Adv. Geosci., 67, 79–89, 2025
https://doi.org/10.5194/adgeo-67-79-2025

Advances in Geosciences
Open Access
EGU

Quantifying carbonate microstructure using classical segmentation pipelines for CCS and radioactive waste applications

Wurood S. Alwan[1,2,★], **Omar Choudhry**[3,★], **Paul Glover**[1], **Louey Tliba**[4], **and Richard Collier**[1]

[1]School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
[2]Ministry of Education, Baghdad, Iraq
[3]School of Computer Science, University of Leeds, Leeds, LS2 9JT, UK
[4]School of Chemical and Process Engineering, University of Leeds, Leeds, LS2 9JT, UK
★These authors contributed equally to this work.

Correspondence: Wurood S. Alwan (eewsa@leeds.ac.uk)

Received: 23 June 2025 – Revised: 19 September 2025 – Accepted: 22 October 2025 – Published: 3 December 2025

Abstract. Digital-rock analysis can now simulate pore-scale flow and geomechanics with high fidelity, yet routine application is still limited by the cost of converting images into pixel-accurate pore masks. Some porosity measurements measure total porosity, but flow requires connected porosity. The need to distinguish between isolated and connected pores is clearly important. The question is why it is difficult to identify connected and unconnected pores. To tackle this, we present a fully labelled 2-D scanning-electron-microscope (SEM) image dataset of outcrop carbonates designed to separate isolated intragranular pores from the connected fracture–pore network. This distinction controls seal performance in carbon-capture and storage (CCS) and radioactive-waste repositories. Each of the 29 056 × 22 952 px images were partitioned into 100 tiles (2048 × 2048 px), annotated by polygon tracing. We then employ and benchmark eight unsupervised computer-vision algorithms: Morphological Gradient, Distance Transform, Local Contrast, Watershed, Global Threshold, Gabor Texture, Refined Morphology, and Edge-based methods. We contribute (i) an open carbonate SEM dataset with labels separating connected and isolated pores, (ii) an efficient polygon-based labelling workflow, and (iii) a quantitative comparison of eight classical pipelines plus a Hybrid voting ensemble. These resources shorten the path from raw images to pore-network predictions, and provide a foundation for future learning-based methods.

1 Introduction

Digital rock physics has emerged as a powerful complement to laboratory core analysis, supplying pore-scale insight that underpins reservoir evaluation, geothermal production and subsurface energy storage (Blunt et al., 2013). By converting high-resolution images into binary maps of solid and void, researchers can simulate single- and multiphase flow, geomechanics and geochemistry with unprecedented fidelity. These capabilities are of growing importance to low-carbon technologies such as carbon capture and storage (CCS) and the geological disposal of radioactive waste, where long-term safety hinges on predicting how CO_2 or brine migrates through complex pore networks (Blunt et al., 2013). Geological CO_2 storage is highly sensitive to micro- and meso-scale heterogeneity because capillary- and gravity-dominated flow regimes amplify the influence of pore connectivity on plume migration (Ringrose et al., 2022).

A persistent bottleneck is segmentation: most workflows still rely on thresholding greyscale scanning-electron-microscope (SEM) or X-ray micro-computed-tomography images, followed by manual post-processing of the masks when pores are filament-thin or sparsely distributed (Sarkar et al., 2018). In these cases, edges lie near the noise level and small porosity channels or bridges of matrix that separate pores must be decided by hand, which increases annotation time and introduces operator bias. Openly available pore-scale datasets with pixel-accurate masks are scarce, limiting reproducible benchmarking (Da Wang et al., 2021). Recent CCS case studies further show that subtle layering and frac-

ture networks can redirect or immobilise dense-phase CO_2 over decadal timescales (Ringrose et al., 2022). To address this bottleneck, we introduce high-quality pixel-accurate labels and benchmark strong classical segmentation baselines for carbonate SEM images.

Recent studies apply convolutional neural networks to porous media and report strong results for delineating pore space in µCT/SEM images (Alqahtani et al., 2022; Bihani et al., 2022; Bangaru et al., 2022). These models can integrate broader context and better preserve thin structures, but they require pixel-accurate labels, topology-aware training, and non-trivial compute, and many published datasets are not class-balanced or publicly reusable. Artificial intelligence model training also takes exponentially more computational resources, which necessitates quicker and more efficient methods.

Classical computer vision segmentation techniques often uses e SEM/µCT (micro Computed Tomography), typically combining thresholding (Otsu, 1979; Sauvola and Pietikäinen, 2000), distance-transform/watersheds (Beucher and Meyer, 1992; Soille, 2003), edge detectors (Canny, 1986) and morphological operations (Soille, 2003). These pipelines are fast and explainable, but performance depends on a number of issues, such as illumination and noise, overlapping pore/matrix histograms, and the requirement for hand-tuned parameters. In addition, thresholding can be unstable under local shading (Sezgin and Sankur, 2004), watersheds can over-segment at weak boundaries unless seeds are carefully controlled (Beucher and Meyer, 1992), edge-based methods often miss pore interiors when boundaries are diffuse (Canny, 1986), and morphology can erode 1–2 px filaments unless parameters are tailored per image (Soille, 2003). This motivates releasing pixel-accurate labels that preserve thin isolated pores and a transparent benchmark across carbonate textures to document both strengths and failure modes of classical approaches. Investigating the optimal uses of these methods can help to inform better architectures in machine learning methods, especially important for very large images with high resolution. In three dimensions this requirement is even more important.

Connected porosity governs flow and thus permeability/relative-permeability upscaling, whereas isolated porosity mainly contributes to capillary trapping and (via wetted surface) adsorption, where the separation has direct consequences for CCS modelling. In practice the distinction is challenging because connectivity is scale- and physics-dependent (sub-pixel throats may allow diffusion but not viscous flow at the prevailing pressure gradient), 2D slices can miss 3D bridges, and thin, low-contrast links near grain contacts make edges ambiguous while intensity/texture overlap undermines simple thresholds. Our pixel-wise labels therefore preserve 1–2 px filaments and topology, and the benchmark explicitly reports the conditions under which classical pipelines succeed or fail.

