Practical Advances in Artificial Intelligence and Computing
Frontier Research Publishing • https://www.frpubs.org
Vol. 5, No. 7, 2026

Volumetric Pore Measurements for Learning in Metal Additive Manufacturing

Thavriel Mornholt[1]*, Elyndra Sablemere[1], Osverin Cairnwick[1]
[1]University of Wollongong, Wollongong, Australia
*Corresponding author: thavrielmornholt786@gmail.com

Abstract: Learning for pore inspection should be tied to a defined three-dimensional engineering quantity. This article develops a volumetric measurement framework using the published extraction example in NIST AMS 100-73 from 2026. Deterministic constructed volumes isolate connected components, voxelization, boundary displacement, and morphological merging. Exact geometric calculations show why small pores and surface-derived descriptors can be sensitive to processing choices. The proposed study design holds out complete builds, preserves voxel calibration and partial-object selection, and evaluates descriptor and endpoint errors alongside segmentation. Generated volumes retain synthetic provenance. No new XCT specimen, material-property estimate, or trained inspection score is claimed.

Keywords: additive manufacturing; XCT; pore metrology; voxelization; connected components

1 The Measurand Is a Three Dimensional Defect

A dark region in an X-ray computed tomography slice is an observation shaped by attenuation, reconstruction, resolution, and segmentation. It is not yet a complete measurement of a material defect. For an additively manufactured metal, the relevant quantity may be total porosity, the largest connected pore, distance to a surface, or the shape and orientation of a lack-of-fusion defect. These quantities respond differently to image processing. This paper derives a measurement framework for learning from pore volumes, with a published NIST example and reproducible geometric calculations.

NIST's 2026 report on generative modeling of lack-of-fusion defects provides the source case [1]. Its published extraction figure shows segmented XCT information, intensity inversion, removal of boundary-connected regions, and isolated three-dimensional pores. Another 2026 study examines automated porosity estimation from CT imagery in cement-based materials [2]. The latter concerns a different material system and is used here as a comparison of measurement structure, not as evidence of performance on metal pores. The present contribution is the explicit link between a learning representation and the quantity an engineering analysis intends to measure.

The broader process-structure-property literature explains why additive manufacturing parameters and resulting defects must be considered together [3]. XCT reviews describe the value and limitations of volumetric inspection [4]. Work linking pyrometry to porosity illustrates the additional inference needed to connect an in-process signal to a later internal measurement [5]. In-situ learning approaches address another part of that chain [6]. The distinction between process observation and defect reference should remain visible. A model predicting porosity from process signals is not directly observing each pore at the time of prediction.

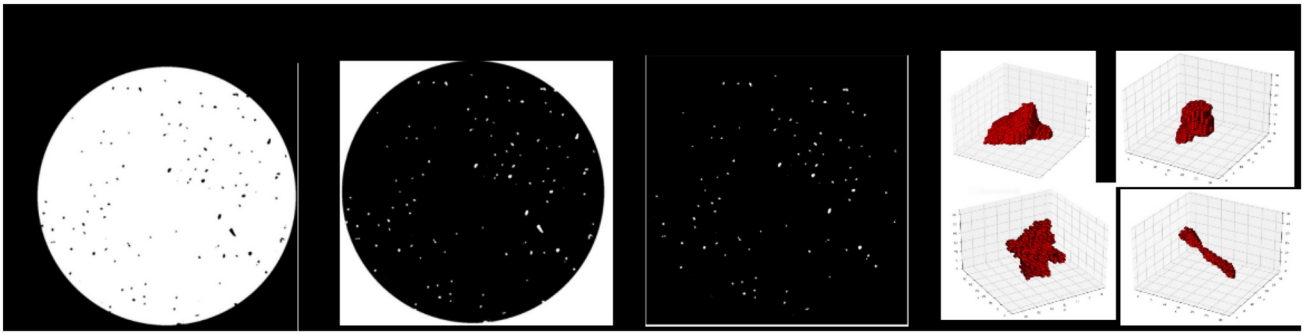


Figure 1. Published pore-extraction display from Hoque and Kim [1], NIST AMS 100-73, Figure 1. Source image extracted unchanged from the report: segmented XCT slice, inversion, boundary clearing, and isolated volumes. It is a source result, not newly processed specimen data.

Figure 1 reproduces the NIST extraction display. Figures 2 through 4 use constructed geometry to isolate connectivity, voxelization, boundary displacement, and morphology. They do not represent newly scanned specimens or measured mechanical properties. No synthetic pore is relabeled as a real XCT observation. The calculations are designed to expose how a seemingly minor image-processing choice can change an engineering quantity before any learning model is fitted.

2 Dimensional Meaning before Algorithm Choice

Let a binary pore mask occupy a region of interest with voxel dimensions delta x, delta y, and delta z. Equation (1) gives the segmented pore volume. It is a geometrical quantity conditional on the mask and physical calibration. If the voxels are anisotropic, replacing the product by a single isotropic spacing changes the volume. If the mask is generated from a cropped or resized image, the transformation must preserve the corresponding physical support. Numerical array dimensions alone do not establish physical dimensions.

$$V_{pore} = N_{pore} \Delta x \, \Delta y \, \Delta z \tag{1}$$

Porosity is a ratio to a specified material region. Equation (2) makes that denominator explicit. A cropped volume can exclude an external surface, include reconstruction background, or remove boundary-connected pores. Each choice can change the result. A paper should state whether it measures the inspected region, the entire specimen, or a selected interior subvolume. Comparing porosity values with different denominators can create an apparent model discrepancy that is actually a difference in the target definition.

$$\varphi = V_{pore} / V_{ROI} \tag{2}$$

Connected components require a neighborhood convention. Face-connected voxels, edge-connected voxels, and corner-connected voxels define different digital topologies. A one-voxel bridge can combine two regions into one defect under one convention. A single slice can show disconnected areas that belong to one connected three-dimensional pore. The analysis should therefore report the three-dimensional connectivity rule and avoid treating two-dimensional components as complete volumetric objects.

