Data-driven Analytical Models for Porosity Detection in Thermal Melt Pool Images

Arizona State University

Pius Gyamenah
Keval Parmar
Karteek Menda

Outline

- Objectives
- Data Collection (Experimental Design)
- Dataset Description
- Data Visualization
- Methodology
- Hyperparameter Tuning/Optimization
- User Interface
- Future Work

Objectives

- Visualize thermal porosity characterization data of Additively Manufactured Ti–6Al–4V Thin-walled Structure via Laser Engineered Net Shaping.

- Perform feature extraction using a pre-trained VGG16 model to extract meaningful features from the images after conversion from csv

- Train machine learning algorithms (Random Forest and XGBoost) to detect the presence of porosity and estimate porosity size.

- Conduct a Random Search hyperparameter search to optimize the performance of the machine learning models.

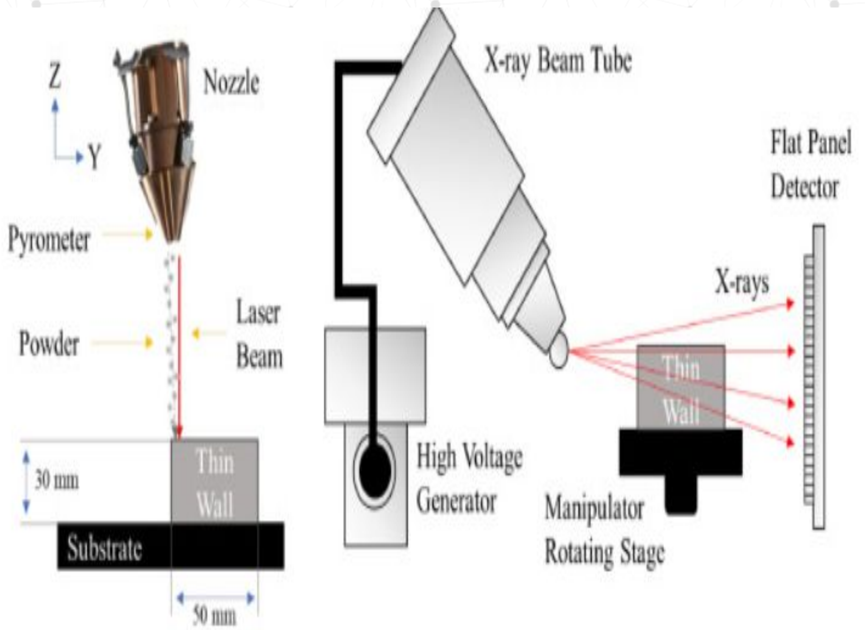
- Evaluate the effectiveness of the trained models in accurately detecting porosity and estimating porosity size in the manufactured structures.

- Provide insights into the potential applications of machine learning models for thermal porosity characterization in additive manufacturing processes.

- Contribute to the advancement of quality control and defect detection methodologies in additive manufacturing industries.

Data Collection (Experimental design)

- Pyrometer Captures melt pool images which are cropped and saved to CSV
- X-ray Computed Tomography camera captures porosity records which are also saved to an Excel CSV



Christian Zamiela, Wenmeng Tian, Shenghan Guo, Linkan Bian, Thermal-porosity characterization data of additively manufactured Ti–6Al–4V thin-walled structure via laser engineered net shaping,Data in Brief,Volume 51,2023,109722,ISSN 52-3409,https://doi.org/10.1016/j.dib.2023.109722.

Dataset Description

- The data presented here consists of a folder with 1,564 cropped pyrometer melt pool images in CSV format and and thermal-porosity table in Excel format.

- The thin wall builds coordinates *Y (mm)* and *Z (mm)* within the thermal-porosity Excel sheet are linked to XCT porosity coordinates within the 0.5 mm range.

- The *Time (Sec)* is based on the Frame Rate of the Pyrometer.