This study aims to (i) publish an open, fully-labelled carbonate SEM data set, (ii) benchmark eight classical segmentation pipelines, and (iii) introduce a Hybrid voting ensemble. To address the identified gaps we present a high-quality, fully labelled 2-dimensional (2D) SEM image set of outcrop carbonates. The original mosaic ($29\,056 \times 22\,952$ px, 8-bit depth) was subdivided into 155 tiles of 2048×2048 px. Each carbonate grain was manually annotated and traced using hand-drawn polygon boundaries, with pixel classification defined as follows:

- Black pixels inside a polygon represent isolated, disconnected, intragranular pores.
- Black pixels outside all polygons represent the connected fracture–pore network.
- Grey pixels form the solid matrix and are masked out as background.

This three phase interpretation is collapsed into a two-class mask (isolated vs. connected porosity) for all analyses in this paper.

Recently there has been a large amount of investment and research in evaluating AI and generative AI in various domains. However, it is important to understand the baseline results of standard and extremely efficient models, such as older computer vision algorithms, to see where improvements can be made. Thus, we evaluate eight representative classical pipelines covering the main families used in computer vision algorithms: thresholding (global/adaptive), distance-transform/watershed, edge-based, morphological and texture filters. Together, they expose complementary strengths and failure modes directly relevant to CCS and radioactive-waste performance metrics, balancing the recovery of thin, isolated intragranular pores (relevant to capillary trapping and radionuclide immobilisation) with precisely mapping the connected network that governs potential migration pathways. Each method is particularly suited for specific use cases.

We have evaluated these algorithms noting their typical behaviour and performance in order to ascertain, which algorithm is best suited to which application. In future, further machine learning approaches can be built upon this more interpretable and explainable understanding.

The eight segmentation pipelines evaluated are:

1. Watershed (Beucher and Meyer, 1992) – a seed-and-grow algorithm on a distance map that excels at splitting touching pores when using markers to control over-segmentation.
2. Distance Transform Thresholding (Soille, 2003) – labels local maxima in the Euclidean distance field as candidate isolated pores (with non-max suppression).
3. Local Contrast (adaptive) thresholding (Sauvola and Pietikäinen, 2000) – per-pixel threshold $T = m + k\sigma$ (m:

local mean, σ: local SD, k: sensitivity). We used $k \in [-0.5, 0.2]$ in 0.1 steps.

4. Canny Edge-based Segmentation (Canny, 1986) – detects strong edges, which can be used to get regions by explicitly closing loops morphologically and fill those enclosed loops to find fully bounded pores.

5. Morphological Gradient (Soille, 2003) – flags any rims with a high grey-level gradient. Thresholding, closing and filling can turn these rims into pore masks. However, it will be prone to false positives at rough grain boundaries.

6. Gabor Texture Filtering (Jain and Farrokhnia, 1991) – multi-orientation Gabor filters emphasise oriented/periodic textures. With magnitude-threshold + cleanup they may capture elongated/fracture-like features more than round isolated pores.

7. Global Threshold (Otsu, 1979) – a single-threshold baseline tuned for high recall on connected pores.

8. Refined Morphology – our conservative baseline that keeps only round, 5–100 px^2 blobs with $\geq 80\,\%$ grey perimeter pixels (Serra, 1983).

Hybrid voting is an ensemble scheme that labels a pixel as "isolated" only when at least two of the three highest-performing filters (Distance Transform, Watershed and Local Contrast) concur, thereby combining their complementary strengths while suppressing individual false positives as discussed in (Isensee et al., 2021) and (Soille, 2003).

The contributions of this study are fourfold:

- Open data: release of the carbonate SEM tiles, polygons and masks, plus full acquisition metadata, under a CC-BY licence.
- Efficient annotation workflow: a polygon-based strategy that preserves pore-edge fidelity while avoiding pixel-wise tracing.
- Baseline model – a reproducible computer vision pipeline that discriminates isolated from connected pores with 93 % accuracy, providing a benchmark for future studies.
- Extension to 3-D – discussion of limitations (single lithology, single scanner) and a forward look to extending the method to µCT volumes and multi-phase mineral segmentation.

By coupling high-resolution imaging with machine learning, the study advances quantitative microstructural analysis for CCS, geothermal and radioactive-waste repositories, thereby supporting the transition to sustainable energy systems.

2 Dataset

The analysis was carried out on ultra-high resolution SEM images from a number of samples. The sample images are from an alluvial sand sheet or sabkha packstone, with grain-size codes ranging from upper-fine (UF) to medium (M) sand, from an industrial partner's field (exact location withheld per confidentiality agreement), but from a depth in the range 4346–4788 m. Its helium porosity at ambient stress lies between 0.016 and 0.02, and Klinkenberg-corrected nitrogen-gas permeability at ambient pressure (k_∞), lies between 0.032 and 0.087 mD. The Klinkenberg correction was carried out by linear extrapolation of apparent permeability (k_{app}) versus $1/P_{\mathrm{m}}$ (with P_{m} the mean gas pressure). k_∞ is reported as metadata only (not used in analyses; see Table 1).

Four back-scatter SEM mosaics were exported as $29\,056 \times 22\,952$ px, 8-bit greyscale, LZW-compressed TIFFs. Each mosaic has a calibrated pixel size of $0.195\,\mu\mathrm{m\,px}^{-1}$, yielding a field of view of $\sim 5.7\,\mathrm{mm} \times 4.5\,\mathrm{mm}$. All four mosaics were captured in the same session, ensuring consistent greyscale calibration. After partitioning the mosaics into 2048×2048 px tiles, we randomly retained 100 tiles, with 25 from each mosaic ($\approx 4.2 \times 108$ px) retained for analysis.