Table 1 distinguishes several pore descriptors and their evidential requirements. Total porosity can be stable while the largest component changes sharply after a bridge is added. A shape parameter can be sensitive to surface estimation even when volume is nearly unchanged. The most appropriate descriptor depends on the engineering question. A model should not be credited with improving defect characterization solely because one convenient image metric improves.

Table 1. Observation quantities and their reference requirements.

Quantity or input	Required evidence	Interpretation
Pore volume	Mask and physical voxel dimensions	Boundary errors disproportionately affect small objects
Porosity	Pore volume and a defined inspected denominator	Different regions of interest measure different targets
Connected pore	Declared three-dimensional neighborhood	A narrow bridge can change component count
Surface and shape	Surface estimator and physical calibration	Smoothing can change the descriptor being measured

3 A Published Extraction Example

The NIST source figure is an instructive sequence because it separates inversion, boundary handling, and component extraction. The original display is retained as a published result of that report. The current paper does not repeat its segmentation or infer new pore statistics from the figure. The relevant lesson is procedural: each transformation changes what counts as an included object and should be recorded with the output. Boundary clearing, for example, is a sampling decision as well as an image operation.

A pore intersecting the inspected boundary is only partially observed. Removing it can produce a cleaner set of complete shapes, useful for a generative shape model. That retained set does not automatically represent the full specimen's defect population. Larger pores can be more likely to intersect a boundary and therefore be excluded. A downstream statistical or mechanical interpretation should account for the selection. The purpose of the dataset should determine whether partial objects are retained, flagged, or excluded.

Image-processing software such as scikit-image makes many of these operations available and reproducible [7]. The software citation should be accompanied by exact functions, parameters, versions, and neighborhood settings where they affect the result. A general algorithm name is insufficient. Two implementations can differ in padding, surface extraction, interpolation, or handling of tied thresholds. These differences can matter when a pore is only a few voxels across.

Generative models introduce a second provenance layer [8]. A synthetic pore can be valuable for testing inspection or training a classifier, but it is not independent empirical evidence that the manufacturing process produced that shape. Its realism should be assessed against held-out measured objects and relevant descriptors. A source-derived shape distribution can be reproduced while important connectivity or surface features remain wrong. Evaluation should therefore follow the intended use of the generated volume rather than relying only on visual resemblance.

4 Exact Geometric Diagnostics

Figure 2 constructs a connected volume from two overlapping spherical regions and shows a central slice. The example has a known construction and no assigned material identity. It demonstrates why a slice and a volume answer different questions. The slice can reveal local cross-sections, while the volumetric representation reveals the connected object. A learning system that consumes slices independently may still

estimate a useful aggregate, but it should be tested for the topology and shape information lost by that representation.

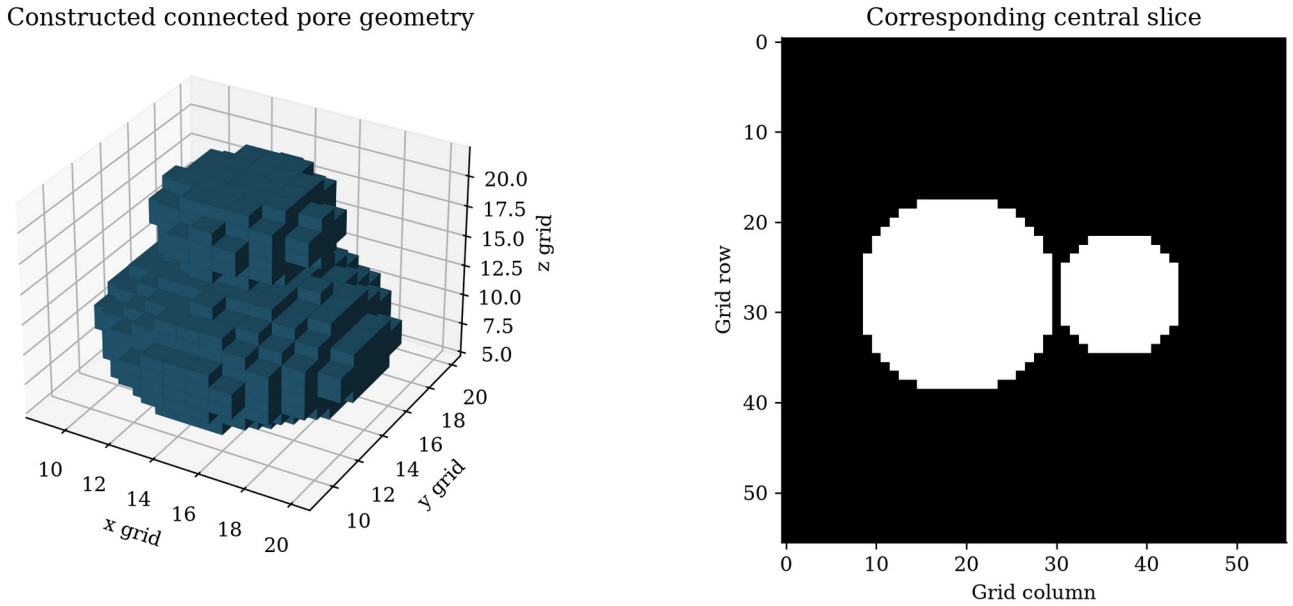


Figure 2. Constructed connected three-dimensional pore geometry and its central slice. Two overlapping spherical regions define the volume. The geometry has no assigned alloy, scan provenance, or measured material-property interpretation.

