Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI-Based Particle Size Distribution Prediction from 3D Packed-Particle Tomographic Volumes: A Comparative Study with classical Watershed Segmentation

Ruhi Gupta
Indian Institute of Technology Roorkee
Mayank Goswami
mayankfph@iitr.ac.in

Indian Institute of Technology Roorkee

Research Article

Keywords: Particle size distribution, 3D Convolutional Neural Networks, watershed segmentation, random sequential addition, optimal transport, X-ray microtomography

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11634141/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

AI-Based Particle Size Distribution Prediction from 3D Packed-Particle Tomographic Volumes: A Comparative Study with Classical Watershed Segmentation

Ruhi Gupta[1] and Mayank Goswami,[1,2]

[1]Department of Physics, IIT Roorkee
[2]MF School of AI and DS, IIT Roorkee
mayank.goswami@ph.iitr.ac.in

***Abstract*—Determining the Particle Size Distribution (PSD) of granular media and powder compacts is vital across condensed matter physics, geological sciences, and materials engineering. While X-ray computed tomography (μCT) captures granular microstructure non-destructively, traditional quantitative PSD estimation relies on classical morphological segmentation (binarization, distance transform, and watershed flooding), which is computationally intensive and sensitive to contact-neck ambiguities. In this work, we investigate whether a compact 3D Convolutional Neural Network (Compact3DCNN, 300,900 parameters) can predict normalized 20-bin PSD histograms directly from unsegmented 3D grayscale volumes without intermediate particle instance extraction. Using a synthetic dataset of 60 independent non-overlapping polydisperse packed-sphere volumes (48^3 voxels) with exact analytic PSD ground truth, we benchmark the CNN against a classical 3D Euclidean distance-transform watershed baseline on an untouched, held-out test split ($N = 10$, 40 train / 10 validation / 10 test). On this low-to-moderate packing regime (15%–25%), the classical watershed baseline achieved superior accuracy, attaining a mean Earth Mover's Distance (EMD) of 0.0974 ± 0.0261 voxels compared to 0.3428 ± 0.0949 voxels for the CNN (paired 95% bootstrap CI: $[+0.1837, +0.3156]$ voxels). Measured runtime was 1.19 ± 0.27 ms per volume for the CNN on an NVIDIA RTX 3050 Laptop GPU versus 33.71 ± 7.50 ms for the watershed pipeline on an Intel CPU; however, we emphasize that this reflects disparate computing architectures rather than a controlled hardware comparison. Furthermore, comparison against an empirical training-set-mean baseline (0.6206 ± 0.2440 vx EMD on validation) confirms that the CNN learns volume-specific morphological features (0.3970 ± 0.1335 vx, a 36.0% error reduction). A controlled validation experiment evaluating direct 1D Wasserstein-1 loss training yielded substantially higher error (1.1382 ± 0.4907 vx EMD) than the primary Kullback-Leibler (KL) divergence configuration, illustrating optimization challenges in cumulative L1 histogram penalties. Our findings characterize the accuracy-throughput profile in direct volumetric PSD regression and delineate key open challenges for dense granular systems.**

***Index Terms*—Particle size distribution, 3D Convolutional Neural Networks, watershed segmentation, random sequential addition, optimal transport, X-ray microtomography.**

I. Introduction

PARTICLE Size Distribution (PSD) is a governing constitutive characteristic of granular systems, influencing bulk packing density, hydraulic permeability, mechanical shear strength, and thermal conductivity [1]. In modern physical sciences and industrial powder metallurgy, 3D X-ray micro-computed tomography (μCT) enables non-destructive volumetric inspection of granular packings [2].

Conventionally, deriving a PSD from 3D tomographic images requires a multi-stage image processing pipeline: intensity filtering, global or local binarization (e.g., Otsu thresholding), Euclidean distance transform computation, marker extraction, and 3D morphological watershed segmentation [3]. While effective under high-contrast conditions with low packing fractions, this classical paradigm exhibits practical and physical constraints:

1) **Computational Bottleneck:** 3D distance transforms and topological watershed flooding require iterative processing that scales with voxel count, posing latency challenges for real-time 4D dynamic tomography.
2) **Contact Sensitivity:** In dense assemblies, contact points form narrow bridging necks, frequently causing over-segmentation from secondary distance peaks or under-segmentation from marker omission.
3) **Accumulated Pipeline Errors:** Thresholding artifacts propagate directly into segmented voxel volumes, distorting equivalent-sphere diameter measurements.

