PelletRayTion: Unsupervised Contaminants Detection in Pelleted Seeds Using High-Throughput Tomography and 3D β-VAE

Sherif.hamdy@geves.fr

Sherif HAMDY[1, 2], Blandine DAVIAUD[1], Laurence LE Corre[1], Ludivine SOUBIGOU-TACONNAT[1], Philippe PANNETIER[1], Pejman RASTI[2], Aurélie CHARRIER[1] and David ROUSSEAU[2]

1 GEVES, National Seed Testing Station, France

2 Université d'Angers, LARIS, France

Introduction: Seed coatings improve handling and germination [1], but purity checks remain complex for this type of seeds, while contaminants pose biosecurity risks. We propose a non-destructive pipeline combining high-throughput X-ray CT with unsupervised learning (3D β-VAE) to detect contaminants trained only on pure seeds. The method requires no massive datasets or exhaustive annotations, ensuring reliable and scalable quality control.

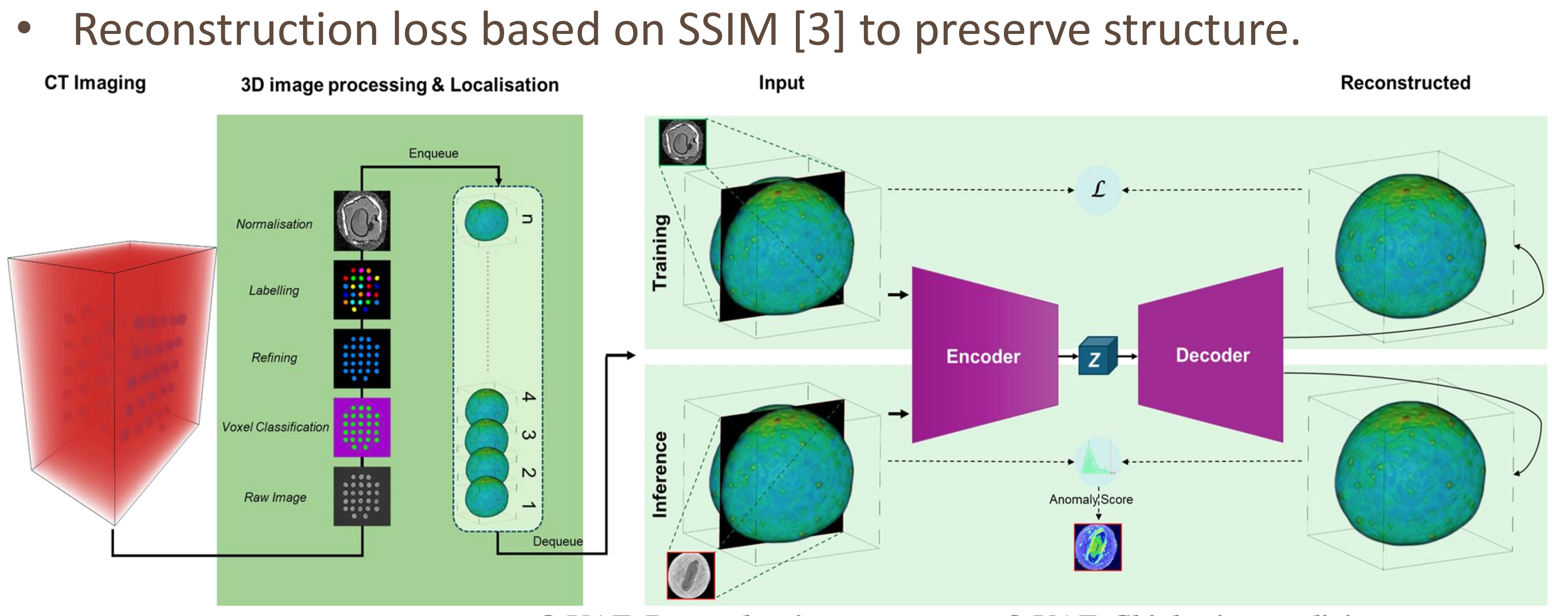
Materials & Methods

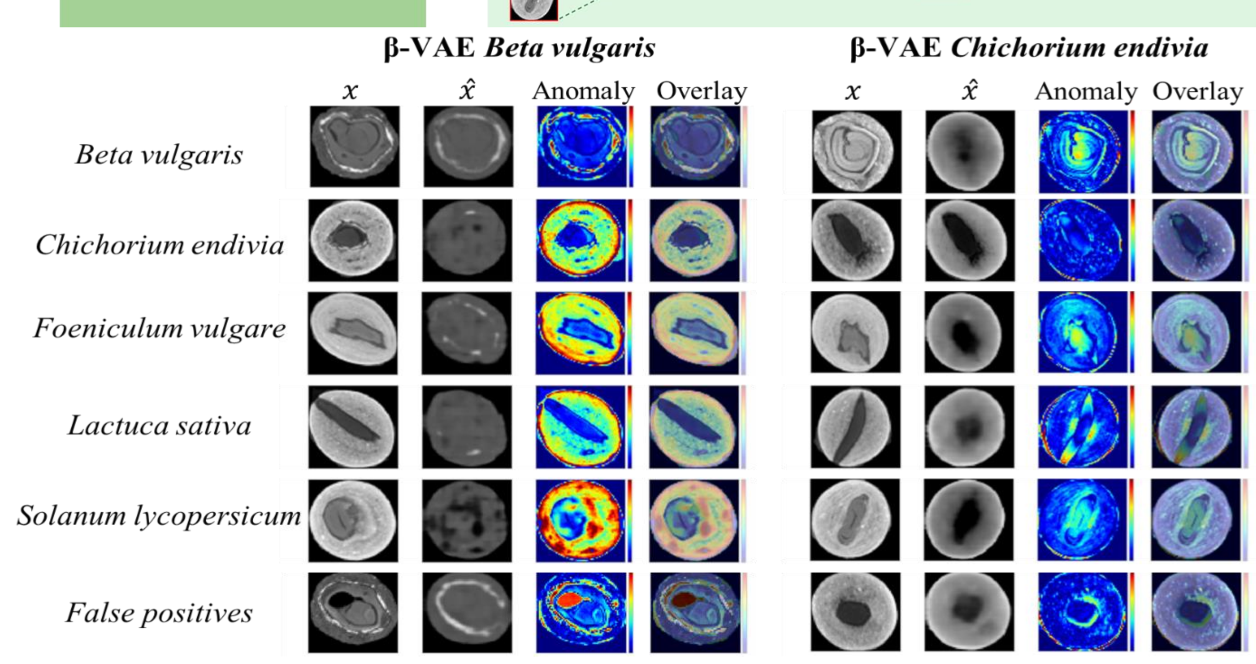
Biological material:

- Pure seeds: *Beta vulgaris,* or *Cichorium endivia*
- Contaminants: *Foeniculum vulgare, Solanum lycopersicum, Lactuca sativa*, and (*Beta vulgaris*, or *Cichorium endivia*).

Model:

- 3D β-VAE [2] trained on pure seeds CT volumes (64×64×64).
- Architecture: encoder (3D conv layers) → latent space z → decoder (3D transposed conv layers).
- Reconstruction loss based on SSIM [3] to preserve structure.