3 Data processing & label generation

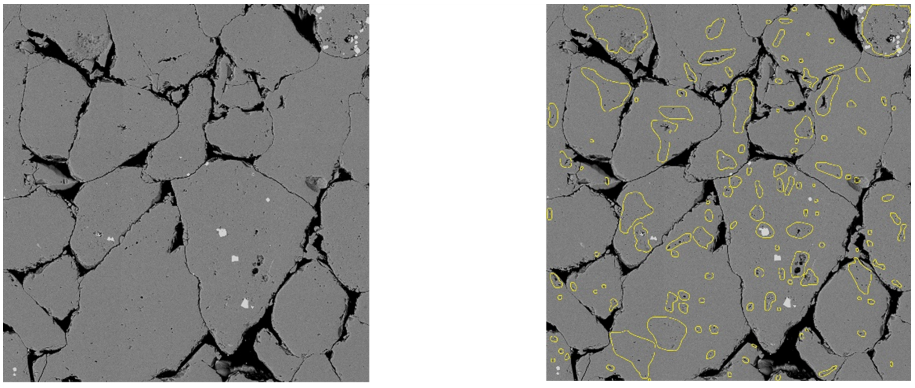
Each raw tile was first binarised with a global greyscale threshold to separate pore space from solid matrix (Fig. 1a). Carbonate grains were then outlined manually in Microsoft Paint using the free-hand polygon tool (Fig. 1b). Black pixels inside a polygon represent isolated (intragranular) pores, black pixels outside all polygons represent the connected fracture–pore network, and grey matrix pixels are ignored. The three-region map is collapsed to a three-class mask (0 for isolated, 1 for connected and 2 for mineral/background) and saved as an 8-bit PNG registered one-to-one with the raw tile. Annotation time per ring was $\sim$ 1–2 min, where some images with more rings took longer to annotate then those with fewer. A simple morphological clean-up filled stray matrix pixels inside wide fractures and removed speckle noise smaller than five pixels. This strategy allows us to highlight easily all the larger regions containing isolated pores, using an algorithm to process the pixels inside the annotated yellow rings. This benefitted us by not having to annotate each pixel for either class. All annotations were first rechecked after labelling all images, and again once using the software to create the isolated and connected classes to further validate the classes, where it was manually corrected if needed.

4 Methodology

This study focuses on unsupervised classical computer-vision pipelines to establish transparent, reproducible base-

Table 1. Specimen-level metadata for the four parent mosaics used in this study.

Sample ID	Depth (m)	Sedimentary facies	Grain size (µm)	Helium Porosity (fractional)	Klinkenberg-corrected N_2 permeability? (mD)
pdo1_7	4764.65	Alluvial sand sheet	UF	0.016	0.087
pdo1_12	4787.89	Alluvial sand sheet	LF	0.02	0.032
pdo2_24	4749.47	Sabka	UF	0.018	0.070
pdo8_21	4346.34	Alluvial sand sheet	M	–	–



a b

Figure 1. Workflow for label generation. **(a)** Raw greyscale tile, **(b)** same tile with yellow polygon annotation.

lines using the released pixel-accurate labels. We used the yellow bounding rings to determine whether pores were isolated or connected. We evaluated eight classical methods, ordered from most to least conservative as summarised in Table 2 and illustrated in Fig. 2. Table 3 provides a summary of pixel class distributions across the dataset the results of which will be described in detail in the next section.

5 Results and Discussion

Pixel-wise performance was assessed with the standard metrics defined below; the numerical values are reported in Table 4 and representative qualitative results are shown in Fig. 3.

- *Accuracy*: Out of all pixels, the percentage the method labelled correctly.
- *Precision (Isolated)*: When the method says, "this pixel is an isolated pore", this is how often it is right. High precision means there are very few false alarms.
- *Recall (Isolated)*: Of all truly isolated-pore pixels, what fraction did the method manage to find? High recall means almost nothing is missed.
- *F1 (Isolated)*: A single score that balances precision and recall, calculated as the harmonic mean of the precision and the recall [$F1 = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$]. High only when the method both finds most isolated pores and avoids false alarms.
- *IoU (Isolated)*: Intersection over Union. How much the predicted isolated-pore area overlaps the true isolated-pore area (0 = no overlap; 1 = perfect match).
- *IoU (Connected)*: The same overlap measure, but for connected-pore pixels.
- *Mean IoU*: The simple arithmetic mean IoU across both classes; a summary of overall region-matching quality.

Watershed delivers the best overall balance, missing only 26 isolated-pore pixels (high recall) while maintaining a moderate false-positive rate. The trade-off is that one in every two "isolated" pixels it proposes is actually connected (precision ~ 0.54). The Refined Morphology method is uncompromising by demanding strict roundness, size of connected pixels (5–100 px^2) and ≥ 80 % grey surrounding pixels, it achieves near-perfect precision (0.97) but sacrifices recall (0.15). This makes it ideal for quality-control checkpoints where only the most certain isolated pores matter.

Regarding computational details, the methods were tested on a M3 Macbook Pro with 16 GB of memory. As the methods are based on parameters and not through machine learning, there is no time taken to train any model, but it is needed for calibration. Inference time with these methods were typically between 356 to 1458 ms per 2048×2048 px image.

Table 2. Classical computer-vision pipelines ordered from most to least conservative.