Recent deep learning approaches in computer vision demonstrate that statistical physical properties can be learned directly from spatial field representations [4]. In this research, we pose a focused question: *Can a compact 3D Convolutional Neural Network directly regress a normalized discrete PSD histogram from a 3D volumetric intensity field, bypassing instance segmentation?*

To address this question under strict scientific constraints, we develop an original, fully reproducible research pipeline comprising:

- A Random Sequential Addition (RSA) generator producing polydisperse, non-overlapping packed spheres with

Corresponding author: mayank.goswami@ph.iitr.ac.in

simulated CT point spread blur and detector noise, coupled with exact analytic ground-truth PSD histograms computed from generating radii.
- A classical 3D distance-transform watershed baseline.
- A lightweight 3D CNN architecture (Compact3DCNN) tailored for 4 GB VRAM environments.
- A leak-free stratified experimental protocol (40 train / 10 val / 10 held-out test).
- A training-set-mean reference baseline confirming non-trivial volumetric learning.
- A controlled validation study testing direct 1D Wasserstein-1 versus KL divergence loss functions.

II. Related Work

Quantitative characterization of granular microstructures from 3D imaging has a rich history across applied mathematics, materials science, and computational physics.

A. Morphological and Geometric Segmentation

The marker-controlled watershed transform, introduced formally by Beucher and Meyer [3], remains the foundation of volumetric granular segmentation. In porous and granular media, Sheppard et al. [2] established standard tomographic enhancement and distance-transform-based separation techniques for X-ray microtomography. While robust in dilute systems, watershed techniques encounter geometric challenges when granular contacts proliferate, as contact necks create saddle points in the Euclidean distance map, often precipitating over-segmentation or under-segmentation.

B. Random Packings and Analytic Phantoms

Analyzing physical properties of disordered media relies heavily on idealized packing models. Torquato et al. [1] rigorously examined random close packings of hard spheres and their structural signatures. Random Sequential Addition (RSA), wherein particles are irreversibly placed into a container under non-overlap exclusion, serves as an established benchmark for generating synthetic granular fabrics with exact ground truth. Such models eliminate circular dependencies in algorithm evaluation, where segmented masks from one tool are otherwise used to validate another.

C. Deep Learning for Granular Characterization

With the advent of deep convolutional networks and autodifferentiation frameworks [4], machine learning has increasingly been applied to tomographic data. Typical deep learning workflows formulate particle analysis as semantic or instance segmentation (e.g., 3D U-Net or Mask R-CNN variants). While granular instance segmentation preserves spatial boundaries, it requires dense voxel-level annotations and post-processing to accumulate particle statistics. In contrast, global parameter regression directly maps the input field to macroscopic or statistical targets. Lin et al. [5] demonstrated that Global Average Pooling (GAP) can effectively aggregate spatial feature maps into structural descriptors while drastically reducing parameter count and mitigating overfitting.

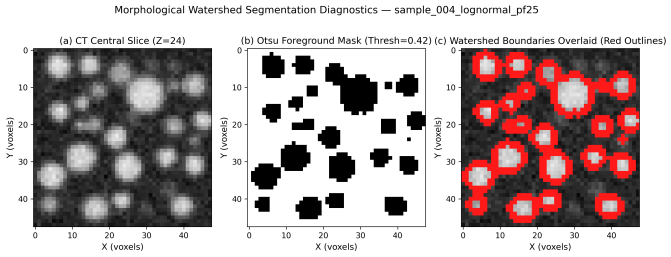


Fig. 1. Morphological watershed pipeline stages: (a) Central synthetic CT grayscale slice ($Z = 24$), (b) 3D Otsu thresholded binary mask, and (c) Watershed separation boundaries overlaid in red.

D. Optimal Transport and Distribution Metrics

Comparing predicted and empirical probability distributions with spatial or metric structure requires metrics that account for bin geometry. While relative entropy (Kullback-Leibler divergence) [6] is canonical in statistical physics and classification, it treats discrete bins as unordered categories. In contrast, the Wasserstein distance (Earth Mover's Distance) [7] incorporates the underlying ground metric between bins. In discrete 1D settings, the Wasserstein-1 distance is analytically tractable via cumulative distribution differences, providing an interpretable physical error metric measured directly in voxel units.

III. Methodology

A. Synthetic Particle Packing and Imaging Simulation

Because experimental μCT scans lack an exact particle-by-particle geometric ground truth, physical verification requires synthetic phantoms with analytically known sphere coordinates and radii.

