Efficient unsupervised segmentation of X-ray fluorescence images in wheat pathology

George L. Malone[1], Adam Sparks[2], Fatima Naim[2], Mark Gibberd[2], and Nicola J. Armstrong*[1]

[1] School of Electrical Engineering, Computing, and Mathematical Sciences (EECMS) | [2] CCDM | * Primary supervisor

ABSTRACT

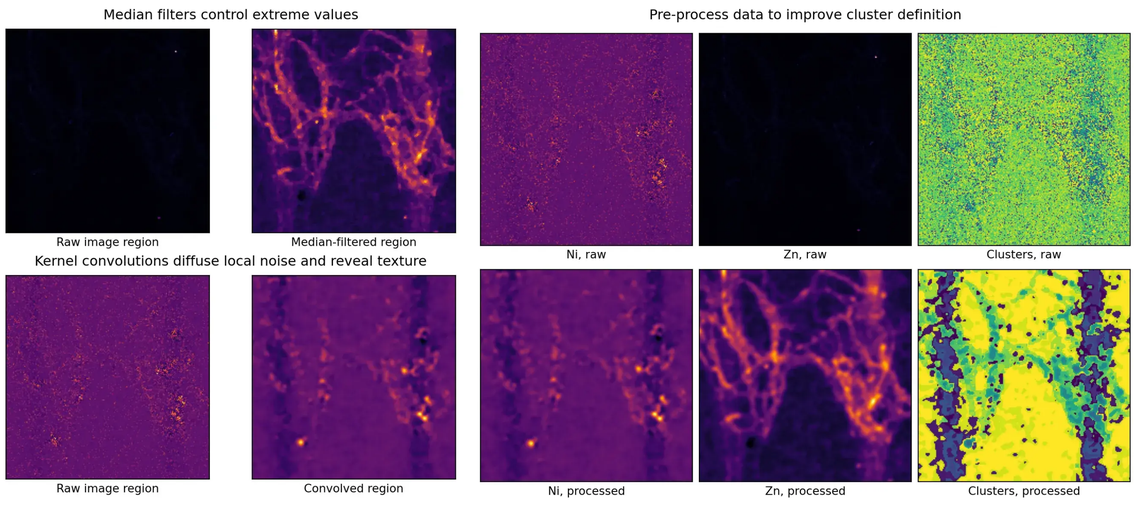
X-ray fluorescence microscopy (XFM) is a highly sensitive imaging method used for measuring metal nutrients in wheat leaves in intensive detail. Unsupervised clustering can segment XFM images with reduced labour and bias, but most methods do not work in XFM due to computational limits. Clustering compressed data may detect interesting features with greatly reduced computational burden, enabling a variety of previously-intractable clustering methods.

X-ray fluorescence microscopy (XFM)

- Exposing metal atoms to a high-energy photon source causes electron excitation, leading to X-ray emission (Kramar 1999).
- Emitted X-rays are captured by a detector array, giving per-pixel metal concentration for multiple metals simultaneously.
- Resulting images are large, noisy, and complex; analysis is often laborious and biased.

Image processing and clustering in XFM

- XFM is highly sensitive, with images generally featuring high-frequency, short-range noise and extreme values.
- Image processing mitigates these by using native image properties, controlling data quality prior to analysis.
- Unsupervised clustering finds distinct regions of similarity in multi-layered image data using inherent properties.
- Clusters are "self-describing", reducing user input and bias, and highlighting regions of interest for targeted analysis.



Benchmarking single XFM image layers

Earlier work tested common image processing and clustering methods for segmenting single XFM image layers, finding that most clustering methods are impractical in XFM:

- Memory needs are large (>480 GB), runtimes are long (18-56+ hours), some methods require biased parameter tuning.
- K-means and bisecting k-means were the only successful clustering methods, and were fast and memory-efficient, but may be biased by prescribing cluster count.
- Clustering scores (Chicco et al. 2025) were most improved by local median filters and kernel convolutions.

Key problem: **XFM images are too large** for most clustering methods.

Can dataset size be reduced and still achieve reasonable clustering?