Method	Description
Refined Morphology	This method applies strict rules about the shape and size of pores. Only those that are small, round, and clearly separated from surrounding features are kept. The effect is conservative: it highlights only the most certain intragranular pores and filters out ambiguous structures.
Gabor Texture	Here the focus is on texture patterns within the image. Gabor filters, which are sensitive to repeating visual structures in different directions, are applied to identify regions with a consistent and uniform appearance. Pores often present as smooth "blobs" against more varied backgrounds, allowing them to be distinguished by this texture-based approach.
Watershed	The watershed method treats the image as though it were a topographic surface, where darker regions resemble valleys and brighter regions resemble ridges. By "flooding" this landscape from seed points, the algorithm grows regions outward until they meet at boundaries. This process is particularly effective at dividing neighbouring pores that touch or overlap, producing clear separations.
Distance Transform	This approach calculates, for every pixel, how far it is from the nearest boundary. Points located at the maximum distance – i.e., at the deepest interior of a pore – are marked as candidate centres. This makes the method well suited to detecting isolated pores and characterising their spatial distribution.
Edge-based	Using edge-detection techniques such as the Canny operator, this method identifies sharp changes in brightness that correspond to boundaries in the image. When these edges form a complete closed loop, the enclosed region can be interpreted as a pore. It relies heavily on well-defined boundaries and is less effective if edges are faint or broken.
Morphological Gradient	This technique measures the difference in brightness across boundaries. Strong changes in intensity are emphasised, producing a rim-like outline around potential pores. Because it highlights even subtle boundaries, the method is highly sensitive, though this can also lead to false detections in noisy regions.
Local Contrast	Rather than considering the whole image at once, this method examines small neighbourhoods and enhances local differences between bright and dark regions. Pores that appear distinct compared to their immediate surroundings are highlighted. This is advantageous when global lighting is uneven, though it may also capture small artefacts.
Global Threshold	A single brightness threshold is applied across the entire image: pixels darker than the threshold are marked as pore, and brighter pixels as solid material. Afterwards, very small features (often noise) are removed. While simple and computationally efficient, this method assumes uniform lighting and contrast across the image, which may not always be the case.

Table 3. Pixel counts for each classical computer-vision pipeline, shown as true predictions/false predictions for connected and isolated pores.

Method	True Connected (higher is better)	False Connected (lower is better)	True Isolated (higher is better)	False Isolated (lower is better)
Morphological Gradient	164	10 550	0	951
Distance Transform	9409	1305	127	824
Local Contrast	8262	2452	714	237
Watershed	9921	793	26	925
Global Threshold	10 183	531	950	1
Gabor Texture	10 677	37	951	0
Refined Morphology	10 709	5	811	140
Edge-based	9515	1199	191	760

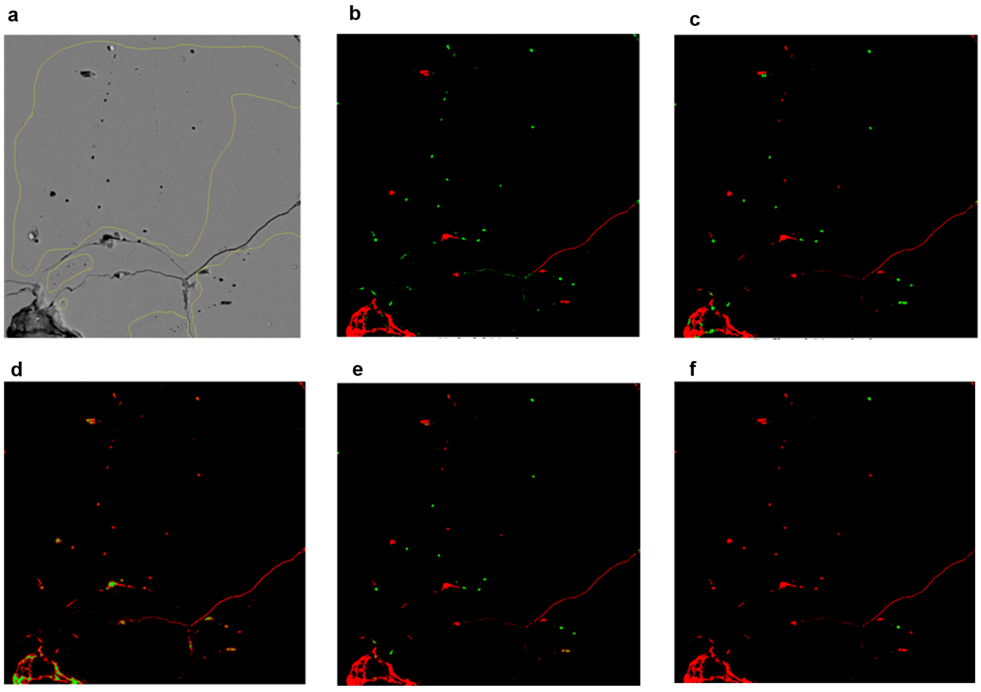


Figure 2. Classical computer-vision pipelines evaluated for isolated-pore segmentation on SEM tile pdo1-7 (top-left quadrant). Red = connected fracture–pore network; green = isolated /disconnected pores. **(a)** Raw greyscale image with hand-traced grain rings. For each automated output the first value is the number of pixels predicted as isolated pores and the percentage in parentheses is the isolated-pore recall (fraction of true isolated-pore pixels that were recovered): **(b)** Distance Transform: 2129 isolated-pore pixels, 18.3 % recall; **(c)** Watershed: 1718, 14.7 %; **(d)** Local-Contrast: 2787, 23.9 %; **(e)** Hybrid voting (two-out-of-three majority of Distance Transform, Watershed and Local Contrast masks): 951, 8.2 %; **(f)** Refined Morphology: 166, 1.4 %.

Table 4. Pixel-wise segmentation performance of the eight classical computer-vision pipelines evaluated in this study.

Metric	Watershed	Refined Morphology	Distance Transform	Edge-Based	Local Contrast	Morph. Gradient	Global Threshold	Gabor Texture
Accuracy	**0.93**	**0.93**	0.88	0.88	0.73	0.10	0.87	0.92
Precision (Isolated)	0.54	**0.97**	0.39	0.39	0.09	0.08	0.00	0.00
Recall (Isolated)	0.97	0.15	0.87	0.80	0.25	**1.00**	0.00	0.00
$F1$ (Isolated)	**0.69**	0.26	0.54	0.52	0.13	0.15	0.00	0.00
IoU (Isolated)	**0.53**	0.14	0.37	0.35	0.07	0.08	0.00	0.00
IoU (Connected)	0.92	**0.93**	0.87	0.87	0.53	0.02	0.07	0.92
Mean IoU	**0.73**	0.54	0.62	0.61	0.30	0.05	0.04	0.46

Notes: Metrics are averaged over the 20-tile test set; boldface marks the best score in each row. $F1$ is the harmonic mean of precision and recall for the isolated-pore class; IoU denotes intersection-over-union. A value of 0.00 or similarly any very small value, indicates that the method could not effectively detect the pore of that class due to being too specific; in this case precision, recall, $F1$, and IoU for that class are reported as 0 by convention.