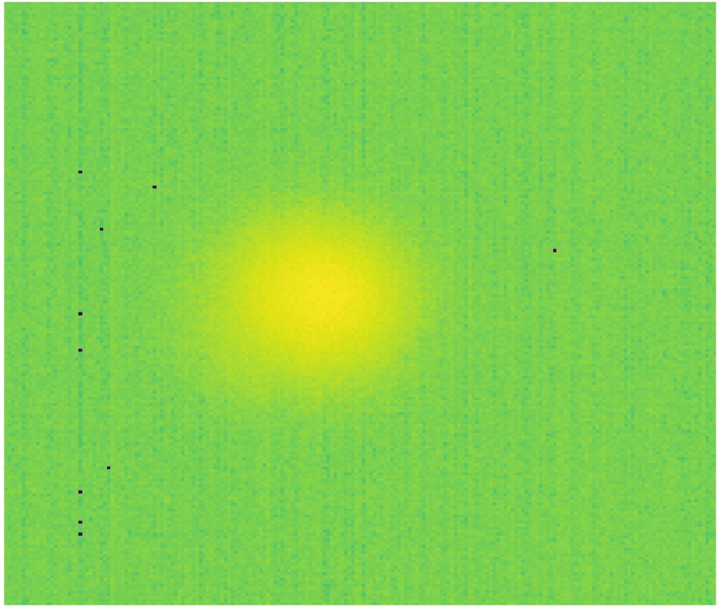
- The x-coordinate is not present in the thermal-porosity Excel sheet because it is constant due to moving in one direction along the y-axis.

- *MP Area, MP Eccentricity, Tmax, Row,* and *Column* provide descriptive characteristics of the heat-affected zone (HAZ) and shape of the melt pool for melt fusion.

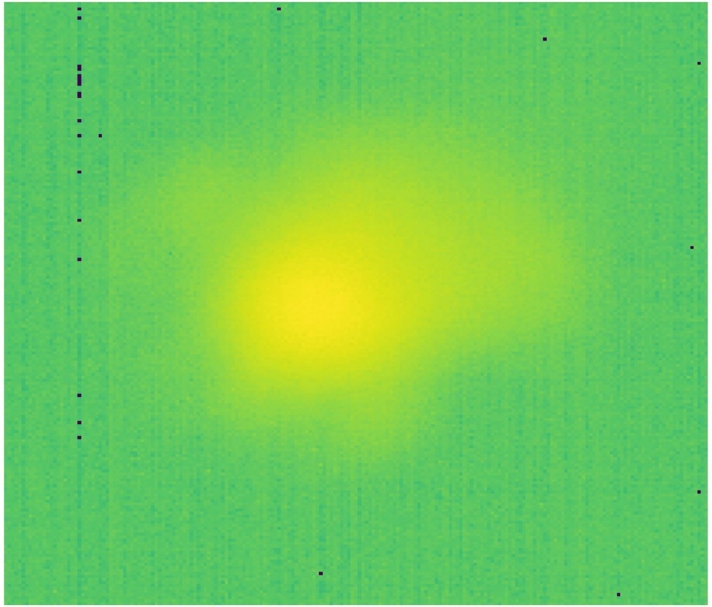
Data Summary

Parameters	Summary Values
Pyrometer CSVs	
Array size	200×200
Pixel pitch	$6.45\ \mu m$
Temperature Range	1000-2500°C
Frame Rate	~ 6.67 fps
Eccentricity Range	$0.13 - 0.95\ mm$
Maximum Temperature Range	1602-2107°C
Melt Pool Area Range	$0 - 7274\ pixels$
Number of CSVs Per Layer Range	16-34 layers
Average Signal-to-Noise Ratio (SNR)	10.97
XCT Porosity Labels	
Binary Labels	0=No Porosity, 1=Porosity
Number of Label Occurrences	0=1490, 1=70
Porosity Size Range	$0.05 - 0.98\ mm$
Image Class Correlation	0.78
Thin Wall Build Parameters	
Build Layers	60
Print Direction	Unidirectional
Print Speed	$12.7\ mm/s$
Layer Thickness	$0.508\ mm$

Data Visualization



Frame with Porosity



Frame without Porosity

Methodology

Algorithms Used:

- Random Forest
- XGBoost

Why Only above Algorithms:

- **Ensemble Learning :** They build **multiple models** and **combine** them to make more accurate predictions. This helps in **reducing overfitting** and **improving generalization**
- **High Accuracy :** They can capture **complex relationships** in the data and handle a **large** number of **features** effectively
- **Handling Non-Linearity :** They can **capture non-linear relationships** in the data, making them suitable for a wide range of problems where the **underlying patterns** are **complex**
- **Robustness to Overfitting :** Both algorithms have mechanisms to **control overfitting**
  - Random Forest uses **random feature selection** and **bagging**
  - XGBoost includes **regularization** terms in its **objective function**
- **Parallelization :** They can be parallelized, allowing for **efficient computation** on **multicore** machines, which can lead to **faster training times**

Hyperparameter Tuning

- Hyperparameter tuning: Finding the **optimal set** of **hyperparameters** for ML Models

Hyper Parameter Tuning for Improved performance

- Adjusting hyperparameters results in **improved performance**. This may involve **increasing** or **decreasing** the **learning rate**, **adjusting the depth of a decision tree, or fine-tuning other parameters**

- **O**verfitting occurs when a model learns the training data too well, including its noise, and performs poorly on new, unseen data. Underfitting occurs when a model is too simple and cannot capture the underlying patterns in the data. **Hyperparameter** tuning helps **strike** a **balance** to avoid these issues

Hyper Parameter tuning for Imbalanced Data

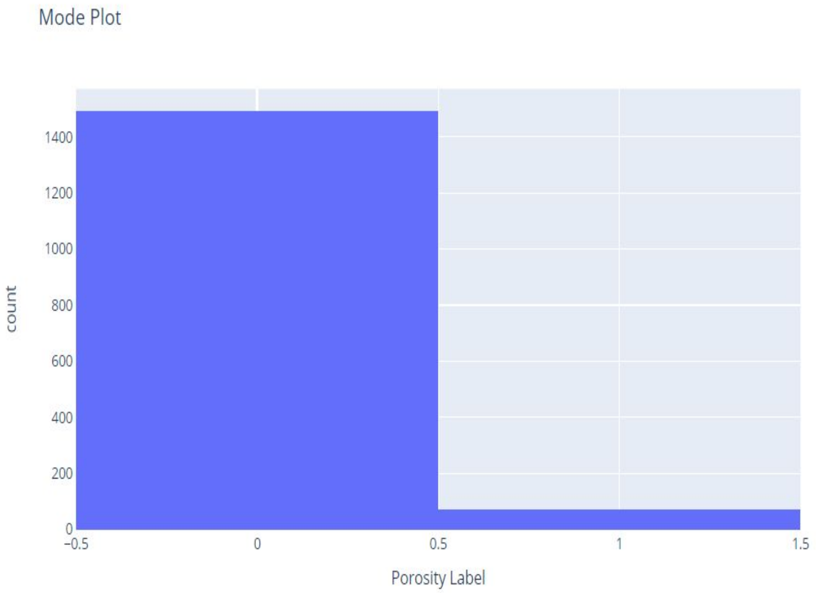
- When dealing with an imbalanced dataset, where the distribution of classes is uneven, hyperparameter tuning becomes even **more important**

- Imbalanced datasets can pose challenges because a model may become **biased** towards the **majority class**, leading to **poor performance** on the **minority class**

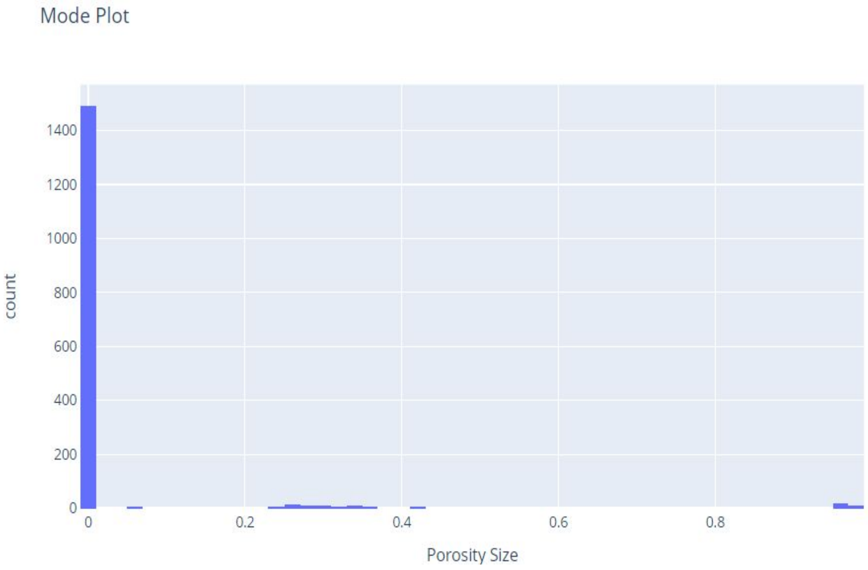
- Hyperparameter tuning can help **mitigate** this issue by **optimizing** the model's ability to handle imbalanced data and contributes to the **generalization ability** of a model

