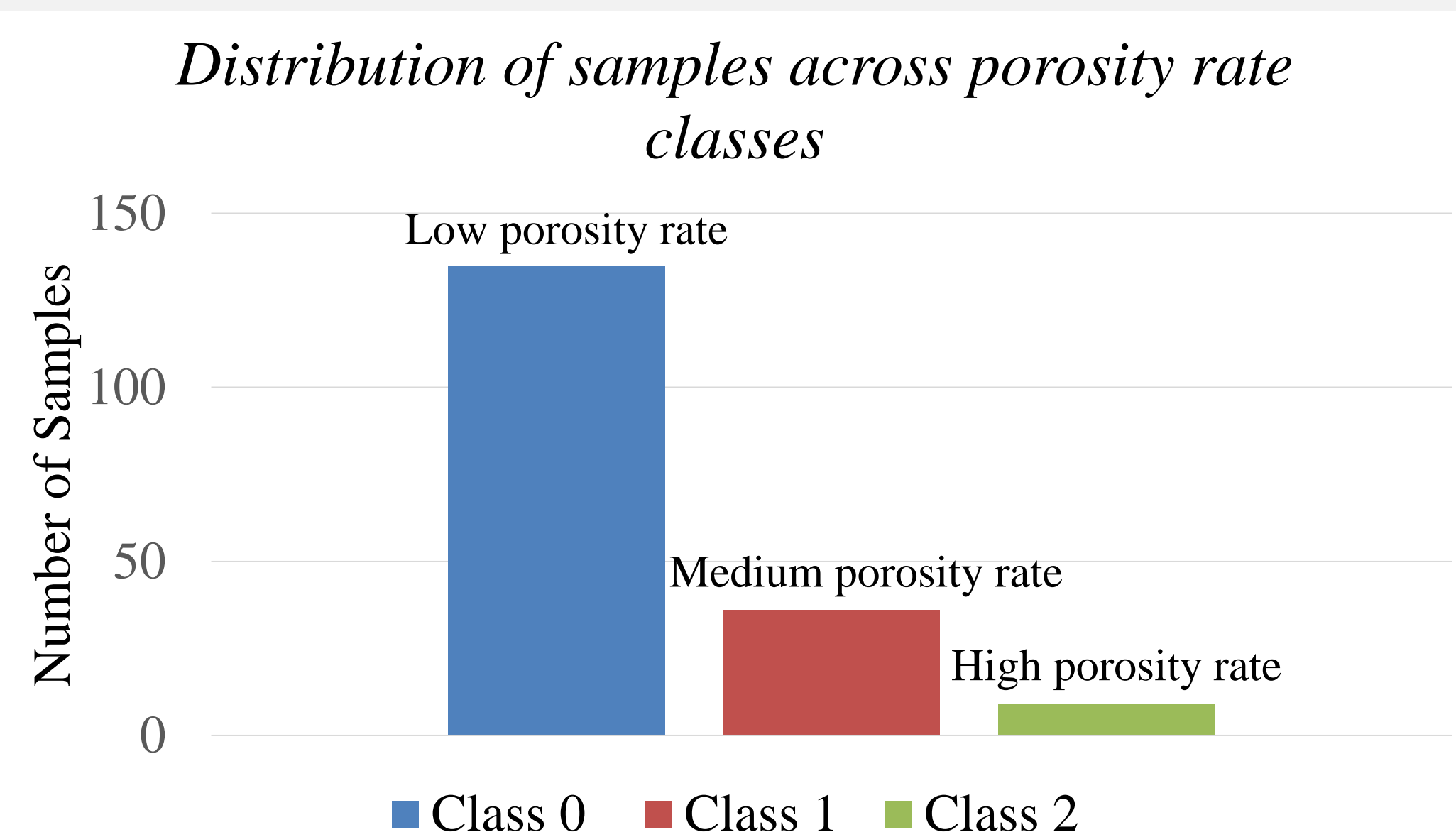
Microstructural view of porosity in composite material, with crack propagation between pores



## Approach & Methodology: Data processing

- \***8 potential predictors:** Data comes from the production process
- ✓ Directly linked to the manufacturing process
- ✗ Noise and imperfections
- ✗ Missing or incomplete values due to sensor limitations and measurement errors.

\***The porosity data** obtained from the **X-ray tomography** of each winding layer of **12 vessels**.



## Tested algorithms:

- **Logistic Regression (LR)**
- **Weighted Logistic Regression (WLR)**
- **Weighted Decision Trees (WDT)**
- **Discrete Minimax Classifier (DMC)**, a robust algorithm developed by Dr. Cyprien GILET (Heudiasyc, UTC); it's designed to handle **imbalanced datasets** by minimizing the maximum classification risk across classes. Its decision rule adapts to uncertain or rare class distributions, making it particularly effective for detecting critical but infrequent defects.