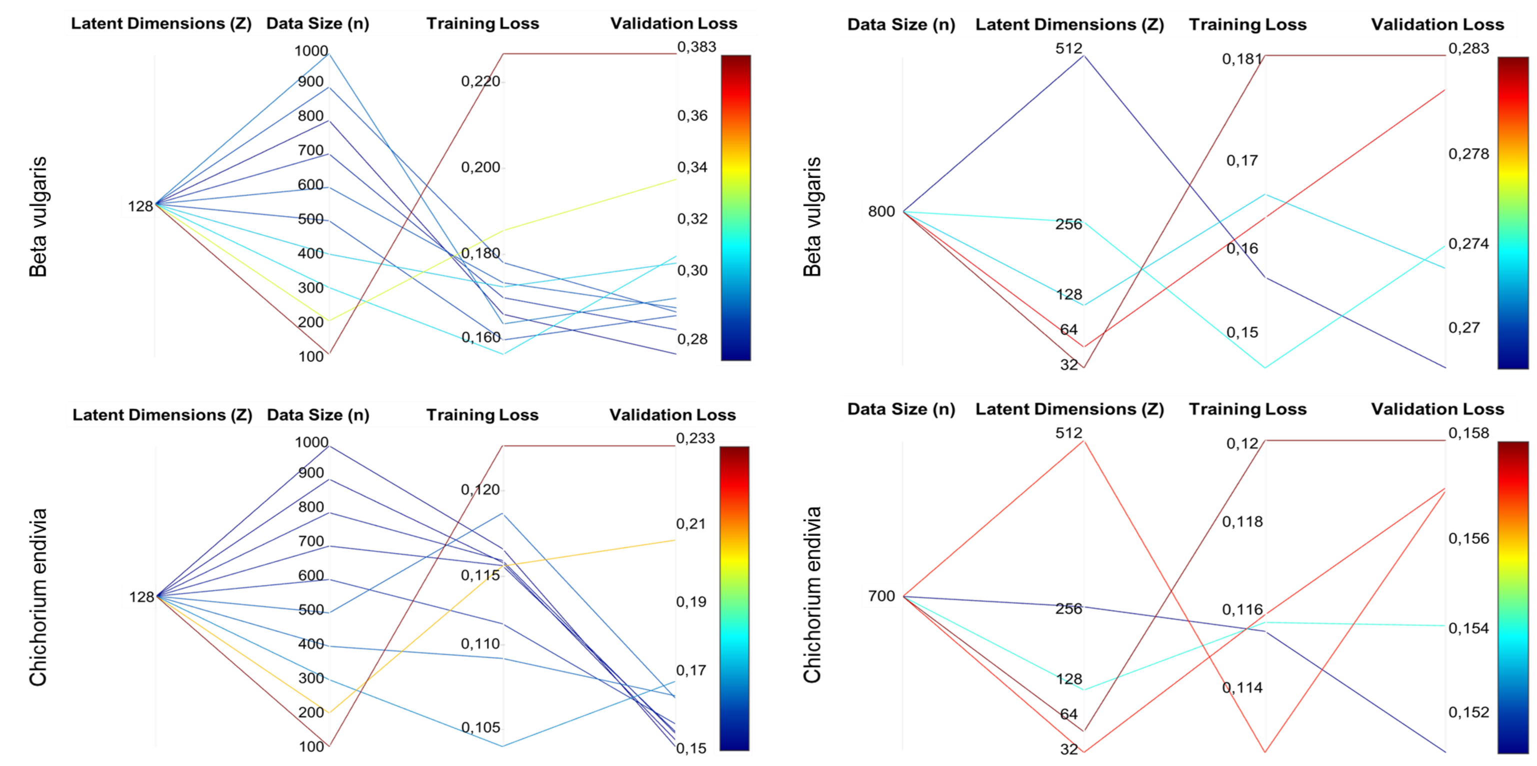
Experiments

Exp 1 – Training Data Size (*n*):

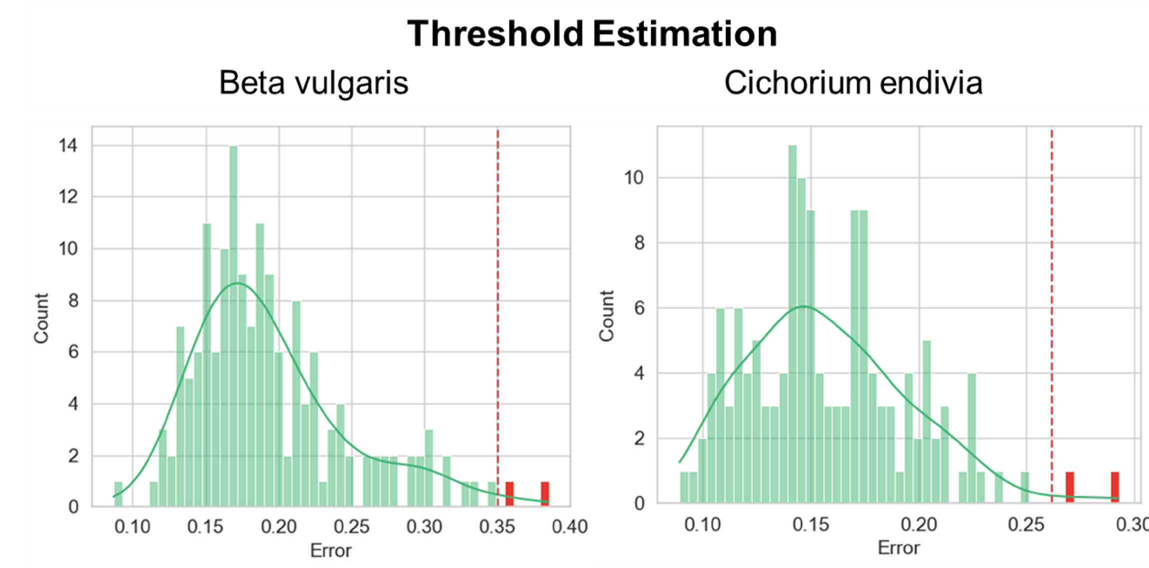
- More data → lower error
- Optimal: n = 700 - 800