Target Feature Analysis

Classification Target(Porosity Label)



Regression Target(Porosity Size(mm))



Classification(Random Forest)

Random Forest with default Parameters

```
Confusion_matrix
 [[299   0]
 [ 10   4]]
Accuracy: 0.9680511182108626
Precision: 1.0
Recall: 0.2857142857142857
F1 Score: 0.4444444444444445
roc_auc_score: 0.6428571428571428

Classification Report:
               precision    recall  f1-score   support

         0.0       0.97      1.00      0.98       299
         1.0       1.00      0.29      0.44        14

    accuracy                           0.97       313
   macro avg       0.98      0.64      0.71       313
weighted avg       0.97      0.97      0.96       313
```

Random Forest with tuned Parameters

```
Confusion_matrix
 [[296   3]
 [  5   9]]
Accuracy: 0.9744408945686901
Precision: 0.75
Recall: 0.6428571428571429
F1 Score: 0.6923076923076924
roc_auc_score: 0.8164118490205446

Classification Report:
               precision    recall  f1-score   support

         0.0       0.98      0.99      0.99       299
         1.0       0.75      0.64      0.69        14

    accuracy                           0.97       313
   macro avg       0.87      0.82      0.84       313
weighted avg       0.97      0.97      0.97       313
```

Classification(XGBoost)

XGBoost with default Parameters

```
Confusion_matrix
 [[297   2]
 [  4  10]]
Accuracy: 0.9808306709265175
Precision: 0.8333333333333334
Recall: 0.7142857142857143
F1 Score: 0.7692307692307692
roc_auc_score: 0.853798375537506

Classification Report:
              precision    recall  f1-score   support

         0.0       0.99      0.99      0.99       299
         1.0       0.83      0.71      0.77        14

    accuracy                           0.98       313
   macro avg       0.91      0.85      0.88       313
weighted avg       0.98      0.98      0.98       313
```

XGBoost with tuned Parameters

```
Confusion_matrix
 [[297   2]
 [  3  11]]
Accuracy: 0.9840255591054313
Precision: 0.8461538461538461
Recall: 0.7857142857142857
F1 Score: 0.8148148148148148
roc_auc_score: 0.8895126612517916

Classification Report:
              precision    recall  f1-score   support

         0.0       0.99      0.99      0.99       299
         1.0       0.85      0.79      0.81        14

    accuracy                           0.98       313
   macro avg       0.92      0.89      0.90       313
weighted avg       0.98      0.98      0.98       313
```

Regression

Random Forest with default Parameters

```
Mean Squared Error (MSE): 0.0060568050777178
Mean Absolute Error (MAE): 0.018287519403401852
R-squared (R2): 0.5686842096026739
Explained Variance Score: 0.5700427303188629
```

XGBoost with default Parameters

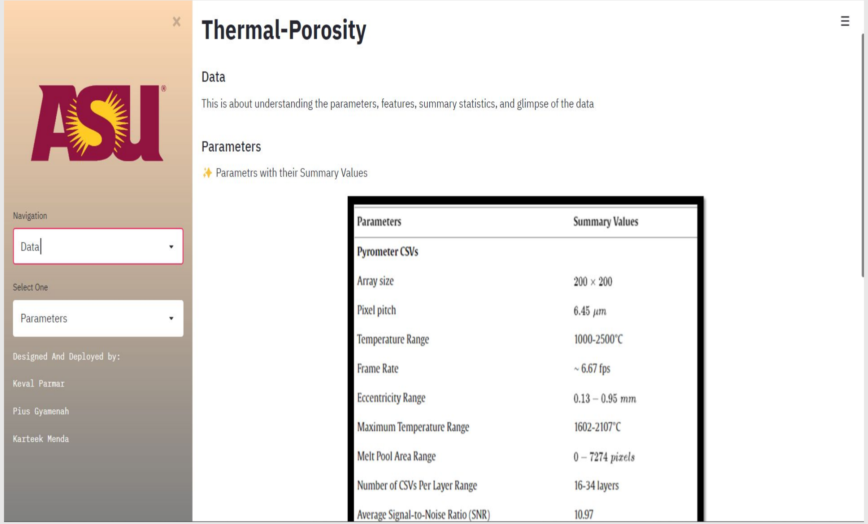
```
Mean Squared Error (MSE): 0.006484027884658444
Mean Absolute Error (MAE): 0.018085941952022514
R-squared (R2): 0.5382609187278748
Explained Variance Score: 0.541304612795196l
```