We implement Random Sequential Addition (RSA) without periodic boundaries. Candidate sphere radii $r \in [r_{\min}, r_{\max}]$ ($r_{\min} = 2.0$ vx, $r_{\max} = 8.0$ vx) are sampled from two target statistical families:

1) **Log-Normal Distribution:** $r \sim \text{LogNormal}(\mu = 1.3, \sigma = 0.35)$, reflecting natural fragmentation and comminution processes.
2) **Bimodal Distribution:** A mixture of two Gaussian modes ($r_1 = 3.0$ vx, $\sigma_1 = 0.5$ vx; $r_2 = 6.0$ vx, $\sigma_2 = 0.7$ vx; equal weights $0.5/0.5$), representative of binary powder mixtures engineered for specific packing properties.

Spheres are sequentially placed within a bounding cube $[0, D_z] \times [0, D_y] \times [0, D_x]$ ($48 \times 48 \times 48$ voxels) under a strict non-overlap clearance rule:

$$\|\mathbf{c}_i - \mathbf{c}_j\|_2 \geq r_i + r_j + \delta, \quad \forall j < i \tag{1}$$

where $\delta = 0.5$ voxels prevents numerical contact singularities during voxelization. Bounding box boundaries constrain sphere centers such that $\mathbf{c}_i \in [r_i, \mathbf{D} - r_i]$. Candidate placement terminates when the target packing fraction $\phi \in [0.15, 0.25]$ is reached or maximum attempts (50,000) are exhausted.

The accepted geometry is rasterized onto a discrete grid:

- **Ground-Truth Label Volume:** Array of unique particle integer IDs used strictly for validation.
- **Synthetic CT Volume:** Attenuation contrast model mapping background to $I_{\text{bg}} = 0.15$ and solid to $I_{\text{solid}} = 0.85$,

followed by 3D Gaussian blur ($\sigma = 0.8$ vx) simulating the beam point spread function (PSF) and additive Gaussian detector noise ($\sigma_{\text{noise}} = 0.04$). Grayscale values are clamped to $[0.0, 1.0]$.

B. Exact Analytic PSD Ground Truth

Ground-truth Particle Size Distributions are computed directly from the generating sphere diameters $d_i = 2r_i$, completely bypassing rasterized voxel approximations. Diameters are partitioned into $K = 20$ logarithmically spaced bins:

$$x_k \in \text{geomspace}(4.0, 13.0, 21), \quad k = 0, \ldots, 20 \tag{2}$$

The discrete histogram is normalized by particle count to form a valid probability simplex vector $\mathbf{y} \in \mathbb{R}^{20}$ ($\sum_{k=1}^{20} y_k = 1.0, y_k \geq 0$).

C. Classical Watershed Baseline Pipeline

The classical baseline operates purely on the noisy grayscale CT volume without accessing ground-truth labels (Fig. 1):

1) **Binarization:** 3D Gaussian smoothing ($\sigma = 0.5$ vx) followed by 3D Otsu thresholding separates solid foreground $\mathcal{M}$ from void space.
2) **Distance Transform:** 3D Euclidean distance transformation $D(\mathbf{x}) = \min_{\mathbf{y} \notin \mathcal{M}} \|\mathbf{x} - \mathbf{y}\|_2$ computes shortest distances to the void interface.
3) **Marker Generation:** Local maxima in $D(\mathbf{x})$ subject to minimum spatial separation ($d_{\min} = 2$ vx) define seed markers.
4) **Watershed Flooding:** Morphological watershed floods the inverted distance map $-D(\mathbf{x})$ constrained by mask $\mathcal{M}$.
5) **PSD Calculation:** Each segmented region's voxel volume V_i is converted to equivalent-sphere diameter:

$$d_i = \left(\frac{6V_i}{\pi} \right)^{1/3} \quad \text{[voxels]} \tag{3}$$

The resulting diameters are binned into the identical 20 logarithmic bins.

