Validation of Generative AI Microstructure Prediction Model Using Optical Micrograph Characterization Metrics

PURDUE UNIVERSITY
School of Materials Engineering

Student Names: Junyeong Ahn, Jacob Goehring, Nathaniel Reynard, Claire Russell
Faculty Advisors: Dr. Arun Mannodi-Kanakkithodi, Dr. Michael S. Titus
Industrial Sponsors: Liam Huston, Stéphane Forsik

ABSTRACT: AI-driven microstructure prediction reduces experimental costs and accelerates metallurgical development; however, validating model accuracy is essential for reliable application. We developed a microstructure quantification workflow to compare statistical features of real micrographs with AI-generated images of PyroMet® 680 Ni-based superalloy. Key metrics include phase fractions, phase distribution, and grain eccentricity. This project aims to expand the training dataset, identify gaps within it, and ultimately improve model performance.

This work is sponsored by Carpenter® Technology Corporation, Reading, PA

CARPENTER® TECHNOLOGY

Background & Objectives

- Carpenter Technology PyroMet 680 is a corrosion resistant, high temperature nickel-based superalloy used for aerospace, energy production, and medical applications.
- Carpenter Technology is developing a neural network based denoising diffusion probabilistic model[1] (DDPM) to predict microstructure development during ingot-to-billet conversion of the alloy.
- Validating the fidelity and accuracy of such generative AI models increases the reliability of its application.
- The project aims to design a microstructure characterization and microstructure statistics comparison workflow to validate the performance of DDPM microstructure generation model. Grain size, phase distribution, and grain eccentricity were calculated from both AI and real micrograph sample spaces and were used as comparison metrics.

PyroMet 680 composition.[3]

Elements	Wt%
Ni	*Balance*
Cr	20.05 - 30.00
Fe	17.00 - 20.00
Mo	8.00 - 10.00
Co	0.50 - 2.50
C	0.05 - 0.15
Si, Mn, W, P, S	≤1.00

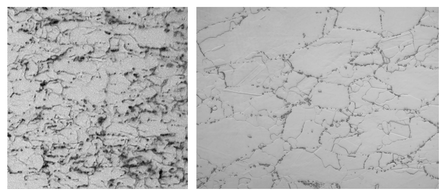

Figure 1. AI-generated (left) and optical (right) micrographs of sample d011 (T = 2000 °F, ε = 0.4, ε̇ = 0.01 s^{-1}, post-forging hold time = 0.1 s).

Methodology

1. Sample Preparation & Micrograph Collection

- Samples experienced isothermal compression from the Gleeble® 3800 thermomechanical processing instrument under various temperature, strain, strain rate, and post-forging hold time.
- Optical micrographs of sample cross section were taken at 1000X and 500X magnifications.
- Samples and AI images were provided by the sponsor.

2. ImageJ Integrated Phase Segmentation

- ML-based ImageJ-FIJI Weka segmentation and Labkit segmentation plugins segmented images into phases: primary carbide, secondary carbide, and grain.
- Each Weka and Labkit classifier was trained by hand drawn segmentation of an image or a set of images before applying classifier to the entire micrograph collection (Figure 2).

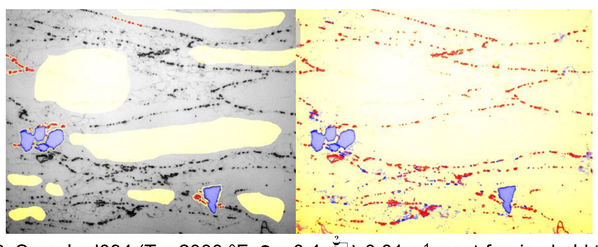

Figure 2. Sample d094 (T = 2000 °F, ε = 0.4, ε̇ = 0.01 s^{-1}, post-forging hold time = 0.1 s) training segmentation for Labkit Classifier 4; hand-segmented (left) vs. trained image (right)

3. Binary Segmentation and Phase Fraction Analysis

- From the segmented images, Python `scikit-image` library calculated the fractions of each phase.
- Phase fractions were calculated as the ratio of the number of pixels occupied by grains to the total number of pixels (2592 × 1944).
- Parity plots were generated for comparison.