Random Forest with tuned Parameters

```
Mean Squared Error (MSE): 0.005851111263269111
Mean Absolute Error (MAE): 0.0185249346559892
R-squared (R2): 0.5833320295375043
Explained Variance Score: 0.5833326530334406
```

XGBoost with tuned Parameters

```
Mean Squared Error (MSE): 0.006493186825438868
Mean Absolute Error (MAE): 0.0181848739508939
R-squared (R2): 0.5376086943734137
Explained Variance Score: 0.5406365154155928
```



User Interface

1. Target Features Distributions(Classification and Regression)
2. Model Metrics Comparison
3. Prediction on New Data

(Will host the web application after implementing the future work and will make it publicly available)

Web App Predictions

Input Images

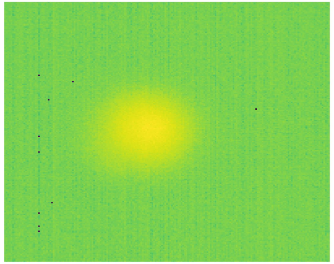


Image-1
Ground Truth : 1(Porosity)

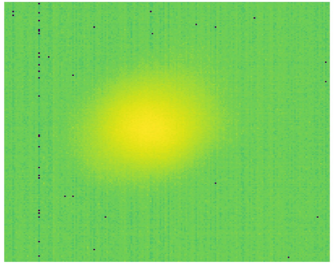


Image-2
Ground Truth : 1(Porosity)

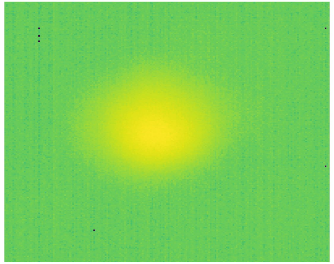


Image-3
Ground Truth : 0 (No Porosity)

Predictions made

The prediction is [1.] which is Porosity Present

The Porosity Size is [0.29502049] mm

The prediction is [1.] which is Porosity Present

The Porosity Size is [0.43790621] mm

The prediction is [0.] which is Porosity Not Present

Future Work

1. As the **Porosity Size(mm)** is a **highly zero-inflated feature**, we need to **come up** with more sophisticated algorithms like **Zero Inflated Poisson Regression**, and **Zero Inflated Negative Binomial Regression** which could very well take care of these scenarios.
2. We want to come up with the **explainability** of the predictions made by our models. So, we need to **highlight** those portions(**some pixels of interest**) of the **thermal images** that **impacted** the **decision** made by the classification algorithm
3. **Finding** and **marking** the exact shape, locations, and distribution of the pores. Spherical pores created by gas are much less detrimental than a lack of fusion between layers - as it is much easier to separate layers that have not fused properly. (also essential for post-processing e.g hot isostatic press (HIP))
4. **Beyond detection**, future work could also focus on finding the **source of the porosity**

References

1. Christian Zamiela, Wenmeng Tian, Shenghan Guo, Linkan Bian,Thermal-porosity characterization data of additively manufactured Ti–6Al–4V thin-walled structure via laser engineered net shaping,Data in Brief,Volume 51,2023,109722,ISSN 352-3409,https://doi.org/10.1016/j.dib.2023.109722.
2. https://iopscience.iop.org/article/10.1088/2053-1591/abcc5d/pdf
3. https://www.ge.com/additive/blog/get-facts-porosity-metal-additive-manufacturing

Thank You