D. Compact 3D Convolutional Neural Network

To ensure execution within a 4 GB GPU memory constraint, we designed **Compact3DCNN** (300,900 parameters) implemented in PyTorch [4]:

- **Backbone:** 4 stages of ConvBlock3D ($1 \to 16 \to 32 \to 64 \to 128$). Each stage comprises Conv3D($3 \times 3 \times 3$, bias = False) $\to$ BatchNorm3D $\to$ LeakyReLU(0.1). Downsampling via MaxPool3D($2 \times 2 \times 2$) follows stages 1, 2, and 3, reducing spatial resolution $48^3 \to 24^3 \to 12^3 \to 6^3$.
- **Global Average Pooling (GAP):** Reduces feature maps $[B, 128, 6, 6, 6] \to [B, 128, 1, 1, 1]$ [5]. GAP imparts spatial translation invariance, eliminates parameter scaling with input volume size, and prevents dense layer parameter explosion.
- **Prediction Head:** Linear($128 \to 64$) $\to$ LayerNorm(64) $\to$ LeakyReLU(0.1) $\to$ Dropout(0.1) $\to$ Linear($64 \to 20$).
- **Output Activation:** A final Softmax(dim $= -1$) guarantees non-negative probability mass summing to 1.0.

E. Training Configuration and Loss Formulations

The network was trained using the Adam optimizer [8] with initial learning rate $\eta = 5 \times 10^{-4}$, weight decay 10^{-3}, batch size 4, and mixed precision enabled. A `ReduceLROnPlateau` scheduler halved the learning rate when validation loss failed to improve for 2 consecutive epochs (decaying to $\eta = 1.25 \times 10^{-4}$ by epoch 14), with early stopping monitoring validation loss.

We formulated two distinct training loss objectives:

1) **Kullback-Leibler (KL) Divergence:**

$$\mathcal{L}_{\text{KL}}(\mathbf{z}, \mathbf{y}) = \sum_{k=1}^{K} y_k \log \left(\frac{y_k}{\text{softmax}(\mathbf{z})_k + \epsilon} \right) \tag{4}$$

which produces continuous, smooth residual gradients $\frac{\partial \mathcal{L}_{\text{KL}}}{\partial z_k} = \hat{p}_k - y_k$.

2) **1D Wasserstein-1 (EMD) Loss:** Using the adjacent physical bin center spacings $\Delta c_k = c_{k+1} - c_k$:

$$\mathcal{L}_{W_1}(\mathbf{z}, \mathbf{y}) = \sum_{k=0}^{K-2} |\hat{F}(k) - F_{\text{target}}(k)| \cdot \Delta c_k \tag{5}$$

where $\hat{F}(k) = \sum_{j=0}^{k} \hat{p}_j$ and $F_{\text{target}}(k) = \sum_{j=0}^{k} y_j$, weighting cumulative differences by physical voxel spacing.

F. Evaluation Metrics and Uncertainty Estimation

Model evaluation relies on four complementary physical and statistical metrics:

1) **Earth Mover's Distance (W_1 EMD):** Evaluated in physical voxel units using geometric bin centers $c_k = \frac{x_k + x_{k+1}}{2}$.
2) **Histogram RMSE and MAE:** Measuring point-wise bin discrepancy.
3) **Characteristic Quantiles (D_{10}, D_{50}, D_{90}):** Extracted via piecewise-linear interpolation over cumulative distribution functions to estimate diameter percentiles in voxels.
4) **Inference Latency:** Mean and standard deviation across test samples, measured using high-precision timers (`time.perf_counter()`) following warm-up iterations.

To assess statistical significance on paired test samples ($N = 10$), we compute non-parametric 95% bootstrap confidence intervals across 2,000 resamples of the paired difference $\Delta_i = \text{Metric(CNN)}_i - \text{Metric(Watershed)}_i$.

IV. Experiments and Results

A. Dataset and Stratified Split

A synthetic pilot dataset of 60 independent volumes (48^3 voxels) was generated across packing fractions 15%–25% (30

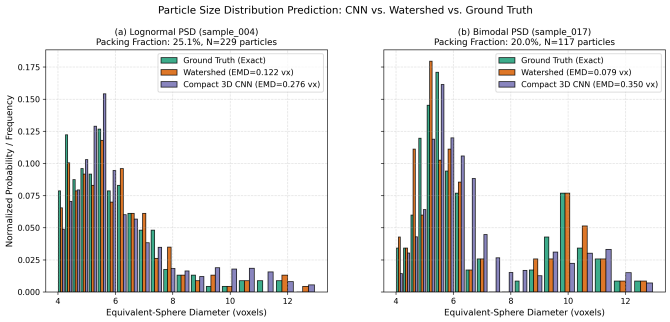


Fig. 2. Representative PSD prediction overlays on held-out test samples: (a) Lognormal distribution (`sample_004`), and (b) Bimodal distribution (`sample_017`).