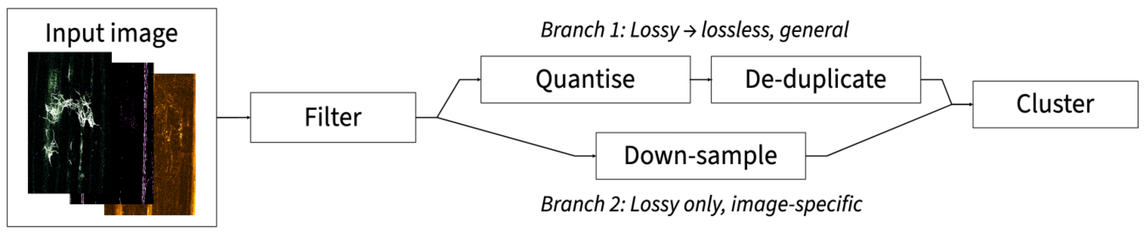
Clustering compressed XFM images

- Distance matrices needed by many clustering methods cannot be stored in main memory in most computers.
- Simple compression methods reduce the complexity and number of data, possibly allowing clustering in a compact dataset.
- Published works describe reducing *dimensions* (e.g. D'Angelo & Fop 2026), but what about reducing *data*? Existing publications do not address this directly.

Hypothesis:

Lossy compression → Lossless compression → Approximated clustering

- Lossy: irreversibly loses resolution but increases compression ratio. Lossless: perfectly reversible, lower compression without lossy step.
- Aim is to achieve similar clustering in compressed vs. full-scale data, reducing computational costs and enabling greater methods variety.
- Begin with simulated data, scale to XFM.



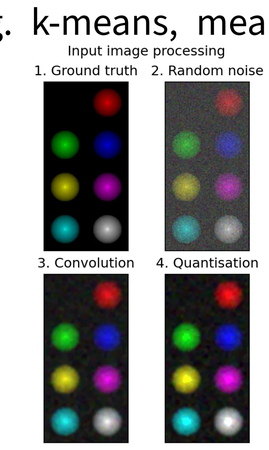
Compressed clustering: Early results

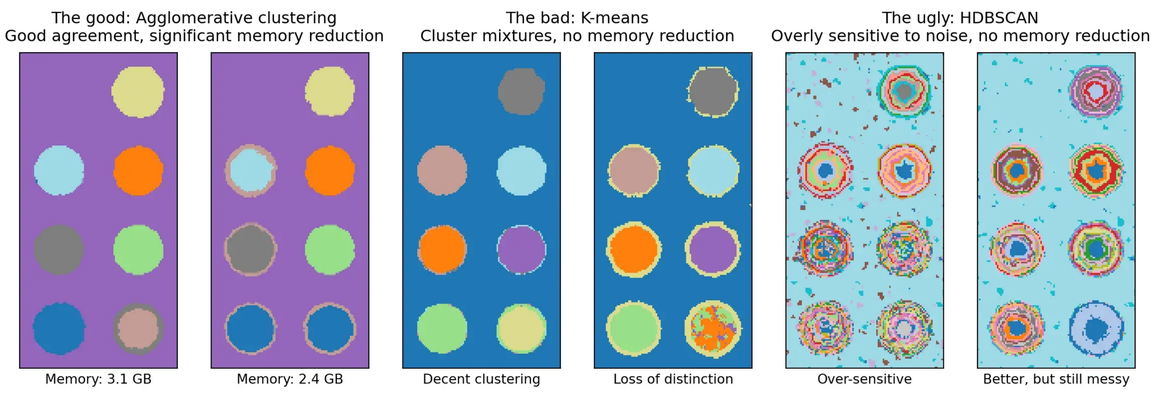
There seems to be some merit!

- Agglomerative clustering (previously impossible) shows similar results at greatly reduced cost.
- However, methods using volume or density (e.g. k-means, mean shift, DBSCAN) are much worse.

Next steps:

- Formally analyse these and other results.
- Scale to other image types including XFM.
- Introduce variations such as dimension reduction.





References

Chicco, D. et al. (2025). The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters. https://doi.org/10.7717/peerj-cs.3309

D'Angelo, S., & Fop, M. (2026). Data compression for fast dimension reduction and clustering of high-dimensional discrete data. https://doi.org/10.48550/arXiv.2606.10593

Kramar, U. (1999). X-Ray Fluorescence Spectrometers. In Encyclopedia of Spectroscopy and Spectrometry (2e). https://doi.org/10.1016/B978-0-12-374413-5.00314-6

I acknowledge the support of the Grains Research and Development Corporation (GRDC) provided through the Analytics for the Australian Grains Industry (AAGI) project.