Parallel computation by splitting each image into 9 tiles of a 3 by 3 grid improved performance to 122 to 714 ms per image. However, as these are static methods, they are not generalisable to other datasets without further calibration.

We report results for eight classical segmentation pipelines and a simple Hybrid voting ensemble on the labelled carbonate SEM dataset. Our comparative analysis demonstrates a clear trade-off between recall and precision, a common challenge in image segmentation. The Watershed algorithm emerges as the most balanced method, achieving the highest mean IoU (0.73) and an exceptional recall for isolated pores (0.97). This indicates that it successfully identifies the vast majority of true isolated pores. Its primary drawback is a modest precision (0.54), meaning it misclassifies a significant number of connected-pore pixels as isolated. For applications in CCS and radioactive waste storage, this behaviour could lead to an overestimation of the rock's capacity for capillary trapping or radionuclide immobilisation. Nonetheless, for a comprehensive first-pass characterisation of a sample,

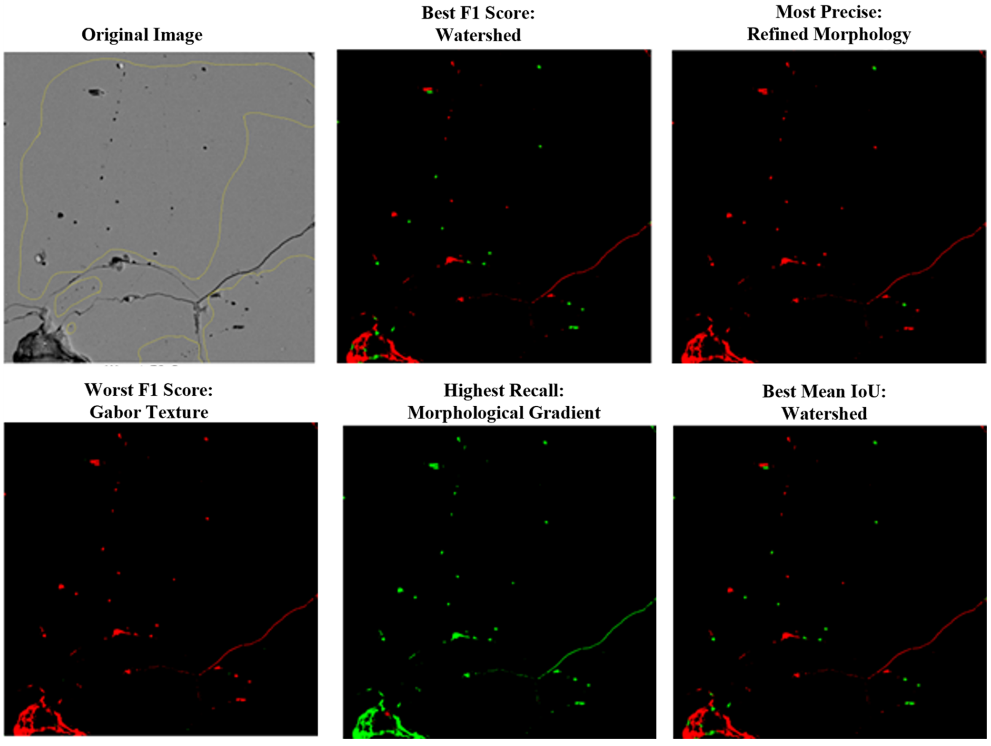


Figure 3. Visual examples of the best- and worst-performing pipelines on a representative tile.

Watershed provides the most complete picture of the potential isolated pore system.

In stark contrast, the Refined Morphology method is defined by its conservatism. It achieves near-perfect precision (0.97) at the cost of very low recall (0.15). This pipeline is engineered to only identify pixels that are unequivocally part of an isolated, intragranular pore. While it misses the majority of such pores, its classifications are highly reliable. This makes it an ideal tool for specific quality-control tasks or for generating a high-confidence "minimum case" estimate of isolated porosity, which is critical for safety case assessments where false positives are unacceptable. The performance of the other methods, such as Distance Transform and Edge-based segmentation, falls between these two extremes, but none offer a more advantageous balance of metrics than Watershed. The poor performance of methods like Morphological Gradient and Global Thresholding highlights the difficulty of relying on simple intensity or gradient features in these complex carbonate images, where pore boundaries can be subtle and indistinct. Global Threshold and Gabor Texture perform poorly for isolated pores (Precision = Recall = 0.00) because they produce no isolated predictions on our carbonate SEM tiles. For Global Threshold, local normalisation suppresses small dark intragranular pores; for Gabor, the filter bank responds to edges rather than interiors, and our conservative post-processing removes the resulting small islands. These two methods are therefore not recommended when isolated-pore quantification is required. By contrast, the Watershed approach is preferable when completeness is the priority (high Recall; good at splitting touching grains), while the Distance Transform is useful when thin filaments must be recovered (with light pre-filtering to limit noise). The Edge-based method is best where boundaries are sharp and continuous, and the Refined Morphology pipeline should be chosen when false positives must be minimised (near-perfect precision at the expense of Recall). These benchmarks provide a transparent reference and document failure modes that can inform future learning-based baselines.

This initial study effectively validates our polygon-based annotation workflow and establishes a robust performance baseline using traditional, explainable methods. It underscores that even without deep learning, significant microstructural insights can be extracted. More importantly, it quantifies the specific challenges, such as distinguishing finely connected fracture networks from truly isolated pores, that a future deep-learning model must overcome.

6 CCUS Applications of Machine Learning-based Pore Segmentation

The injection of CO_2 into underground reservoirs for carbon capture and underground storage (CCUS) is very different from the well-known process of extracting gas (Ringrose et al., 2022). There are many reasons for this, of which three of

the most important are (i) that supercritical CO_2 injection is highly sensitive to heterogeneity and anisotropy (Reynolds and Krevor, 2015; Trevisan et al., 2017), (ii) the CO_2 flow regime varies from the well, to the near-wellbore space, then more distally (Rui et al., 2025), and (iii) variations in rock microstructure at small scales control the overall fluid flow (Al-Zainaldin et al., 2017). All three of these issues are controlled by the amount of void space in the rock (porosity) and the size and connectedness of the void space; in other words, the microstructure of the rock. Therefore, general pore-segmentation without taking into account the different kinds of pores is important for all aforementioned applications.