4. 2-point Correlation for Phase Distribution

- To analyze the clustering of each phase, the `TwoPointCorrelation` function within Materials Knowledge System in Python (`PyMKS`) package[4] computed the 0-0, 0-1, and 1-1 correlation of each phase.
- Correlation results from AI images and real images were compared by cosine similarity scores.

$$Cosine\ similarity(x, y) = \frac{x \cdot y}{|x| \cdot |y|}$$

where x, y are vectors (matrices) to compare

5. Angular Resolved Chord Length Distribution[5] (ARCLD)

Chords are defined as the lines that cross the area of interest. Bresenham's line algorithm is used to find the chords that cross each phase at varying angles, and distributions of the chord lengths at each angle were derived.

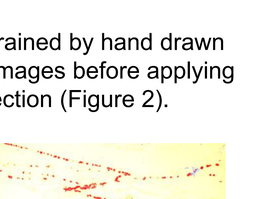

Figure 3. Bresenham's algorithm.

Results & Discussion

1. Classifier Development History

Plug-in	Name	# of Images in Training	Description
Weka	Classifier 1	1	Single image outline experimentation; not applied to other images
Weka	Classifier 2	1	Features were outlined
Weka	Classifier 3	1	Features were shaded; image application failed due to crashing
Labkit	Classifier 1	1	Shaded; applied to 1000x images
Labkit	Classifier 2	1	Shaded; applied to 1000x and AI-gen images
Labkit	Classifier 3	1	Shaded; applied to 1000x and AI-gen images
Labkit	Classifier 4	6	Shaded image stack; applied to 1000x and AI-gen image stacks

2. Comparison of Phase Fractions

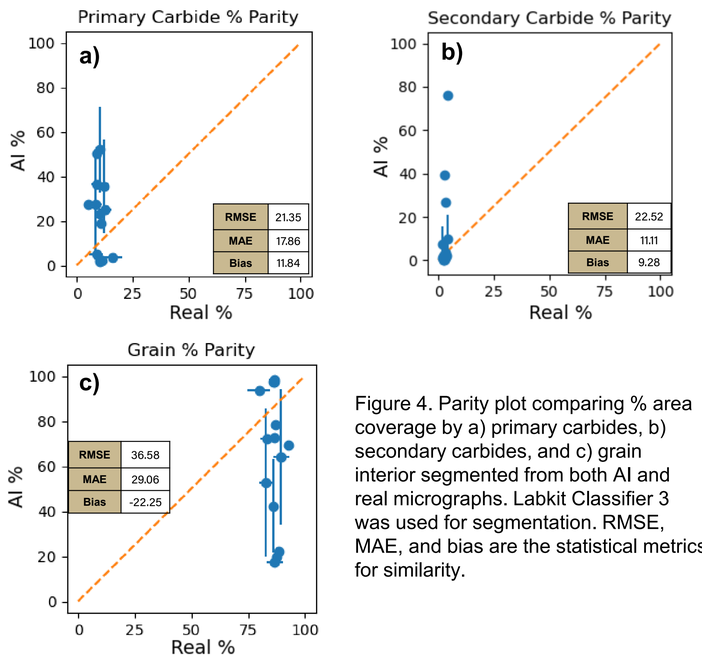

Figure 4. Parity plot comparing % area coverage by a) primary carbides, b) secondary carbides, and c) grain interior segmented from both AI and real micrographs. Labkit Classifier 3 was used for segmentation. RMSE, MAE, and bias are the statistical metrics for similarity.

Parity plots and histograms were created to show the similarity between the optical and synthetic microstructures. This does not directly represent the classifier performance.