## Obtained results

### Average confusion matrix for each method

True class	Predicted class		
	0	1	2
0	0.94	0.06	0
1	0.79	0.20	0
2	1	0	0

**Logistic regression (LR)**

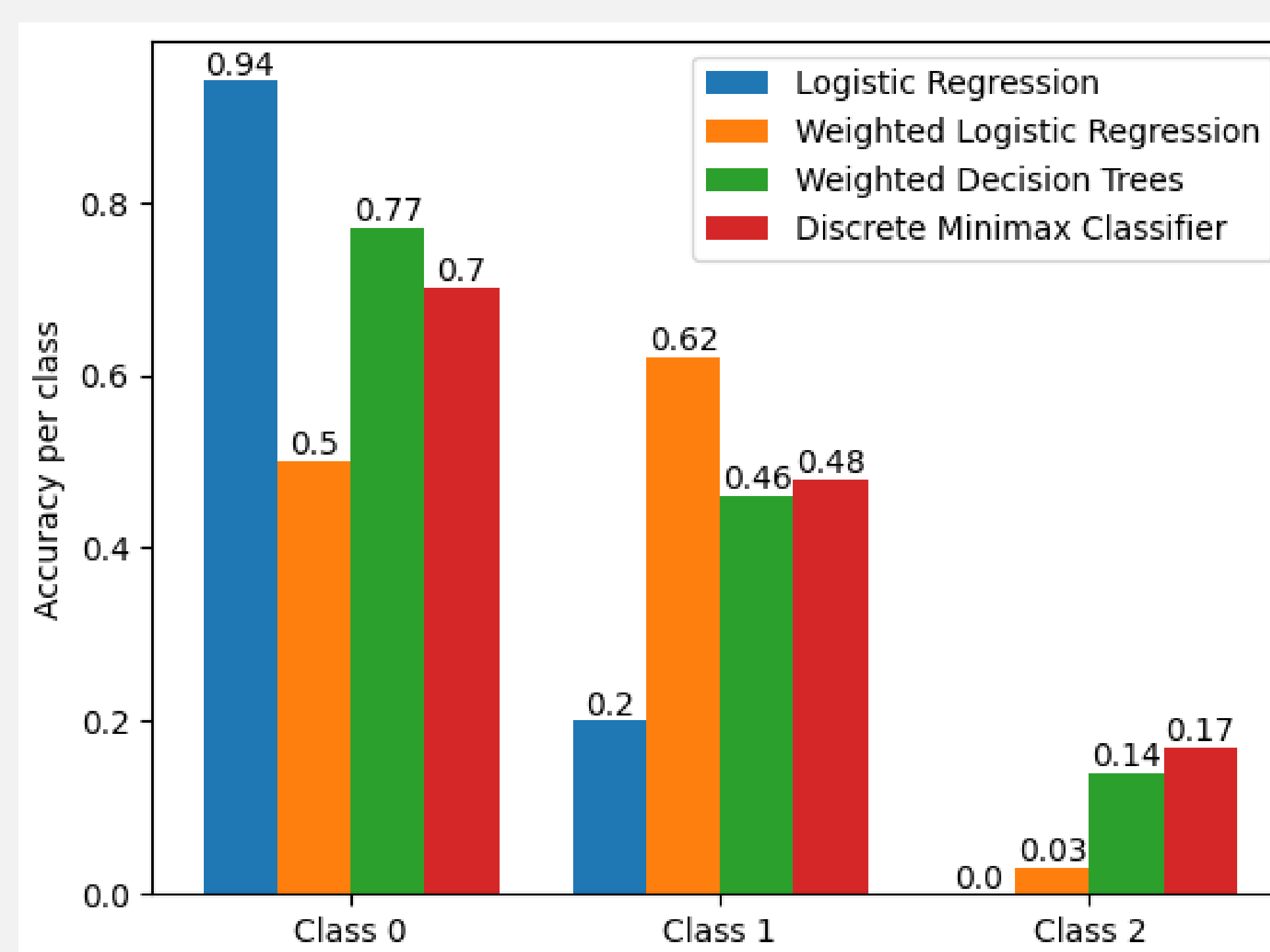
True class	Predicted class		
	0	1	2
0	0.77	0.10	0.13
1	0.38	0.46	0.16
2	0.63	0.23	0.14

**Weighted tree**

True class	Predicted class		
	0	1	2
0	0.50	0.27	0.24
1	0.26	0.62	0.12
2	0.70	0.27	0.03

**Discrete minimax**

### Accuracy per class for each method with three classes



## Conclusion & perspectives

Our study shows that the Discrete Minimax Classifier (DMC) improves critical defect detection in highly imbalanced datasets, ensuring greater safety and reliability of hydrogen storage systems, and opening opportunities for wider defect detection applications.

Future work will focus on enriching the dataset with additional samples and input variables to enhance model robustness and improve the detection of critical defects. Additionally, further application of these methods to other datasets and challenges is expected.