The pixel-wise masks produced here separate connected from isolated pores. From these masks one can derive simple inputs for modelling: (i) the connected-pore fraction (share of pixels labelled connected) to constrain permeability/relative-permeability upscaling by excluding non-flowing porosity; (ii) the isolated-pore fraction to bound capillary-trapped volumes and residual saturations; and (iii) where surface-area information is available (e.g., from 3-D μCT or stereological estimates), the connected wetted surface area to parameterise adsorption-capacity models (e.g., Langmuir-type).

Heterogeneity in the lithology and particularly in microstructure and its influence on transport processes has a strong effect on the capillary- and gravity-dominated flow processes which are typical of CO_2 storage projects (Reynolds and Krevor, 2015; Trevisan et al., 2017). A good example of this is the effect of the complex layering of the Sleipner field, which dominated the shape and development of the CO_2 plume, which represents the effects of strong vertical heterogeneity (Furre et al., 2017, 2020). Flow-in gas and oil production systems is much less affected by rock heterogeneity because fluid transport occurs in the viscous-dominated flow regime (Bentley and Ringrose, 2021).

Consequently, CCUS processes can only be managed if the CCUS geological modelling and simulation takes account of the rock microstructure, and that in turn requires us to be able to segment any image or measurement of a rock into its different phases as that small scale.

This paper has attempted to use computer vision algorithms to automate and accelerate the ability to segment the rock into various mineral phases, into connected and unconnected pore space. Future work would involve using artificial intelligence and machine learning approaches, which have previously been used in CCUS. One approach used machine learning techniques to help upscale core data so that it could be used in a conventional large-scale reservoir model and simulation campaign (Yu et al., 2022). Unfortunately, such an approach has the outcome of averaging the small-scale heterogeneities which control the supercritical CO_2 flow.

A combination of CT-scan images and various machine learning approaches, including convolutional neural networks (CNNs), was proposed by (Panaitescu et al., 2024). They recognised the importance of understanding how multi-scale, image-derived data can enhance understanding of porous features can be unlocked with the use of machine learning, but did not specifically focus their work on CCUS reservoirs, targeting instead marginal oil and gas reservoirs. Building on their earlier work (Panaitescu et al., 2024), they do not attempt to segment connected and unconnected porosity, as in our case, but porosity from fractures. They noted that simple histogram thresholding is not sufficient in accurately perform semantic even if threshold optimisation techniques (Otsu, 1975) or Gaussian Mixture Models (Huang and Chau, 2008) are used, because simple differences in grey scales does not provide enough information to separate phase.

Consequently, they used CNNs, especially using U-net-inspired architectures (Ronneberger et al., 2015), providing results that are extremely useful for application to CCUS reservoirs in coal (Karimpouli et al., 2020) and other sequestration reservoirs (Kim et al., 2020; Reinhardt et al., 2022; Pham et al., 2023).

Elsewhere in CCUS science and engineering, deep learning is also making an impact. These applications include simulation (Chu et al., 2022), process engineering (Sabeena, 2023) and optimising carbon storage (Wang et al., 2025).

7 Limitations and Future Work

The findings and methodologies presented in this paper provide a strong foundation for microstructural analysis, but they are subject to several limitations that also define the direction of our future research.

7.1 Limitations

- 2D Analysis: The use of 2D SEM images provides a cross-sectional view that may not fully capture the complex, three-dimensional nature of pore connectivity. Pores that appear isolated in a 2D slice may be connected in the third dimension.
- Single Lithology: Our analysis was performed exclusively on outcrop carbonate samples from a single geological setting. The performance of the segmentation methods may vary significantly with other rock types, such as sandstones or shales, or with carbonates possessing different diagenetic histories.
- Classical Methods: As detailed, this study focused on classical computer-vision algorithms. While insightful, these methods lack the learning capacity and potential for higher accuracy and generalisation offered by machine learning algorithms.
- Manual Annotation Effort: The label generation, though efficient, still requires significant manual input, making it a bottleneck for analysing much larger datasets.

7.2 Future Work

The following future work is based on the strengths, contributions and limitations of this paper:

1. Machine Learning: Using the insights from the classical method benchmark, we will train machine learing models to exceed the performance of all computer vision algorithms.

2. Extension to 3D: This segmentation workflow should be extended to 3D X-ray μCT volumes of the same or similar rock types. This will enable a true 3D quantification of pore network topology and provide more realistic inputs for flow simulation models.

3. Generalisation and Robustness Testing: Further machine learning models should be tested on a wider range of lithologies and on images acquired from different imaging systems to assess and improve its generalisability.

4. Multi-Class Segmentation: We will expand the model's capabilities beyond a binary pore classification to a multi-class segmentation problem that also identifies and quantifies different mineral phases within the solid matrix.

8 Conclusions

This study addressed a critical gap in digital rock physics: the quantitative segmentation of isolated versus connected porosity in carbonate reservoir analogues. Our primary contributions are:

1. An Efficient Annotation Workflow: We detailed a polygon-based method that preserves pore-edge fidelity more effectively than pixel-wise tracing, reducing annotation time while maintaining high quality.

2. A Classical Method Benchmark: We systematically evaluated eight classical computer-vision pipelines, establishing that the Watershed algorithm provides the best overall balance of recall and precision for this task. This analysis serves as a foundational benchmark for future work.

3. The Watershed algorithm is the strongest single classical method with the best overall performance for isolated pores (precision/recall/$F1 = 0.54/0.97/0.69$) and highest mean IoU (Table 4). Whilst the Hybrid voting method improved upon the mean across the eight single-method baselines of the isolated-pore $F1$ scores from 0.29 to 0.62 for pore classification and segmentation accuracy, it still performed poorer than the Watershed algorithm.