3. 2-point Correlation and Similarity Scores

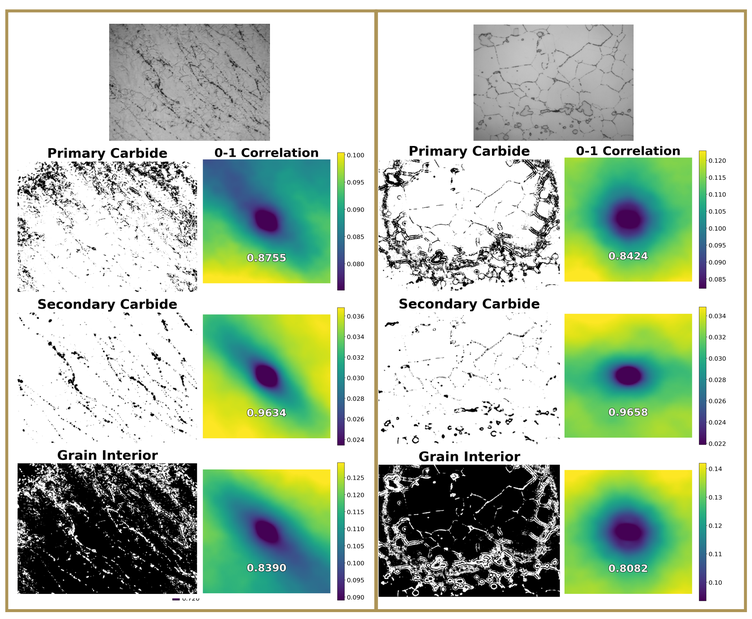

Figure 5. 2-point correlation heatmap of classified phases with high eccentricity (left) and low eccentricity microstructures. Labkit Classifier 3 was used for segmentation.

0-1 corresponds to the probability of finding another phase at a given displacement vector. Shades around the center represent grain elongation.

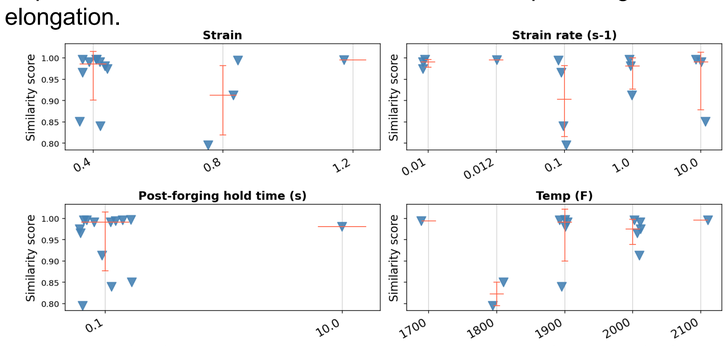

Figure 6. Similarity scores between 0-0, 0-1, and 1-1 correlation results from AI and real images categorized by processing parameters.

Due to limited number of synthetic images, 14 samples were compared. All similarity scores from correlation results lie between 0.8 and 1, indicating high similarity of phase distributions.

Results & Discussion Cont.