Figure 3 compares digital sphere counts with the continuous sphere volume over declared voxel radii. The calculation is deterministic. Its oscillations reflect the discrete grid rather than measurement noise. Small objects have a large boundary-to-volume ratio, making them especially sensitive to voxelization. A minimum-size filter can remove the most unstable objects, but it also changes the defect population represented by the dataset. That exclusion should be part of the reported target, not an invisible preprocessing convenience.

Equation (3) gives the relative volume increase for an ideal sphere whose boundary moves outward by b. The expression is exact for that constructed geometry. Real pores are not generally spherical, and a morphological operation need not correspond to a uniform physical offset. The calculation nevertheless provides a dimensional warning: a small boundary displacement can cause a large relative change for a small object. A segmentation score should therefore be connected to descriptor error at the relevant size scale.

$$\Delta V / V = \left(1 + b / r\right)^3 - 1 \tag{3}$$

Figure 4 applies declared binary dilation to two constructed regions. The component count changes when the gap closes. The transformation can make an image look smoother while changing the number of defects. In a real pore network, that could alter a connectivity-related descriptor substantially. A postprocessing operation should be evaluated with the same engineering quantities as the raw prediction. Visual cleanup alone is not evidence of improved physical characterization.

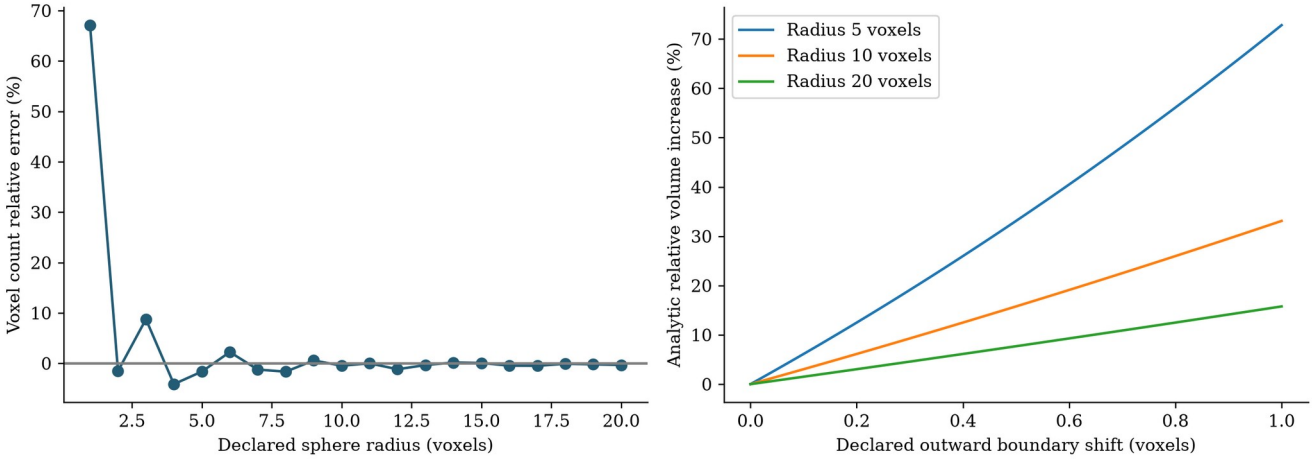


Figure 3. Deterministic sphere voxelization error and exact relative sphere-volume change under declared outward boundary offsets. These curves describe constructed geometry and are not empirical XCT uncertainty estimates.

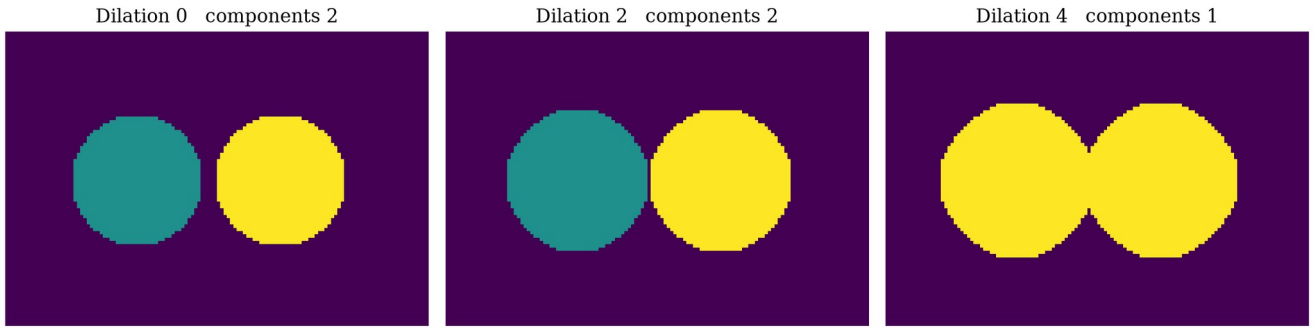


Figure 4. Declared binary dilation of two constructed regions at zero, two, and four iterations, with connected-component counts under the script's default neighborhood. The example demonstrates topology changes rather than improved defect metrology.