By successfully distinguishing between the connected pore network and the isolated intragranular pores, this work enhances our ability to classify formations (Glover et al., 2022), and to quantify key microstructural properties that govern fluid flow, capillary trapping, and solute transport, ultimately allowing useful properties such as permeability to be predicted accurately (Al Khalifah et al., 2020; Rashid et al., 2015). These findings have direct relevance for improving the security and efficiency of geological carbon storage and radioactive waste disposal.

Code availability. It is not possible to share a copy of the code used in this work at this time because the code is currently still under development and is not in a state that it can be released publicly. We hope to make the code available with our forthcomming paper on AI processes.

Data availability. Original image data used in this work belongs to a third party and cannot be released due to contractual restrictions. The labelled data created in this paper could, in principle, be published. However, it forms the input to current AI procedure research and will be published in the paper concerning that research.

Author contributions. Wurood S. Alwan: Methodology, Investigation, Data curation, Writing – original draft. Omar Choudhry: Methodology, Software, Formal analysis, Visualisation, Writing – original draft, Writing – review & editing. Paul Glover: Conceptualisation, Methodology, Project administration, Supervision, Writing – review & editing. Louey Tliba: Methodology, Writing – review & editing, investigation. Richard Collier: Supervision.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Special issue statement. This article is part of the special issue "European Geosciences Union General Assembly 2025, EGU Division Energy, Resources & Environment (ERE)". It is a result of the EGU General Assembly 2025, Vienna, Austria & Online, 27 April–2 May 2025.

Acknowledgements. The authors would like to thank the reviewers for their constructive comments which have much improved the pa-

per. We would also like to thank Prof. Quentin Fisher (University of Leeds) for allowing the use of the original image data.

Review statement. This paper was edited by Michael Kühn and reviewed by Roberto Emanuele Rizzo and one anonymous referee.

References

Al Khalifah, H., Glover, P., and Lorinczi, P.: Permeability prediction and diagenesis in tight carbonates using machine learning techniques, Mar. Petrol. Geol., 112, 104096, https://doi.org/10.1016/j.marpetgeo.2019.104096, 2020.

Alqahtani, N. J., Niu, Y., Wang, Y. D., Chung, T., Lanetc, Z., Zhuravljov, A., Armstrong, R. T., and Mostaghimi, P.: Super-resolved segmentation of X-ray images of carbonate rocks using deep learning, Transp. Porous Media, 143, 497–525, https://doi.org/10.1007/s11242-022-01781-9, 2022.

Al-Zainaldin, S., Glover, P. W., and Lorinczi, P.: Synthetic fractal modelling of heterogeneous and anisotropic reservoirs for use in simulation studies: implications on their hydrocarbon recovery prediction, Transp. Porous Media, 116, 181–212, https://doi.org/10.1007/s11242-016-0770-3, 2017.

Bangaru, S. S., Wang, C., Zhou, X., and Hassan, M.: Scanning electron microscopy (SEM) image segmentation for microstructure analysis of concrete using U-net convolutional neural network, Autom. Constr., 144, 104602, https://doi.org/10.1016/j.autcon.2022.104602, 2022.

Bentley, M. and Ringrose, P.: The rock model, in: Reservoir Model Design: A Practitioner's Guide, 2nd edn., Springer, Cham, 11–63, https://doi.org/10.1007/978-3-030-70163-5, 2021.

Beucher, S. and Meyer, F.: The morphological approach to segmentation: the watershed transformation, in: Mathematical Morphology in Image Processing, edited by: Dougherty, E. R., CRC Press, 433–481, ISBN 978-0-8247-8724-0, 1992

Bihani, A., Daigle, H., Santos, J. E., Landry, C., Prodanović, M., and Milliken, K.: MudrockNet: Semantic segmentation of mudrock SEM images through deep learning, Comput. Geosci., 158, 104952, https://doi.org/10.1016/j.cageo.2021.104952, 2022.

Blunt, M. J., Bijeljic, B., Dong, H., Gharbi, O., Iglauer, S., Mostaghimi, P., Paluszny, A., and Pentland, C.: Pore-scale imaging and modelling, Adv. Water Resour., 51, 197–216, https://doi.org/10.1016/j.advwatres.2012.03.003, 2013.

Canny, J.: A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 679–698, https://doi.org/10.1109/TPAMI.1986.4767851, 1986.

Chu, A. K., Benson, S. M., and Wen, G.: Deep-learning-based flow prediction for CO_2 storage in shale–sandstone formations, Energies, 16, 246, https://doi.org/10.3390/en16010246, 2022.

Da Wang, Y., Blunt, M. J., Armstrong, R. T., and Mostaghimi, P.: Deep learning in pore scale imaging and modeling, Earth-Sci. Rev. 215, 103555, https://doi.org/10.1016/j.earscirev.2021.103555, 2021.

Furre, A.-K., Eiken, O., Alnes, H., Vevatne, J. N., and Kiær, A. F.: 20 years of monitoring CO_2-injection at Sleipner, Energy Procedia, 114, 3916–3926, https://doi.org/10.1016/j.egypro.2017.03.1523, 2017.

Furre, A.-K., Meneguolo, R., Pinturier, L., and Bakke, K.: Planning deep subsurface CO_2 storage monitoring for the Norwegian full-scale CCS project, First Break, 38, 55–60, https://doi.org/10.3997/1365-2397.fb2020074, 2020.

Glover, P. W., Mohammed-Sajed, O. K., Akyüz, C., Lorinczi, P., and Collier, R.: Clustering of facies in tight carbonates using machine learning, Mar. Petrol. Geol., 144, 105828, https://doi.org/10.1016/j.marpetgeo.2022.105828, 2022.

Huang, Z.-K. and Chau, K.-W.: A new image thresholding method based on Gaussian mixture model, Appl. Math. Comput., 205, 899–907, https://doi.org/10.1016/j.amc.2008.05.130, 2008.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nat. Methods, 18, 203–211, https://doi.org/10.1038/s41592-020-01008-z, 2021.