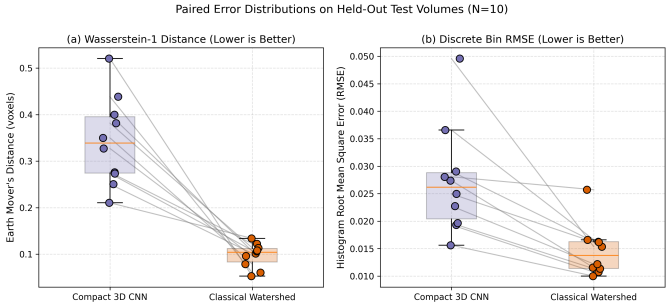


Fig. 3. Paired test error distributions ($N = 10$): (a) Earth Mover's Distance (EMD), and (b) Histogram RMSE.

lognormal, 30 bimodal). The dataset was partitioned by unique parent volume ID:

- **Training Set:** 40 parent volumes (66.7%)
- **Validation Set:** 10 parent volumes (16.7%)
- **Held-Out Test Set:** 10 parent volumes (16.7%)

Strict disjointness audits confirmed zero sample ID overlap across splits. The held-out test set remained sealed until final benchmarking.

B. Verification of Learning vs. Training-Set-Mean Baseline

To verify that the CNN acquired genuine morphological representations rather than predicting a static mean distribution, we constructed an empirical training-set-mean baseline $\bar{\mathbf{y}}_{\text{train}} = \frac{1}{40}\sum_{i\in\text{train}} \mathbf{y}_i$. Evaluating this static distribution on the 10 validation volumes yielded:

- **Training-Mean Baseline:** W_1 EMD $= 0.6206 \pm 0.2440$ vx, RMSE $= 0.0361 \pm 0.0080$, D_{50} abs err $= 0.4724$ vx.
- **Trained Compact3DCNN:** W_1 EMD $= \mathbf{0.3970 \pm 0.1335}$ vx, RMSE $= \mathbf{0.0247 \pm 0.0064}$, D_{50} abs err $= \mathbf{0.1983}$ vx.

The trained CNN reduced validation EMD error by 36.0% and median diameter error by 58.0% relative to the static training mean (25.7% EMD reduction on lognormal samples, 48.7% on bimodal samples). This result is consistent with the network acquiring input-specific morphological information, although identifying precisely which spatial features (e.g., local surface curvature, pore-throat geometry, or high-order contrast moments) the network exploits remains an open question that would require spatial gradient saliency or feature-ablation studies.

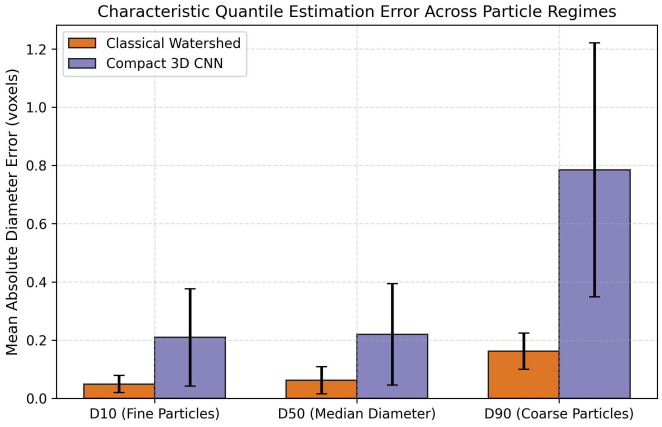


Fig. 4. Mean absolute diameter error across characteristic quantiles (D_{10}, D_{50}, D_{90}) in physical voxel units.

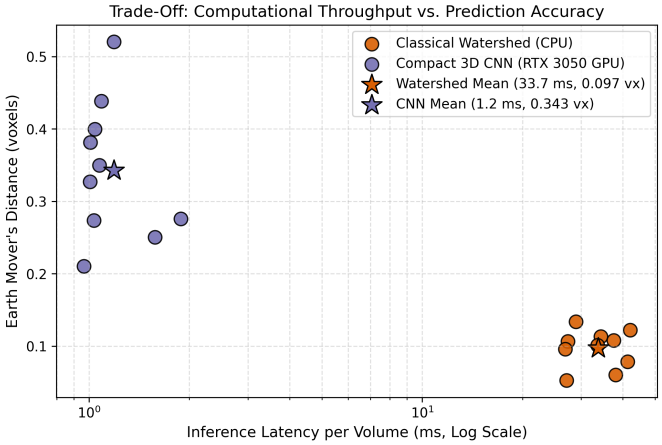


Fig. 5. Accuracy versus computational latency trade-off. Compact3DCNN achieves low latency via GPU execution, whereas watershed executes on CPU.