5 Learning Representations and Baselines

A random forest on explicit pore descriptors tests whether the available shape summaries already explain a target [9]. A support vector machine provides another compact comparison [10]. Residual networks can learn local image structure [11]. Patch transformers and hierarchical encoders can represent broader spatial organization [12], [13]. Dense encoder-decoders can segment pore regions [14]. These methods should be compared using the same inspected volumes and physical support. Otherwise, an apparent architecture gain can originate from different crops, resolutions, or labels.

Class imbalance occurs because solid material often occupies most voxels. Focal loss can emphasize difficult or rare observations during fitting [15]. It does not define which small pores are physically resolvable. Language-aligned representations can assist image retrieval or descriptive grouping [16], but words such as spherical or lack of fusion require a documented morphological interpretation. A descriptive label should not be treated as a measured manufacturing mechanism unless independent evidence supports that assignment.

Transfer between machines, alloys, scan settings, and reconstruction pipelines is a domain problem [17]. Boosted trees can combine process parameters with measured descriptors [18]. The data split should hold out complete specimens or builds for the corresponding claim [19]. Randomly distributing slices from the same

volume across training and testing shares much of the physical object and imaging noise. Model-comparison uncertainty should be based on independent builds or specimens where those are the experimental replicates [20].

A second comparison should distinguish detection from metrology. A detector can locate a candidate pore while estimating its boundary poorly. A segmentation can have high overlap yet distort the surface area or narrow neck. The primary metric should follow the engineering use. For a screening task, missed critical defects may dominate. For a descriptor regression, size- and shape-dependent bias may be more relevant. The paper should report both the intermediate image measure and the final quantity when the latter is its claim.

6 Uncertainty of the Derived Quantity

Calibration can be assessed for predicted pore presence or class probabilities [21]. Proper scoring rules evaluate those probability statements [22]. They do not by themselves calibrate a total-volume interval. Voxel errors are spatially correlated and can cluster at boundaries. Treating each voxel as an independent Bernoulli variable can yield an unrealistically narrow interval for a connected defect. A useful uncertainty analysis should preserve plausible spatial structure or use independent repeat measurements when available.

Ensembles can expose variation among fitted segmenters [23]. Attribution can help investigate dependence on reconstruction artifacts or surrounding material [24]. Shared shortcuts remain possible [25]. For example, a model can associate a build condition with a particular scan artifact rather than the defect morphology of interest. A held-out scan configuration and a controlled reconstruction audit are more direct tests of that dependence than an explanation map alone.

Equation (4) gives a dimensionless compactness-like descriptor in terms of volume and surface area. Its value depends on the surface estimator as well as the mask. A staircase digital boundary and an interpolated triangulated surface can produce different areas. The descriptor should therefore be reported with the surface extraction method and physical calibration. A model cannot be meaningfully compared on this quantity if each pipeline computes it under a different geometric convention.

$$Q_{shape} = 36\pi V^2 / A_{surface}^{\ 3} \tag{4}$$

Table 2. Prospective evaluation units and independent references.

Evidence or condition	Required comparison	Interpretation
A complete manufactured build and specimen	Hold out complete builds before producing image patches	Reference inspection of held out specimens
An in process sensor history and later xct volume	Compare mappings with declared alignment uncertainty	Synchronized process logs and independent volume registration
A specimen with a sparse defect population	Evaluate specimen and defect outcomes alongside voxel scores	Independent defect reference under the engineering criterion
A model update for a changed xct condition	Compare frozen and adapted models on the same later specimens	Independent later reference measurements and calibration checks

7 Experimental Design for a New Inspection Study

A new study should begin by declaring the specimen population and defect quantities. Record build parameters, alloy, specimen geometry, scan settings, voxel dimensions, reconstruction, and reference procedure. Split complete builds before deriving slices or patches. Hyperparameter search should occur inside the training builds [26]. A final independent build should remain unavailable to threshold, denoising, and morphology decisions. Repeatedly adjusting those decisions against the final volume turns it into development data.

Distribution tests can compare descriptor or image-feature distributions across incoming scans [27]. An alarm may indicate a new manufacturing condition, a new scanner response, or a changed reconstruction procedure. It is a prompt for investigation, not a diagnosis of process failure. A regularized linear baseline can help assess whether a complex predictor improves a defined property estimate beyond a stable transparent alternative [28]. Its residuals should be examined by pore size, specimen, and scan condition.

A useful repeatability study scans the same object under declared conditions. It can separate variation in acquisition and reconstruction from variation in learning. A reference study can compare segmentation-derived measurements with an appropriate independent method, taking account of its own support and uncertainty. Neither comparison should assume that one instrument is exact. Agreement needs a common measurand and a documented relationship between the sampled volumes.

Table 3. Failure mechanisms and records needed for interpretation.

Evidence or condition	Required comparison	Interpretation
Partial volume effects and uncertain thresholds	Descriptor variation across plausible masks	Reference instructions and threshold histories
Border clearing and limited field of view	Size dependent inclusion and descriptor bias	Boundary flags and all exclusion rules
Reconstruction artifacts and new material appearances	Paired descriptor errors and valid coverage	Scan versions training cutoffs and calibration records
Volume shape location and connectivity contributing differently	Property residual sensitivity to each descriptor	Property model assumptions and specimen linkage
Mode collapse and unrepresented shape variation	Shape topology and downstream sensitivity	Generation seeds model version and selection criteria

8 Dimensional Checks and Specimen Studies

The following dimensional and experimental checks define a new inspection programme. They are prospective comparisons and do not add measured specimens to the published NIST example. Their sequence links voxel calibration, boundary interpretation, independent builds, process alignment, connected components, and engineering endpoints. Each check identifies the quantity that changes when a processing choice changes. This is necessary before a learning improvement can be interpreted as better volumetric metrology.

