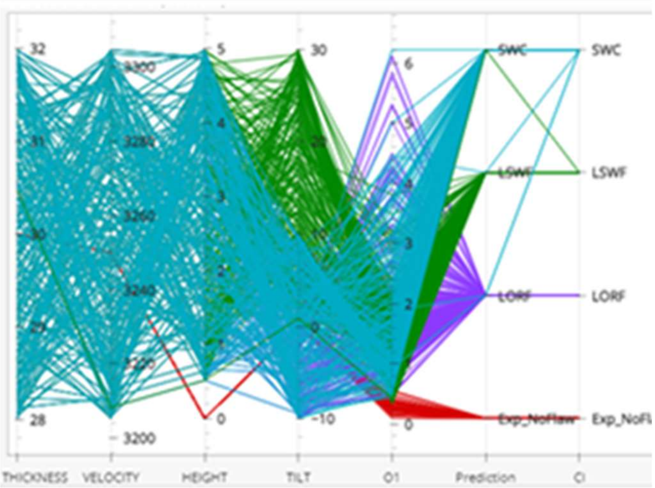
CIVA DATA SCIENCE

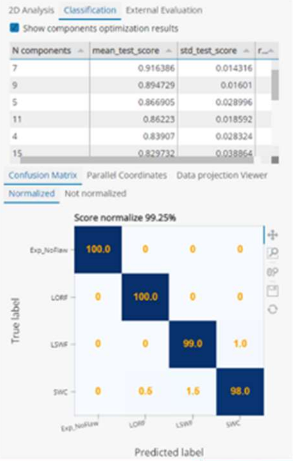
DATA management for Machine Learning NDE assisted diagnostics

Real use of “AI” in NDE suffers from the difficulty of getting sufficient and relevant data to train algorithms. **Simulation can overcome this situation by efficiently providing massive relevant data sets.** This new CIVA Data Science module allows you to bring together CIVA simulations, CIVA Analysis, and data management tools to design and validate **Machine-Learning based diagnostic models** in an NDE environment.

Train then Evaluate classification and detection prediction models:

Parametric simulations in CIVA will let you build relevant databases for aided diagnostic strategies. CIVA Data Science will then provide you the tools **to manipulate databases** of simulated and experimental data. Script your own extraction criteria **with Python**, and/or use the so-called projection algorithms (PCA and UMAP) to determine relevant parameters that split learning data into **classification** groups. CIVA Data Science includes visualization maps and graphs such as 2D plots, parallel plots, and confusion matrices to set up training parameters. Another module allows you to train a **detection** model for one class success/fail evaluation. Once the diagnostic model is trained, you can compare it to an external (experimental) database that was previously imported into CIVA Data Science to check its efficiency.



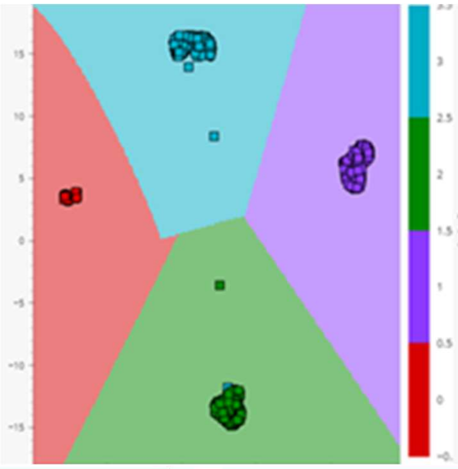


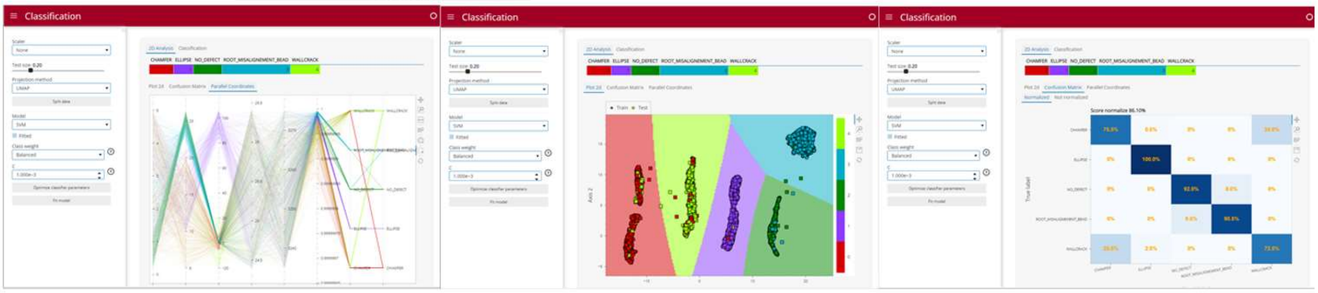
☑ Show components optimization results

N components	mean_test_score	std_test_score	r...
7	0.916386	0.014316	
9	0.894729	0.01601	
5	0.866905	0.028996	
11	0.86223	0.018592	
4	0.83907	0.028324	
15	0.829732	0.038864	

Score normalize 99.25%

True label \ Predicted label	Exp_NoFlaw	LORF	LSWF	SWC
Exp_NoFlaw	100.0	0	0	0
LORF	0	100.0	0	0
LSWF	0	0	99.0	1.0
SWC	0	0.6	1.5	98.0





INSIDE CIVA DATA SCIENCE…

CIVA Data Science includes the following tools:

- **DATABASES & METAMODELS:** Collect simulated and experimental data, define **output criteria**, generate metamodels, check data dimensions and consistency.

- **DATABASES FUSION:** **Merge** different data sources (different parametric studies or acquisition data files, etc.) to increase the number of samples or stack the criterion.

- **PYTHON NOTEBOOKS:** Use **Python script** to customize data sets to your needs.

  CIVA Data Science is available as an add-on to the CIVA Script module.

- **CLASSIFICATION:** Define the target classes for the defects to be characterized. **Pre-analyze** data on 2D plots then compare different Scaler, Projection, and Classifier algorithms and parameters to **train your data**. Finally, **evaluate** the diagnostic model on an external data set.

- **OUTLIER DETECTION:** Train the model with one class and set the threshold for **success/fail evaluation**. It can be for instance flaw / no flaw detection by training it on no flaw databases.

- **PREDICTION ACCURACY:** Evaluate the accuracy of the metamodels you generated from the databases and simulation variations.