C. Held-Out Test Benchmark Results

Table I presents the comparative evaluation on the 10 held-out test volumes evaluated across identical bins.

TABLE I
COMPARATIVE BENCHMARK ON HELD-OUT TEST SET ($N = 10$)

Metric	Compact3DCNN	Watershed	Paired Diff 95% CI
W_1 EMD [vx]	0.3428 ± 0.095	$\mathbf{0.0974 \pm 0.026}$	$[+0.184, +0.316]$ vx
Histogram RMSE	0.0273 ± 0.010	$\mathbf{0.0146 \pm 0.005}$	$[+0.008, +0.019]$
Histogram MAE	0.0200 ± 0.008	$\mathbf{0.0101 \pm 0.004}$	$[+0.007, +0.013]$
D_{50} Abs Err [vx]	0.2198 ± 0.126	$\mathbf{0.0622 \pm 0.047}$	$[+0.045, +0.281]$ vx
D_{50} Rel Err [%]	$4.00\% \pm 2.80\%$	$\mathbf{1.14\% \pm 0.89\%}$	$[+0.84\%, +5.08\%]$
Execution Hardware	NVIDIA RTX 3050	Intel Core CPU	—
Reported Latency [ms]	$\mathbf{1.2 \pm 0.3}$ ms	33.7 ± 7.5 ms	Disparate Hardware

As shown in Table I and Figures 2–4, classical watershed achieved superior distributional fidelity on this benchmark (W_1 EMD $= 0.0974$ vx vs. 0.3428 vx for CNN; paired 95% CI: $[+0.1837, +0.3156]$ vx). Watershed also achieved lower error in D_{10} (0.0492 vs 0.2094 vx) and D_{90} (0.1621 vs 0.7847 vx).

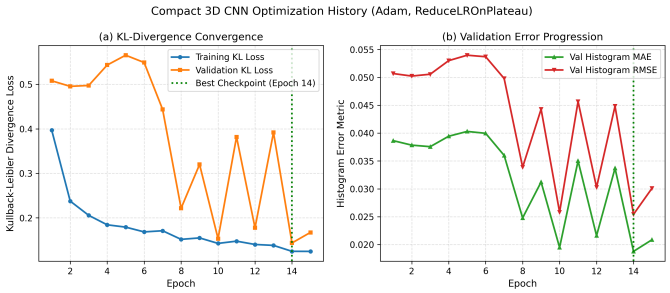


Fig. 6. Training history of Compact3DCNN: (a) KL-divergence loss convergence, and (b) Validation MAE and RMSE across epochs.

D. Controlled Validation Experiment: Wasserstein vs. KL Loss

To investigate whether training directly on the evaluation metric improves performance, we executed experiment `exp1_wasserstein_pilot`. An identical network was trained on the 40 training volumes using `Wasserstein1DPSDLoss` (Adam, $\eta = 5 \times 10^{-4}$, weight decay 10^{-3}, 15 epochs) and evaluated strictly on the 10 validation volumes:

- **KL-Loss Checkpoint:** Validation EMD $= \mathbf{0.3970 \pm 0.1335}$ vx, RMSE $= \mathbf{0.0247 \pm 0.0064}$, D_{50} abs err $= \mathbf{0.1983}$ vx.
- **Wasserstein Loss Checkpoint:** Validation EMD $= 1.1382 \pm 0.4907$ vx, RMSE $= 0.0536 \pm 0.0076$, D_{50} abs err $= 0.9883$ vx.

Under this specific experimental configuration, the cumulative L1 Wasserstein loss performed substantially worse (186% higher validation EMD). In this discrete cumulative setting, the subgradient of the L1 term $\text{sign}(\hat{F}(k) - F(k)) \in \{-1, +1\}$ produces constant-magnitude updates regardless of the residual size, whereas the softmax cumulative sum links all preceding logits, inducing high gradient variance during backpropagation. While this mathematical characteristic offers a compelling explanation for the erratic optimization dynamics observed, the experiment did not fully isolate this cause from potential interactions with learning rate schedules or momentum buffers; dedicated loss-schedule co-ablation would be required to verify this hypothesis definitively. Consequently, the Wasserstein model was not evaluated on the held-out test set, and the KL-trained model was retained as the primary architecture.