8.1 Physical voxel scale

The first requirement in physical voxel scale is a traceable observation. The experimental unit is a reconstructed specimen volume, and the quantity of interest is pore volume with calibrated dimensions. The main factor to audit is anisotropic voxels and resampling. This test should begin with the original measurement record, before fitted scores or attractive examples are examined. A source value should retain its units, sampling support, and availability status. If a derived quantity changes, the study should be able to identify the transformation responsible. Otherwise, a difference attributed to learning can originate in the measurement pipeline. The reference should be scan calibration and known dimensional references, with its own uncertainty stated rather than assumed away.

The proposed comparison is to compare measurements before and after documented geometric transforms. This is a prospective protocol, not an additional experiment reported as completed here. It should use the same eligible units and the same target convention on both sides. Acquisition and transformation choices should be fixed before the final evaluation. The diagnostic is volume consistency and dimensional residual. Examine its distribution across the actual independent units, retaining failures and unsupported observations in the denominator where the policy requires them. Report the raw counts or residuals alongside any normalized summary, because an identical percentage can represent very different evidential support in two collections.

The key interpretive limit is that array shape alone does not establish physical scale. If the comparison changes both physical support and the learned representation, it cannot isolate an architectural benefit. A useful sensitivity analysis changes one documented component at a time and reports whether the conclusion persists under reasonable alternatives. Preserve voxel dimensions and coordinate transforms. These records connect the calculation to the source and make an independent reconstruction possible. A technically successful result would identify the conditions under which the observation remains usable, the conditions under which it fails, and the additional measurement needed to resolve those failures. It would not infer a wider application merely from a clean numerical output.

8.2 Reference boundaries

The reference study for reference boundaries concerns a pore boundary and surrounding material. It asks whether a consistent segmented defect surface can be supported consistently when the evidence is influenced by partial volume effects and uncertain thresholds. Begin with a written reference convention and a sample chosen independently of model confidence. The reference is repeat scans or justified independent reference measurements. Reviewers should have the information needed for that convention, but an intended independent judgment should not be anchored by the model's proposed answer. Retain disagreements and the reason for uncertainty. Consensus can be useful, yet it should not erase the fact that some observations support more than one reasonable interpretation.

To investigate this question, compare independent segmentations with boundary uncertainty retained. Score descriptor variation across plausible masks, and distinguish uncertainty in the reference from disagreement with a definite label. An ambiguous case can be evaluated through admissible alternatives, an uncertainty interval, or an explicitly excluded category, provided the choice is fixed in advance. Exclusion changes the evaluated population and must remain visible. The study should also measure the effect of the reference policy on the substantive conclusion. A small average disagreement can matter if it concentrates in rare or consequential cases. Conversely, tiny discrepancies may be scientifically immaterial under the intended measurement tolerance.

Interpretation should respect the fact that a visually smooth boundary is not automatically a more accurate surface. The purpose of review is to establish what the observation supports, not to make every case fit a binary success story. Compare the result under reasonable reference conventions without selecting the most favorable one after seeing the model. Preserve reference instructions and threshold histories, including the unresolved cases and the final adjudication history. These materials allow later researchers to distinguish a model improvement from a change in annotation practice. They also identify where a better instrument, another observation, or domain expertise would provide more value than further optimization against an uncertain label.

8.3 Independent builds

The generalization test for independent builds uses a complete manufactured build and specimen as the unit that must remain intact. Its target is transfer to an unseen manufacturing replicate, while the dependence of concern is shared slices and process conditions. A random split of convenient numerical rows can conceal this dependence. The partition should therefore be created at the source-unit level before overlapping crops, windows, or augmented records are derived. Each child record should inherit its parent's assignment. The independent reference is reference inspection of held out specimens. This establishes an evaluation population that corresponds to the claim rather than to the easiest available data arrangement.

The proposed experiment is to hold out complete builds before producing image patches. Evaluate build conditional descriptor and detection errors for the held-out units, showing their range as well as a pooled value. The split should also state which kinds of novelty remain shared across the boundary. Holding out one axis does not automatically hold out every other axis. If the intended use requires several changes at once, assess that combined condition explicitly and use additional comparisons to diagnose the source of any deterioration. Development decisions should be made without inspecting the locked outer reference labels.

The main limit is that many slices share the same physical object. An uncertainty calculation should resample or otherwise model the units that are independent for the scientific question. Treating correlated children as independent trials can make a small difference appear decisive. Preserve build identifiers and patch ancestry, including the construction of every derived record. A successful test would support a statement about the chosen transfer axis and the represented population. It would also identify whether errors concentrate in particular groups. This is more informative than a single favorable split that cannot be related to a concrete future observation setting.

8.4 Process to inspection timing

Timing in process to inspection timing should be defined for an in process sensor history and later XCT volume. The intended statement is a defensible mapping from observed process data to defects, and the potential mismatch is misaligned process and inspection coordinates. Record when the physical observation occurred, when the processed input became available, and when the model or decision was issued. These timestamps can differ substantially. A retrospective array can contain information unavailable at the nominal issue time even when its elements carry earlier labels. The timing reference is synchronized process logs and independent volume registration. The test should reconstruct the allowed information state rather than assuming that a correctly sorted table is necessarily causal.