Exp 2 – Latent Space Size (*z*):

- *B. vulgaris*: best at *z* = 512
- *C. endivia*: best at *z* = 256

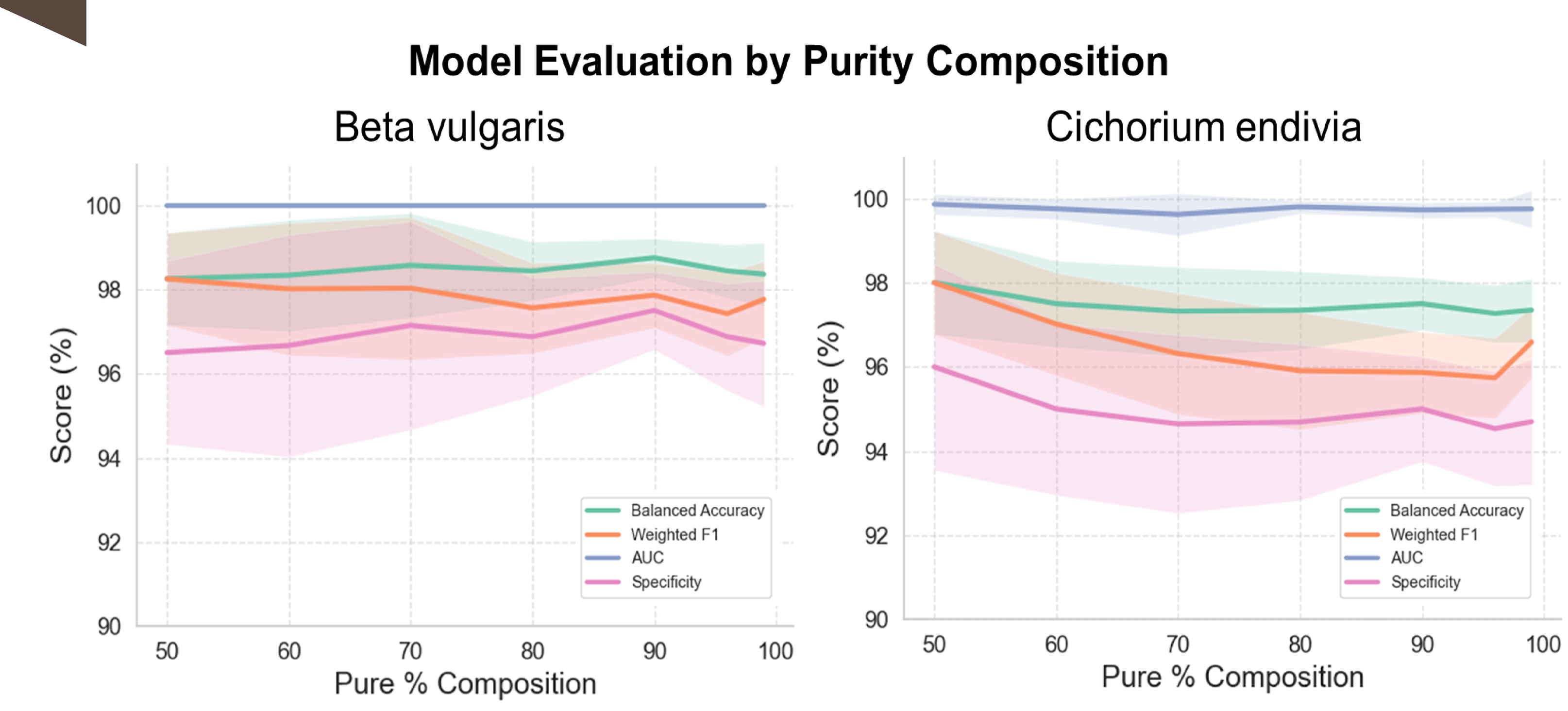


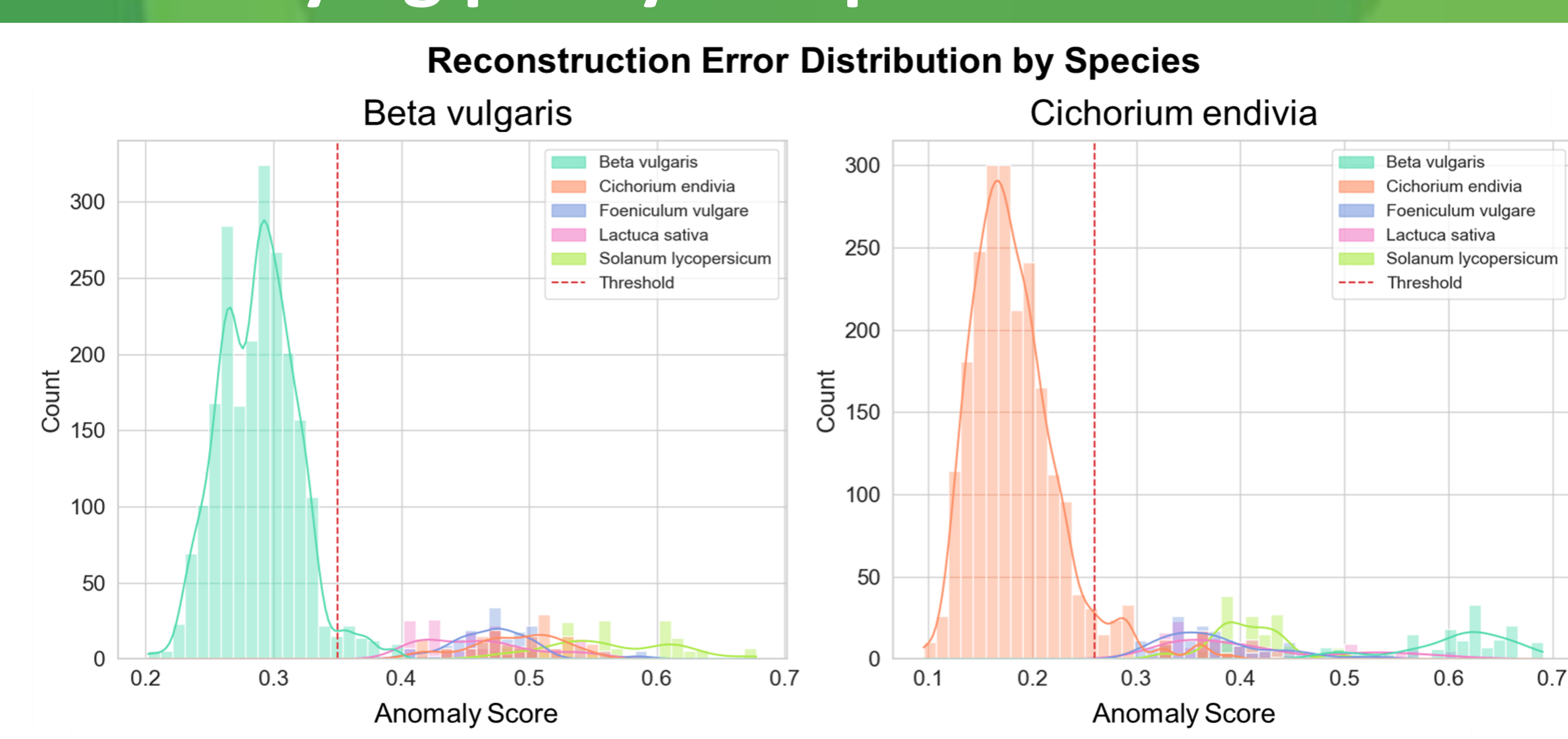
- Data efficiency: an optimal data range exists.
- Latent space dimensionality is species-dependent; selecting appropriate ***z*** improves generalisation.