4. Chord Length Distributions and Dimensionality Reduction

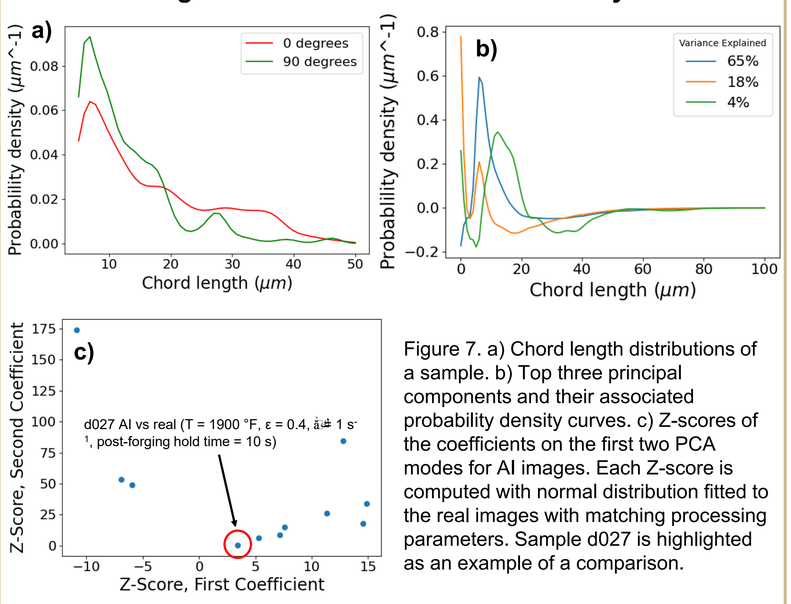

Figure 7. a) Chord length distributions of a sample. b) Top three principal components and their associated probability density curves. c) Z-scores of the coefficients on the first two PCA modes for AI images. Each Z-score is computed with normal distribution fitted to the real images with matching processing parameters. Sample d027 is highlighted as an example of a comparison.

The first three principal components cover 87% of the variance among CLDs. The linear combination of these three will replicate 87% of the curves shown in Figure 7a.

We validated AI images using two normal distributions fitted to the coefficients on the first two principal components of real images for each set of processing parameters. From these we computed Z-scores for the AI images' coefficients. This highlights the offset of CLDs between the AI and real images of the same processing history. Extreme Z-scores indicate a large discrepancy of CLDs between AI and real images.

Conclusion & Recommendations

1. Different plugins and number of images used for each classifier training resulted in classifiers with varying strengths and weaknesses.
   - ➢ Ensemble of multiple classifiers trained on specific processing parameters or microstructural characteristics could improve model performance and versatility.
2. Phase fractions show large discrepancy between AI and real micrographs.
3. Similarity scores comparing correlation results indicate that the phase distribution is similar between AI and real micrographs even though the images are visibly different.
   - ➢ Other metrics (phase fractions, chord length distribution) must be included for comparison.
4. Chord length distributions from real images do not demonstrate a strong trend, as the first couple of distribution modes capture less than 90% of total variance.
   - ➢ Running linear regression on the coefficients where the independent variables are the processing parameters could capture processing parameter dependency of CLDs.

Acknowledgements

The senior design team would like to thank Liam Huston, Stéphane Forsik, and Wesley Roth from Carpenter Technology and Prof. Mannodi and Prof. Titus for their guidance throughout the project. We would like to also thank Austin Dicus and Mario Epler for a successful visit to Carpenter headquarters.

References

1. Azqadan, E., Jahed, H., & Arami, A. (2023). Predictive microstructure image generation using denoising diffusion probabilistic models. *Acta Materialia*, 261, 119406. https://doi.org/10.1016/j.actamat.2023.119406
2. Mendez, J. G., Heffern, W. J., Hagaman, S. D., Troper, M. F., Tucker, V., Huston, L., Dicus, A. D., Epler, M. E., Krane, M. J. M., Titus, M. S., & Forsik, S. A. J. (2024). Modeling microstructural development during hot working of Ni-based superalloy Alloy 680. *Minerals, Metals and Materials Series*, 1081–1090. https://doi.org/10.1007/978-3-031-63937-1_100
3. Carpenter Technology Corporation. (1991). *PyroMet® 680 datasheet*. Retrieved from https://www.carpentertechnology.com/alloy-finder/pyromet-680
4. Brough, D. B., Wheeler, D., & Kalidindi, S. R. (2017). Materials knowledge systems in Python—a data science framework for accelerated development of hierarchical materials. *Integrating Materials and Manufacturing Innovation*, 6, 36–53. https://doi.org/10.1007/s40192-017-0089-0
5. Kamat, S., Tucker, V., Titus, M. S., & Wagner, G. J. (2025). A high-throughput physics- and data-driven framework for high-entropy alloy development. *Acta Materialia*, 292, 121045. https://doi.org/10.1016/j.actamat.2025.121045