The proposed comparison is to compare mappings with declared alignment uncertainty. Measure prediction error by mapping confidence under the actual buffering and processing policy. A result should retain the exposure and issue schedule, including intervals that produce no usable output. Delayed reference information must not enter an earlier model update. If the target interval overlaps successive issues, the

dependence should remain part of verification. A useful ablation changes the permitted history or waiting rule while holding other conditions fixed. This separates an improvement due to additional observation from an improvement due to a better representation of the same observation.

The interpretation is constrained because a later internal measurement was not directly observed during printing. Preserve registration transforms and sensor timestamps, so that another analysis can replay the issue sequence and verify what the system could have known. Report the distribution of delays, not merely the average forward-pass time of one model call. Long tails and missed deadlines can determine usefulness for a time-sensitive action. A defensible result identifies the time at which information becomes actionable and the conditions under which the intended interval is no longer supported. It does not infer timely operation solely from accurate retrospective localization or classification.

8.5 Partial defects

The incomplete-observation study for partial defects concerns a pore intersecting the inspected boundary. It tests an explicitly defined population of complete or partial pores when observations are affected by border clearing and limited field of view. Missingness should have an explicit representation and provenance. A numeric placeholder is not automatically a physical zero, and an absent output is not automatically a negative observation. Define the eligibility rule and fallback before evaluating results. The reference is known inspected geometry and source boundary intersections, which should distinguish unavailable measurement from a genuine observed absence. This separation is necessary to keep both coverage and accuracy interpretable.

The experiment should compare retaining flagging and excluding boundary objects. Evaluate size dependent inclusion and descriptor bias jointly, because a method can improve accepted accuracy simply by declining difficult inputs. Record the reasons for rejection and examine whether they are associated with the target or with a particular population. A constructed deletion test is useful for isolating an algorithmic dependency, but it should remain identified as constructed. Naturally occurring interruption may have a different mechanism and may correlate with difficult environmental or operational conditions. Both forms of evidence can be useful if their roles are kept separate.

The limit is that excluding partial objects can change the defect population. A robust system should make its unsupported interval visible and retain enough context to choose a more informative next observation. Preserve boundary flags and all exclusion rules. Report the number and distribution of eligible, accepted, rejected, and recovered units under the same accounting rule. This prevents a clean-looking map or sequence from implying coverage it does not have. The substantive conclusion should concern reliability under the tested missingness conditions, with a clear account of how the fallback changes the scientific or operational target. An untested imputation strategy should not silently turn absent evidence into confident fact.

8.6 Rare critical objects

The prevalence test for rare critical objects takes a specimen with a sparse defect population as its exposure population. The intended outcome is reliable detection of the relevant critical defect class. The difficulty is large solid background and class imbalance, which can make the evaluation mixture unlike the mixture encountered in use. State how the source sample was selected and which conditional rates are being transported to a new scenario. The independent reference is independent defect reference under the engineering criterion. A balanced development set can be appropriate for fitting, but its class proportions should not be mistaken for a measurement of the future population.

The proposed analysis is to evaluate specimen and defect outcomes alongside voxel scores. Measure missed critical objects and false candidates per volume, retaining the eligible exposure denominator and the decision aggregation rule. A probability sample can estimate operating performance directly within its supported population. Scenario reweighting can expose consequences under alternative assumed proportions, but the assumed stability of conditional errors must be stated. It is often useful to report both the empirical source result and the scenario calculation. They answer different questions. Do not select only periods with known target events when the claimed workload concerns a continuous stream that also contains long target-free intervals.

Interpret the result with the limit that pooled voxel accuracy can hide missed pores. Preserve defect size classes and the evaluated physical volume, including the source sampling and any weights. A useful conclusion connects false detections and misses to a defined amount of observation, review, or action. It should show where uncertainty in prevalence materially changes the decision. Strong performance on an enriched collection can justify further study while remaining insufficient for a rare-event deployment claim. Making the population and exposure explicit helps readers distinguish those levels of evidence without dismissing the value of a carefully controlled benchmark.

8.7 Digital topology

The controlled perturbation in digital topology is applied to a constructed or measured connected pore network. The scientific target remains stable connected component and shape measurements; the factor varied is one voxel bridges and morphological operations. Begin by explaining why the intervention represents a meaningful change in the observation process. A numerical alteration can diagnose dependence without reproducing a real physical condition. Its value is greatest when the preserved information and the changed information are both explicit. The reference is known constructed geometry and reviewed measured objects, and any constructed component should retain a known design rather than being described as a new empirical observation.

The proposed experiment is to compare raw masks and declared connectivity or dilation conventions. Assess component count surface and largest pore changes using the complete processing chain that the claim requires. If an intervention changes raw observations, recompute the actual derived inputs rather than changing a convenient proxy and assuming equivalence. Include a control that leaves the relevant physical quantity unchanged where that is possible. Compare errors across the declared perturbation range, avoiding a single unusually favorable setting. The test can reveal a dependence, but identifying its mechanism may require additional observations that vary the suspected factors independently.

The interpretive boundary is that image cleanup can change the engineering measurand. Preserve neighborhood rules and morphology parameters, and keep the transformed observations distinguishable from the originals. A successful diagnostic explains which part of the model or measurement becomes unstable and at what support scale. It should not infer broad robustness from an algebraic property of one feature or an attractive transformed display. The next step can be a physical replication under the corresponding condition, with independent reference measurements. This progression connects a reproducible analytic test to the real observation without claiming that the former has already substituted for the latter.