V. Discussion and Limitations

A. Accuracy vs. Throughput Profile

In low-to-moderate packing densities (15%–25%), classical watershed segmentation demonstrates low distributional error on these synthetic phantoms (W_1 EMD $= 0.0974 \pm 0.0261$ vx) because non-overlapping spheres provide distinct, unambiguous distance maxima. We emphasize that an aggregate distribution error below one voxel does not imply sub-voxel spatial segmentation accuracy for individual particle boundaries; rather, it reflects accurate volumetric integration and grain accounting on this benchmark. The classical pipeline operates as a localized measurement engine: by identifying individual distance peaks and flooding the inverted distance map, it computes the volume of each particle directly. On this synthetic phantom, where spheres are separated by at least $\delta = 0.5$ voxels, watershed correctly segmented nearly all individual grains (for instance, detecting 229 out of 229 particles in `sample_004` and 194 out of 195 in `sample_013`).

Regarding execution speed, the CNN processed volumes in 1.19 ± 0.27 ms on an NVIDIA RTX 3050 Laptop GPU, compared to 33.71 ± 7.50 ms for the watershed pipeline on an Intel CPU. We explicitly highlight that this is **not a controlled hardware comparison**: the CNN leveraged GPU tensor parallelism with mixed precision, whereas the watershed baseline utilized standard CPU-based morphological operators from `scipy.ndimage`. A controlled comparison would evaluate a GPU-accelerated watershed implementation (such as CUDA distance transforms) or run both pipelines on the CPU. Nonetheless, the measured inference time illustrates the practical throughput advantage of forward CNN passes on commodity GPU accelerators.

B. Observed Error Modes and Physical Mechanisms

Analysis of CNN predictions indicates two primary error modes:

1) **Distribution Over-Smoothing and Entropy Inflation:** CNN predictions exhibit elevated Shannon entropy (2.689 nats on validation vs. 2.382 nats for ground truth), broadening sharp modal peaks into adjacent bins. A plausible hypothesis for this behavior is the interplay between Global Average Pooling (GAP) and the KL divergence objective. GAP collapses spatial dimensions into a single 128-dimensional vector by averaging across all spatial positions; while this enforces translation invariance and curtails parameter count, it also discards localized spatial coordinates. Combined with a modest training cohort of 40 volumes, minimizing relative entropy strongly penalizes predicting near-zero probability in populated bins ($\log \hat{p}_k \to -\infty$), which may encourage the model to converge toward a broader, more conservative probability distribution. While this interpretation aligns with the observed entropy increase, isolating the respective contributions of spatial pooling versus the loss formulation remains an open question requiring ablation studies with coordinate-preserving heads and alternative divergences.
2) **Valley Bridging in Bimodal Packings:** For bimodal targets, the network reliably identifies both modal locations but partially fills the intermediate trough (Fig. 2b). Predicting a smooth bridge between two peaks represents a safer local minimum under KL loss than risking misaligned sharp peaks when training volume diversity is bounded.
3) **Absence of Global Diameter Bias:** The mean signed D_{50} error across test samples is $+0.043$ vx, confirming that the CNN prediction head does not suffer from systematic diameter shrinkage or dilation.

C. Scientific Limitations

We delineate three key limitations of this study:

1) **Synthetic Spherical Phantoms:** Real granular materials (such as sands, ores, and pharmaceutical powders) possess angular, ellipsoidal, or irregular morphologies with surface roughness and internal porosity that alter X-ray attenuation.
2) **Packing Density Regime:** Evaluated packing fractions were constrained to 15%–25%. In dense assemblies ($\phi > 0.40$), particle contacts proliferate and distance transform maxima merge, which is known to degrade classical watershed performance.
3) **Sample Size and Uncertainty:** The pilot dataset comprises 60 synthetic volumes (48^3 voxels). While sufficient for an exploratory feasibility benchmark, 40 training volumes is a small cohort for 3D representation learning, and evaluating on 10 held-out test volumes yields relatively wide bootstrap confidence intervals ($[+0.184, +0.316]$ vx for EMD difference).