- Threshold = 99th percentile of pure-seed errors

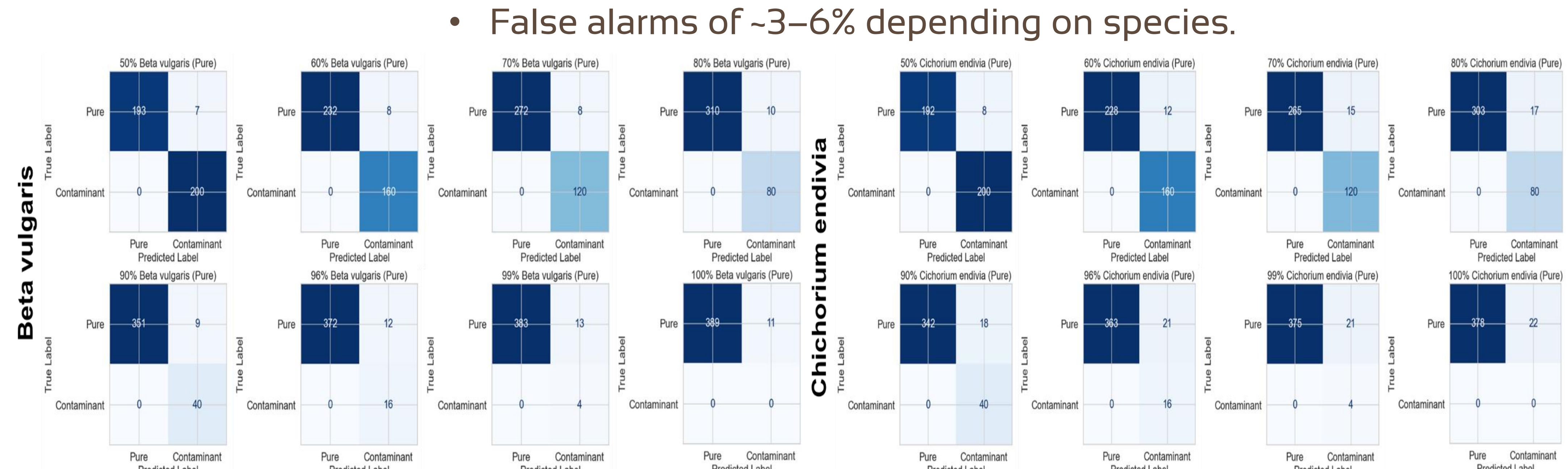
Contaminant detection performance on varying purity compositions





- Tested 1–50% contamination with four replicates per composition.
- High performance:
  - *B. vulgaris*: Acc 98.5 ± 1%, F1 97.8 ± 1.2 %, AUC 100 ± 0.0%, Specificity 96.9 ± 1.9%.
  - *C. endivia*: Acc 97.5 ± 1%, F1 96.5 ± 0.4%, AUC 99.8 ± 0.3%, Specificity 94.9 ± 1.9%.
- Steady performance across compositions.
- Low variability at each composition.

- Contaminants never classified as pure.
- False alarms of ~3–6% depending on species.



Contributions

- ✓ High-throughput, non-destructive method enables robust, unsupervised contaminant detection in pelleted seeds.
- ✓ No massive datasets or exhaustive annotations.
- ✓ Demonstrates high performance across contamination levels, with low false positives and zero false negatives for pure seeds.
- ✓ Scalable tool for seed lot quality assurance.

References

[1] Halmer, P. (2000). Commercial seed treatment technology. Seed Technology, 22(1), 13–20.Kingma, D. P., & Welling, M. (2014).

[2] Higgins, I., Matthey, L., Pal, A., et al. (2017). β-VAE: Learning disentangled representations. ICLR.

[3] Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process., 13(4), 600–612.