8.8 Scanner adaptation

The adaptation study for scanner adaptation evaluates a model update for a changed XCT condition. Its desired outcome is improved measurement on subsequent independent scans, and the main risk to interpretation is reconstruction artifacts and new material appearances. A change detector and a model-

quality comparison answer different questions. The former identifies a departure in a selected representation; the latter asks whether a candidate improves the intended output. Define the training cutoff, permitted labels, update budget, and promotion rule before the later evaluation. The independent reference is independent later reference measurements and calibration checks. Both an incumbent and a challenger should be assessed on the same eligible future observations.

The proposed comparison is to compare frozen and adapted models on the same later specimens. Evaluate paired descriptor errors and valid coverage with paired records wherever possible. Labels acquired after an update cannot be treated as if they had been available for fitting it. Pseudo-labels should retain their generated provenance and should not silently become independent reference truth. An annotation budget should include the actual selection and review process, because targeted easy examples can make adaptation look cheaper than it is. A frozen reference is valuable even when it is ultimately inferior, because it exposes whether change was needed and what the update actually contributed.

The interpretation must recognize that a distribution shift can originate in scanning rather than manufacturing. Preserve scan versions training cutoffs and calibration records, including rejected candidates and the conditions that triggered review. An adaptation result should report stability, coverage, and relevant errors rather than only the best post-update score. If the update is harmful for a rare subgroup, a pooled improvement can hide that consequence. A defensible promotion decision rests on independent later evidence and a declared operating objective. It should not be justified merely by a distribution alarm, a newer architecture, or a visually persuasive explanation of an individual prediction.

8.9 Engineering endpoint

The decision-level test for engineering endpoint uses a pore representation used in a property analysis. It asks whether a defined defect related engineering quantity remains supported after the model output enters the intended analysis or action. The factor of concern is volume shape location and connectivity contributing differently. Define the downstream quantity before examining model results, including any aggregation, threshold, matching, or cost assumptions. The reference is independent material tests with matched inspected support. This creates a link between an intermediate score and the consequence that motivates the work. Without that link, a favorable technical metric can be assigned practical meaning that the experiment never measured.

The proposed comparison is to compare reference and predicted descriptors under the same property model. Report property residual sensitivity to each descriptor alongside the intermediate errors, so that their relationship can be inspected rather than presumed. Different error types can have unequal effects on the endpoint. A small number of consequential errors may dominate a summary that looks stable on average. Conversely, some image or sample discrepancies may have little effect on the intended quantity. A sensitivity analysis should vary defensible decision assumptions without choosing the one that makes the model look best. Retain the underlying counts and source-supported measurements needed to assess alternatives.

The key limit is that segmentation overlap does not determine mechanical property accuracy. Preserve property model assumptions and specimen linkage, including unresolved or declined decisions. A useful result identifies the conditions under which the model changes an endpoint, the uncertainty of that change, and the additional evidence required for a stronger interpretation. This is a substantive validation question, not an administrative final step. It allows a complex learning system to be judged on the scientific or operational contribution it actually makes while keeping broader claims available for separate, appropriately designed studies.

8.10 Synthetic volume use

The external study for synthetic volume use concerns a generated pore and its intended inspection experiment. Its claim is useful synthetic support with preserved provenance, under changes in mode collapse and unrepresented shape variation. Replication should begin from a locked feature, label, and decision specification, while documenting any necessary local mapping. The independent reference is independent measured pore populations under matched selection. A new environment may use the same field names or category words while measuring something different. Check those meanings before comparing numerical scores. Unsupported mappings should remain visible so that a transfer result is not created by silently redefining the task.

The proposed experiment is to evaluate generated objects against held out measured descriptors. Measure shape topology and downstream sensitivity, including the availability of valid inputs and the effort needed to obtain the reference. Report local conditions alongside the aggregate result. A model may be useful within a constrained envelope even when it fails elsewhere. The aim is to identify that envelope precisely, not to turn every local success into a universal claim. A reproducibility package should retain the configuration needed to reconstruct preprocessing, fitting, issue, and scoring, while preserving the permitted provenance of any source material that cannot be redistributed freely.

The limiting explanation is that visual realism is not independent evidence of a manufactured defect. Preserve generation seeds model version and selection criteria. The final account should distinguish what was observed, what was fitted, what was constructed for diagnosis, and what remains a proposed next test. Numerical reproducibility is necessary for the executed calculations, but it does not by itself establish external scientific validity. An informative replication can therefore confirm a useful result, reveal a boundary, or identify a measurement mismatch. Each outcome contributes knowledge when the source records and target definitions make the reason assessable.

9 Conclusions

Learning for pore inspection should be organized around a three-dimensional engineering quantity. The NIST extraction example shows why boundary selection and connected components are part of the observation process. Constructed volumes then isolate voxelization, boundary displacement, and topology without claiming new specimen data. These calculations explain why a cleaner-looking mask or a stronger image score can still yield a worse defect measurement.

A reproducible next experiment should hold out complete builds, retain physical calibration, evaluate raw and postprocessed outputs, and measure descriptor errors across pore sizes and scan conditions. Synthetic volumes should retain their generated provenance and be tested for the intended use. Under that approach, artificial intelligence can contribute to volumetric inspection while leaving the material and measurement assumptions available for engineering review.