D. Physical and Practical Engineering Implications

In materials engineering and granular physics, Particle Size Distribution directly governs continuum transport properties:

1) **Hydraulic Permeability Estimation:** The Kozeny-Carman relationship dictates that porous permeability scales quadratically with median diameter: $k \propto D_{50}^2$. For Compact3DCNN, the observed D_{50} relative error on test volumes is $4.00\% \pm 2.80\%$. In continuum fluid modeling, an error of 4% in D_{50} propagates to approximately $(1 + 0.04)^2 - 1 \approx 8.2\%$ error in estimated permeability. For rapid in-situ screening of borehole core samples or filtration packed beds, this level of fidelity provides viable order-of-magnitude transport characterization in milliseconds.
2) **Sintering Kinetics and Powder Compaction:** In powder metallurgy and ceramic processing, initial-stage neck growth and thermal shrinkage rates scale inversely with particle radius ($d(\Delta L/L_0)/dt \propto r^{-p}$, where $p \in [3, 4]$ depending on whether lattice or grain-boundary diffusion dominates). Because the CNN tends to over-smooth bimodal distributions and broaden fine-particle tails, relying directly on CNN outputs for fine-fraction estimation could lead to overestimating sintering reactivity, underscoring the necessity of classical segmentation when tail precision is critical.

E. Computational Complexity and Scaling Analysis

The two methodologies exhibit fundamentally distinct asymptotic complexity regimes:

- **Morphological Watershed:** 3D Euclidean distance mapping (via separable parabolic envelopes or exact vector transforms) scales with spatial voxel count $\mathcal{O}(N_x N_y N_z)$, while topological marker-directed flooding requires priority-queue operations scaling as $\mathcal{O}(V \log V)$ where V is foreground voxel volume. On full-scale synchrotron tomograms (1024^3 voxels), classical execution requires several minutes of CPU time.
- **Compact3DCNN:** The network forward pass comprises a fixed series of 3D tensor contractions whose depth is invariant to volume contents. On GPU architectures supporting tensor parallelism, batched 3D convolutions process volumes in constant wall-clock latency (1.19 ms for 48^3 grids). This high throughput enables continuous inline quality inspection and real-time 4D tomography streaming where multi-minute offline segmentation is infeasible.

VI. Conclusion and Future Outlook

This research presented a rigorous, leak-free comparative benchmark evaluating direct 3D CNN volumetric histogram regression against classical 3D morphological watershed segmentation on synthetic packed-particle microtomography. Our empirical findings demonstrate:

1) Classical morphological watershed attained superior distribution fidelity (W_1 EMD $= 0.0974 \pm 0.0261$ vx vs. 0.3428 ± 0.0949 vx for CNN; paired 95% CI: $[+0.1837, +0.3156]$ vx) on dilute-to-moderate spherical packings ($\phi \in [15\%, 25\%]$) where non-overlapping spheres provide clean, unambiguous distance maxima.
2) Compact3DCNN (300,900 parameters) successfully demonstrated volume-specific representation learning, achieving a 36.0% validation EMD reduction over an empirical training-set-mean baseline.
3) Forward inference of the CNN executed in 1.19 ± 0.27 ms on an NVIDIA RTX 3050 Laptop GPU compared to 33.71 ± 7.50 ms for the watershed pipeline on an Intel CPU, highlighting a practical throughput advantage for online screening.
4) Direct training on discrete cumulative L_1 Wasserstein loss yielded substantially worse validation performance (1.1382 vx EMD) compared to Kullback-Leibler divergence (0.3970 vx EMD), reflecting optimization challenges induced by piecewise-constant subgradients.

Future research should expand in three key directions: (i) benchmarking dense assemblies near random close packing ($\phi \approx 0.64$) where touching grain contacts induce severe watershed under-segmentation; (ii) evaluating non-spherical, faceted, and irregular mineral geometries; and (iii) investigating coordinate-preserving multi-scale feature pyramids and entropic regularized optimal transport (Sinkhorn divergence) to prevent peak over-smoothing while retaining gradient stability.

References

[1] S. Torquato, T. M. Truskett, and P. G. Debenedetti, "Is random close packing of spheres well defined?" *Physical Review Letters*, vol. 84, no. 10, pp. 2064–2067, 2000.

[2] A. P. Sheppard, R. M. Sok, and H. Averdunk, "Techniques for image enhancement and segmentation of tomographic images of porous materials," *Physica A: Statistical Mechanics and its Applications*, vol. 339, no. 1-2, pp. 145–151, 2004.

[3] S. Beucher and F. Meyer, "The morphological approach to segmentation: the watershed transformation," in *Mathematical Morphology in Image Processing*, E. R. Dougherty, Ed. Marcel Dekker, 1993, pp. 433–481.

[4] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[5] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[6] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
[7] C. Villani, *Optimal Transport: Old and New*. Springer Science & Business Media, 2009, vol. 338.
[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.