Jain, A. K. and Farrokhnia, F.: Unsupervised texture segmentation using Gabor filters, Pattern Recognit., 24, 1167–1186, https://doi.org/10.1016/0031-3203(91)90143-S, 1991.

Karimpouli, S., Tahmasebi, P., and Saenger, E. H.: Coal cleat/fracture segmentation using convolutional neural networks, Nat. Resour. Res., 29, 1675–1685, https://doi.org/10.1007/s11053-019-09536-y, 2020.

Kim, Y., Ha, S. J., and Sup Yun, T.: Deep learning for extracting micro-fracture: Pixel-level detection by convolutional neural network, E3S Web Conf., 205, 03007, https://doi.org/10.1051/e3sconf/202020503007, 2020.

Otsu, N.: A threshold selection method from gray-level histograms, Automatica, 11, 23–27, 1975.

Otsu, N.: A threshold selection method from gray-level histograms, IEEE Trans. Syst., Man, Cybern., 9, 62–66, https://doi.org/10.1109/TSMC.1979.4310076, 1979.

Panaitescu, C., Wu, K., Kartal, M., Tanino, Y., Starkey, A., Qin, G., Zhao, L., Cao, Z., and Wu, S.: Exploring morth sea fractured sandstone properties: artificial intelligence, multiscale imaging, pore-fracture network analysis and experimental results, in: SPE Europe Energy Conference and Exhibition, Turin, Italy, 10–13 June 2024, paper SPE-220049-MS, https://doi.org/10.2118/220049-MS, 2024.

Pham, C., Zhuang, L., Yeom, S., and Shin, H.-S.: Automatic fracture characterization in CT images of rocks using an ensemble deep learning approach, Int. J. Rock Mech. Min. Sci., 170, 105531, https://doi.org/10.1016/j.ijrmms.2023.105531, 2023.

Rashid, F., Glover, P., Lorinczi, P., Hussein, D., Collier, R., and Lawrence, J.: Permeability prediction in tight carbonate rocks using capillary pressure measurements, Mar. Petrol. Geol., 68, 536–550, https://doi.org/10.1016/j.marpetgeo.2015.10.005, 2015.

Reinhardt, M., Jacob, A., Sadeghnejad, S., Cappuccio, F., Arnold, P., Frank, S., Enzmann, F., and Kersten, M.: Benchmarking conventional and machine learning segmentation techniques for digital rock physics analysis of fractured rocks, Environ. Earth Sci., 81, 71, https://doi.org/10.1007/s12665-021-10133-7, 2022.

Reynolds, C. and Krevor, S.: Characterizing flow behavior for gas injection: Relative permeability of CO_2-brine and N2-water in heterogeneous rocks, Water Resour. Res., 51, 9464–9489, https://doi.org/10.1002/2015WR018046, 2015.

Ringrose, P., Andrews, J., Zweigel, P., Furre, A.-K., Hern, B., and Nazarian, B.: Why CCS is not like reverse gas engi-

neering, First Break, 40, 85–91, https://doi.org/10.3997/1365-2397.fb2022088, 2022.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, Medical image computing and computer-assisted intervention–MICCAI 2015, edited by: Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., Lecture Notes in Computer Science, 9351, Springer, Cham, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Rui, Z., Zeng, L., and Dindoruk, B.: Challenges in the large-scale deployment of CCUS, Engineering, 44, 17–20, https://doi.org/10.1016/j.eng.2024.11.031, 2025.

Sabeena, J.: Enhancing Carbon Capture, Utilization, and Storage (CCUS) Through AI-Enabled CNN and Bayesian Networks, in: 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 15–17 November 2023, 1547–1551, https://doi.org/10.1109/ICSCNA58489.2023.10370286, 2023.

Sarkar, P., Kumar, A., Singh, K. H., Ghosh, R., and Singh, T. N.: Pore system, microstructure and porosity characterization of Gondwana shale of Eastern India using laboratory experiment and watershed image segmentation algorithm, Mar. Petrol. Geol., 94, 246–260, https://doi.org/10.1016/j.marpetgeo.2018.04.006, 2018.

Sauvola, J. and Pietikäinen, M.: Adaptive document image binarization, Pattern Recognit., 33, 225–236, https://doi.org/10.1016/S0031-3203(99)00055-2, 2000.

Serra, J.: Image Analysis and Mathematical Morphology, Academic Press, Inc., United States, 610 pp., ISBN 0126372403,1983.

Sezgin, M. and Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation, J. Electron. Imaging, 13, 146–165, https://doi.org/10.1117/1.1631315, 2004.

Soille, P.: On the morphological processing of objects with varying local contrast, in: Discrete Geometry for Computer Imagery (DGCI 2003), Naples, Italy, 19–21 November 2003, 52–61, https://doi.org/10.1007/978-3-540-39966-7_4, 2003.

Trevisan, L., Pini, R., Cihan, A., Birkholzer, J. T., Zhou, Q., González-Nicolás, A., and Illangasekare, T. H.: Imaging and quantification of spreading and trapping of carbon dioxide in saline aquifers using meter-scale laboratory experiments, Water Resour. Res., 53, 485–502, https://doi.org/10.1002/2016WR019749, 2017.

Wang, Y., Thanh, H. V., Al-Mudhafar, W. J., Dai, Z., Hemeng, Z., Davoodi, S., and Zhang, T.: Data Driven Based Deep Learning for Optimizing Carbon Storage and Methane Adsorption in Unconventional Shale Gas Reservoirs, Journal of Environmental Chemical Engineering, 13, 116901, https://doi.org/10.1016/j.jece.2025.116901, 2025.

Yu, X., Butler, S. K., Kong, L., Mibeck, B. A., Barajas-Olalde, C., Burton-Kelly, M. E., and Azzolina, N. A.: Machine learning-assisted upscaling analysis of reservoir rock core properties based on micro-computed tomography imagery, J. Petrol. Sci. Eng., 219, 111087, https://doi.org/10.1016/j.petrol.2022.111087, 2022.