References

[1] M. Hoque, F. H. Kim, "Generative modeling of additive manufacturing lack-of-fusion defects using three-dimensional adversarial network", National Institute of Standards and Technology (U.S.), 2026, doi: 10.6028/nist.ams.100-73.

[2] B. Kim, S. P. K.R., Y. Natarajan, Danushkumar, S. Park, S. Pyo, "Advanced stacked modeling techniques for material porosity estimation via high-resolution computed tomography imaging", Scientific Reports, vol. 16, no. 1, Art. no. 12714, 2026, doi: 10.1038/s41598-026-39748-1.

[3] T. DebRoy et al., "Additive manufacturing of metallic components – Process, structure and properties", Progress in Materials Science, vol. 92, pp. 112-224, 2018, doi: 10.1016/j.pmatsci.2017.10.001.

[4] A. du Plessis, I. Yadroitsev, I. Yadroitsava, S. G. Le Roux, "X-Ray Microcomputed Tomography in Additive Manufacturing: A Review of the Current Technology and Applications", 3D Printing and Additive Manufacturing, vol. 5, no. 3, pp. 227-247, 2018, doi: 10.1089/3dp.2018.0060.

[5] J. A. Mitchell, T. A. Ivanoff, D. Dagel, J. D. Madison, B. Jared, "Linking pyrometry to porosity in additively manufactured metals", Additive Manufacturing, vol. 31, Art. no. 100946, 2020, doi: 10.1016/j.addma.2019.100946.

[6] W. Wang, P. Wang, H. Zhang, X. Chen, M. Chen, J. Li, "An In-Situ Deep Learning-Based Defect Detection Technology for Additive Manufacturing Process", 2023 8th International Conference on Integrated Circuits and Microsystems (ICICM), pp. 65-70, 2023, doi: 10.1109/icicm59499.2023.10366010.

[7] S. van der Walt et al., "scikit-image: image processing in Python", PeerJ, vol. 2, pp. e453, 2014, doi: 10.7717/peerj.453.

[8] I. Goodfellow et al., "Generative Adversarial Nets", Advances in Neural Information Processing Systems, vol. 27, pp. 2672-2680, 2014, [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

[9] L. Breiman, "Random Forests", Machine Learning, vol. 45, no. 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.

[10] C. Cortes, V. Vapnik, "Support-vector networks", Machine Learning, vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.

[11] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016, doi: 10.1109/CVPR.2016.90.

[12] A. Dosovitskiy et al., "An Image is Worth 16x16 Words Transformers for Image Recognition at Scale", International Conference on Learning Representations, 2021, [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy.

[13] Z. Liu et al., "Swin Transformer Hierarchical Vision Transformer Using Shifted Windows", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012-10022, 2021, [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html.

[14] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Lecture Notes in Computer Science, pp. 234-241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[15] T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, "Focal Loss for Dense Object Detection", 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999-3007, 2017, doi: 10.1109/ICCV.2017.324.

[16] A. Radford et al., "Learning Transferable Visual Models From Natural Language Supervision", Proceedings of the 38th International Conference on Machine Learning, vol. 139, pp. 8748-8763, 2021, [Online]. Available: https://proceedings.mlr.press/v139/radford21a.html.

[17] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J. W. Vaughan, "A theory of learning from different domains", Machine Learning, vol. 79, no. 1-2, pp. 151-175, 2010, doi: 10.1007/s10994-009-5152-4.

[18] T. Chen, C. Guestrin, "XGBoost: A Scalable Tree Boosting System", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016, doi: 10.1145/2939672.2939785.

[19] D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure", Ecography, vol. 40, no. 8, pp. 913-929, 2017, doi: 10.1111/ecog.02881.

[20] C. Nadeau, Y. Bengio, "Inference for the Generalization Error", Machine Learning, vol. 52, no. 3, pp. 239-281, 2003, doi: 10.1023/a:1024068626366.

[21] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, "On Calibration of Modern Neural Networks", Proceedings of the 34th International Conference on Machine Learning, vol. 70, pp. 1321-1330, 2017, [Online]. Available: https://proceedings.mlr.press/v70/guo17a.html.

[22] T. Gneiting, A. E. Raftery, "Strictly Proper Scoring Rules, Prediction, and Estimation", Journal of the American Statistical Association, vol. 102, no. 477, pp. 359-378, 2007, doi: 10.1198/016214506000001437.

[23] B. Lakshminarayanan, A. Pritzel, C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles", Advances in Neural Information Processing Systems, vol. 30, pp. 6402-6413, 2017, [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html.

[24] S. M. Lundberg, S. I. Lee, "A Unified Approach to Interpreting Model Predictions", Advances in Neural Information Processing Systems, vol. 30, pp. 4765-4774, 2017, [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

[25] R. Geirhos et al., "Shortcut learning in deep neural networks", Nature Machine Intelligence, vol. 2, no. 11, pp. 665-673, 2020, doi: 10.1038/s42256-020-00257-z.

[26] J. Bergstra, Y. Bengio, "Random Search for Hyper-Parameter Optimization", Journal of Machine Learning Research, vol. 13, pp. 281-305, 2012, [Online]. Available: https://jmlr.org/papers/v13/bergstra12a.html.

[27] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, A. Smola, "A Kernel Two-Sample Test", Journal of Machine Learning Research, vol. 13, pp. 723-773, 2012, [Online]. Available: https://jmlr.org/papers/v13/gretton12a.html.

[28] A. E. Hoerl, R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems", Technometrics, vol. 12, no. 1, pp. 55-67, 1970, doi: 10.1080/00401706.1970.10488634.